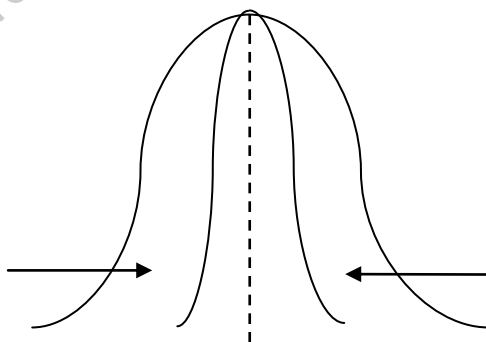
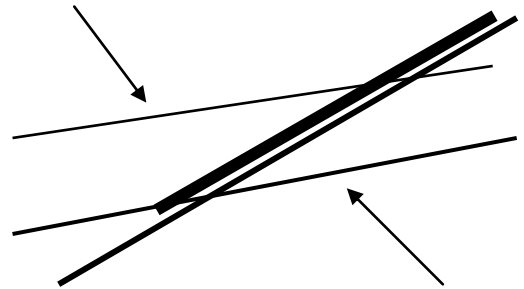
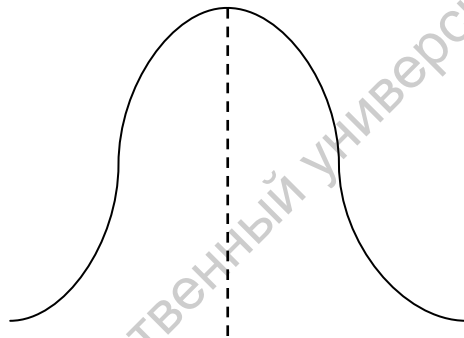


ЮЛИЙ  
СТАВРОПОЛЬСКИЙ

# МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИИ



$$y_i = B_0 + B_1X_i + B_2Z_i + B_2X_iZ_i + B_2Z_i^2$$

$$y_i = \beta_0 + \beta_1X_i + \beta_2Z_i + e_i$$

САРАТОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
им. Н. Г. ЧЕРНЫШЕВСКОГО

Ю. В. СТАВРОПОЛЬСКИЙ

**МАТЕМАТИЧЕСКИЕ  
МЕТОДЫ В  
ПСИХОЛОГИИ**

*Учебное пособие для студентов  
психологических отделений вузов*

Издательство Саратовского университета  
2009

УДК 159.9:51:681.3

ББК 88.4 я73

С 76

Ставропольский Ю. В. Математические методы в психологии. Саратов: Изд-во Сарат. ун-та, 2009. 357 с.  
ISBN 978-5-292-03933-4

В пособии представлены основные проблемные области современной математической статистики. Пособие содержит вопросы для самопроверки, тестовые задания, задания для семинаров и контрольных работ, список литературы.

Учебное пособие предназначено для студентов психологических отделений вузов.

Рецензент:

доцент кафедры психологии СГУ им. Н. Г. Чернышевского,  
кандидат психологических наук Е. В. Рягузова

ISBN 978-5-292-03933-4

© Ставропольский Ю. В., 2009

# УСТАНОВОЧНЫЙ МОДУЛЬ

## Описание структуры курса или дисциплины

### Учебник

Учебное пособие «Математические методы в психологии» предназначено для студентов отделений психологии, изучающих одноименную учебную дисциплину.

Структура учебного пособия соответствует структуре изучаемого курса и включает следующие разделы:

### УСТАНОВОЧНЫЙ МОДУЛЬ

#### МОДУЛЬ 1

#### ТЕМА 1. ТЕРМИНОЛОГИЯ СТАТИСТИЧЕСКОЙ ВЫБОРКИ

Выборочное распределение

Выборочная ошибка

Правило 65%, 95% и 99%

Вероятностная выборка

Простая случайная выборка

Стратифицированная случайная выборка

Систематическая случайная выборка

Кластерная (территориальная) случайная выборка

Многоэтапная выборка

Невероятностная выборка

Произвольные выборки

Преднамеренные (целенаправленные) выборки

Модальная выборка

Экспертная выборка

Гетерогенная выборка

Выборка «снежный ком»

Вопросы для самоконтроля по теме «Терминология статистической выборки»

Тестовые задания к теме «Терминология статистической выборки»

#### ТЕМА 2. ВАЛИДНОСТЬ КОНСТРУКТОВ

Измерение различных типов валидности

Валидность транслирования

Лицевая валидность

Содержательная валидность

Критериальная валидность

Прогностическая валидность

Одновременная валидность

Конвергентная валидность

Дискриминантная валидность

Понятие о валидности конструкторов  
Конвергентная и дискриминантная валидность  
    Конвергентная валидность  
    Дискриминантная валидность  
Угрозы валидности конструкторов  
    Монооперациональный уклон  
    Монометодический уклон  
    Взаимодействие различных экспериментальных воздействий  
    Взаимодействие тестирования и воздействия  
    Ограниченная обобщаемость конструкторов  
    Неразделенность конструкторов и уровней конструкторов  
    «Социальные» угрозы валидности конструкторов  
        Формулирование гипотез  
        Предвзятое оценивание  
        Ожидания экспериментатора  
Номологическая сеть  
Матрица мультисвойств-мультиметодов  
    Принципы интерпретации  
    Преимущества и недостатки матрицы мультисвойств-мультиметодов  
    Модифицированная матрица мультисвойств-мультиметодов - исключается фактор методов  
    Что такое конвергентная валидность?  
    Что такое дискриминантная валидность?  
Согласование паттернов конструктивной валидности  
    Согласованность паттернов и валидность конструкторов  
    Преимущества и недостатки согласования паттернов  
Вопросы для самоконтроля по теме «Валидность конструкторов»  
Тестовые задания к теме «Валидность конструкторов»

### ТЕМА 3. НАДЕЖНОСТЬ

Теория истинного значения  
Ошибка измерения  
    Что такое случайная ошибка?  
    Что такое систематическая ошибка?  
    Уменьшение ошибки измерения  
Теория надежности  
Типы надежности  
    Надежность внешних наблюдений  
    Тест-ретестовая надежность  
    Надежность параллельных форм  
    Внутренне согласованная надежность  
        Средняя корреляция между пунктами методики  
        Средняя корреляция по общему баллу

Расколота пополам надежность  
Альфа Кронбаха ( $\alpha$ )

Надежность и валидность

Вопросы для самоконтроля по теме «Надёжность»

Тестовые задания к теме «Надёжность»

#### ТЕМА 4. ШКАЛИРОВАНИЕ

Уровни измерения

Почему уровень измерения важен?

Шкалирование

Общие вопросы шкалирования

Цели шкалирования

Метрика

Одномерность или многомерность?

Основные типы одномерных шкал

Шкалирование Тёрстоуна

Шкалирование Лайкерта

Шкалирование Гуттмана

Вопросы для самоконтроля по теме «Шкалирование»

Тестовые задания к теме «Шкалирование»

Семинар 1. Измерения, шкалы и статистика

#### МОДУЛЬ 2

#### ТЕМА 5. ВНУТРЕННЯЯ ВАЛИДНОСТЬ

Установление причины и следствия

Предшествование во времени

Циклические флуктуации

Ковариация причины и следствия

Отсутствие достоверных альтернативных объяснений

Причины, связанные с единственной группой

Регрессия к среднему

Причины, связанные с множественностью групп

Социальные причины

Вопросы для самоконтроля по теме «Внутренняя валидность»

Тестовые задания к теме «Внутренняя валидность»

#### ТЕМА 6. ВАЛИДНОСТЬ ВЫВОДОВ

Валидность выводов

Угрозы валидности выводов

Улучшение валидности выводов

Статистическая мощность

Матрица принятия статистического решения

Вопросы для самоконтроля по теме «Валидность выводов»

Тестовые задания к теме «Валидность выводов»

#### ТЕМА 7. ДЕСКРИПТИВНЫЕ СТАТИСТИКИ

Подготовка данных

Проверка и отслеживание данных

Проверка данных на точность  
Формирование структуры базы данных  
Ввод данных в компьютер  
Преобразования данных  
Дескриптивные статистики  
Одномерный анализ  
Частот распределения

Вопросы для самоконтроля по теме «Дескриптивные статисти-  
стики»

Тестовые задания к теме «Дескриптивные статистики»

#### ТЕМА 8. КОРРЕЛЯЦИЯ

Пример корреляции

Расчет корреляции

Проверка значимости корреляции

Другие корреляции

Вопросы для самоконтроля по теме «Корреляция»

Тестовые задания к теме «Корреляция»

#### ТЕМА 9. t-КРИТЕРИЙ

Статистические выводы

t-критерий

Статистический анализ с применением t-критерия

Формула стандартной ошибки различия между средни-  
ми

Вопросы для самоконтроля по теме «t-критерий»

Тестовые задания к теме «t-критерий»

Семинар 2. Меры центральной тенденции и показатели ва-  
риации

### МОДУЛЬ 3

#### ТЕМА 10. ИНДИКАТОРНЫЕ (ФИКТИВНЫЕ) ПЕРЕМЕННЫЕ

Индикаторные (фиктивные) переменные

Вопросы для самоконтроля по теме «Индикаторные (фик-  
тивные) переменные»

Тестовые задания к теме «Индикаторные (фиктивные) пере-  
менные»

#### ТЕМА 11. ОБЩАЯ ЛИНЕЙНАЯ МОДЕЛЬ

Двумерная линейная модель

Общая линейная модель

Вопросы для самоконтроля по теме «Общая линейная мо-  
дель»

Тестовые задания к теме «Общая линейная модель»

#### ТЕМА 12. ПОСТЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ

Постэкспериментальный анализ

Вопросы для самоконтроля по теме «Постэксперименталь-  
ный анализ»

Тестовые задания к теме «Постэкспериментальный анализ»

### ТЕМА 13. АНАЛИЗ ПО ФАКТОРНОМУ ПЛАНУ

Анализ по факторному плану

Вопросы для самоконтроля по теме «Анализ по факторному плану»

Тестовые задания к теме «Анализ по факторному плану»

Семинар 3. Дисперсионный анализ и факторный анализ

### МОДУЛЬ 4

### ТЕМА 14. РАНДОМИЗИРОВАННЫЙ АНАЛИЗ БЛОКОВ

Рандомизированный анализ блоков

Вопросы для самоконтроля по теме «Рандомизированный анализ блоков»

Тестовые задания к теме «Рандомизированный анализ блоков»

### ТЕМА 15. АНАЛИЗ КОВАРИАЦИЙ

Анализ ковариаций

Вопросы для самоконтроля по теме «Анализ ковариаций»

Тестовые задания к теме «Анализ ковариаций»

### ТЕМА 16. АНАЛИЗ НЕЭКВИВАЛЕНТНЫХ ГРУПП

Требования к анализу

Наглядный пример

Проблема

Регрессия и ошибка измерения

Каким образом регрессия вписывается в линии

Каким образом ошибка измерения влияет на наклон

Почему эта проблема не возникает при рандомизированных планах?

Обобщение проблемы

Решение проблемы

Ковариационный анализ (ANCOVA) со скорректированной надежностью

Повторное рассмотрение примера

Какую надежность предпочесть?

Вопросы для самоконтроля по теме «Анализ неэквивалентных групп»

Тестовые задания к теме «Анализ неэквивалентных групп»

### ТЕМА 17. АНАЛИЗ ПЕРЕГИБА ЛИНИИ РЕГРЕССИИ

Требования к анализу

Допущения при анализе

Отсекающий критерий

Распределение до и после эксперимента

Дисперсия доэкспериментальных значений в контрольной группе

Непрерывность доэкспериментального распределения

Реализация эксперимента

Проблема криволинейности



Конкретизация модели

Анализ стратегии

Этапы анализа

Преобразование доэкспериментальных данных

Визуальное рассмотрение взаимосвязи

Конкретизация членов высшего порядка и взаимодействий

Оценка исходной модели

Усовершенствование модели

Пример анализа

Вопросы для самоконтроля по теме «Анализ перегиба линии регрессии»

Тестовые задания к теме «Анализ перегиба линии регрессии»

## ТЕМА 18. АНАЛИЗ ВЫТЕСНЕНИЯ ТОЧЕК РЕГРЕССИИ

Статистические требования

Анализ вытеснения точек регрессии

Вопросы для самоконтроля по теме «Анализ вытеснения точек регрессии»

Тестовые задания к теме «Анализ вытеснения точек регрессии»

Семинар 4. Алгебраические модели психических процессов

## МОДУЛЬ 5

Контрольные задания по дисциплине «Математические методы в психологии»

Вариант I

Вариант II

Вариант III

Вариант IV

Вариант V

### Описание дисциплины

Современное развитие психологии связано с междисциплинарным подходом, в котором большое место занимает математика. Задачи поиска функциональных и качественных решений в психологии решаются на уровне дескриптивной статистики, корреляционного, дисперсионного и факторного анализа. Чем сложнее психологическая задача, тем более сложные разделы математики привлекают внимание психологов. Процесс математизации психологической проблемы всегда протекает в борьбе двух тенденций. С одной стороны, необходимо выделить форму изучаемого явления и проанализировать её логически. С другой стороны, не все психологические гипотезы укладываются в известные математические формы, требуя от исследователя способности мыслить гибко. В данном пособии ставилась задача помочь студентам психологических отделений усвоить наиболее важные разделы математической статистики на конкретных примерах из области психологии. Уме-

ние представлять психологические данные в числовом виде развивает у студента способность к количественному анализу психических явлений. Речь идет и о приборных измерениях, и о подсчёте экспертных оценок и тестовых баллов. Переход от количественных критериев к качественным развивает у студентов способность проводить интерпретацию и психологический анализ изучаемого явления.

### **Требования**

Необходимо до 12 числа каждого месяца выбрать тему для подготовки к семинарскому занятию, до 16 числа каждого месяца необходимо ознакомиться с изучаемым материалом в виде модуля по дисциплине, для того, чтобы на последующей интерактивной консультации после 20 числа каждого месяца по дисциплине задать возникшие вопросы своим одногруппникам или непосредственно преподавателю в ходе обсуждения. До 22 числа каждого месяца необходимо подготовить выбранные темы к семинарскому занятию, которое также как и консультация пройдет в интерактивном (голосовом) режиме. Окончание изучения дисциплины оформляется в виде контрольной работы, которую следует оформлять в соответствии с установленными требованиями и направлять в отдел дистанционного обучения не позднее установленного срока.

### **Система оценивания**

Выполнение каждого модуля совокупно оценивается в 100 баллов. Эта суммарная оценка складывается из оценок за выполнение отдельных видов заданий: вопросов для самоконтроля после каждой темы, тестовых заданий после каждой темы и семинарских заданий по завершению каждого модуля.

Ответы на задания для самоконтроля в модуле оцениваются в 40 баллов.

Выполнение тестовых заданий модуля оценивается в 40 баллов.

Подготовленная тема семинарского занятия оценивается в 20 баллов.

Сумма = 100 баллов на один модуль.

Итого: 4 учебных модуля = 400 баллов + 1 модуль промежуточной аттестации (100 баллов) = 500 баллов по дисциплине за один семестр.

Выполненная контрольная работа максимально оценивается в 100 баллов.

### **Задачи**

Завершив изучение курса «Математических методов в психологии», студент должен уметь самостоятельно оперировать математическим аппаратом психологического исследования в следующих пределах:

- ❖ уметь самостоятельно рассчитать выборку испытуемых, исходя из целей и задач собственного исследования;
- ❖ определить валидность исследуемых конструктов;
- ❖ оценить надежность используемого измерительного инструментария;

- ❖ устранить причины, угрожающие валидности проводимого исследования;
- ❖ применять в исследовании одномерное шкалирование;
- ❖ уметь самостоятельно вычислить дескриптивные статистические показатели;
- ❖ уметь вычислять корреляции;
- ❖ вычислять статистические критерии (t-критерий);
- ❖ знать общую линейную модель и основные подходы к её анализу.

### **Умения**

Усвоенные навыки и умения позволят студенту

- ❖ самостоятельно организовывать психологическое исследование;
- ❖ интерпретировать результаты проведенного исследования;
- ❖ избегать распространенных ошибок при проведении психологического исследования;
- ❖ понимать опубликованные отчеты и интерпретации данных психологических исследований;
- ❖ мыслить самостоятельно.

### **Сведения об авторах**

Автор-составитель данного учебного пособия Ю. В. Ставропольский. При составлении пособия использованы библиографические источники, указанные в списке литературы.

### **Методические рекомендации**

В процессе изучения предмета следует сосредоточить свое внимание на главных, основополагающих моментах, а именно:

- уяснить основную терминологию математической статистики, имеющую отношение к психологическим исследованиям, понимать её назначение и точный смысл и уметь ею оперировать;
- усвоить статистические приёмы количественного исследования человеческого поведения;
- изучить формулы и примеры вычислений и оценок основных показателей математической статистики, таких как корреляции, статистические критерии, шкалирование, дескриптивные показатели.
- представлять себе особенности организации психологического измерения и интерпретации полученных данных;
- выработать у себя способность систематически мыслить и анализировать психологические явления с применением усвоенных знаний;
- сформировать в себе культуру математического мышления при обращении к психологическим явлениям.

В процессе самостоятельной работы студенты не должны ограничиваться материалами настоящего пособия. Следует шире обращаться к опубликованным отчетам о психологических исследованиях и пытаться самостоятельно разобраться в приводимых в них данных.

## Самостоятельная работа

Самостоятельная работа студентов – это многообразные виды индивидуальной и коллективной деятельности студентов, осуществляемые под руководством, но без непосредственного участия преподавателя в специально отведенное для этого аудиторное и внеаудиторное время. Это особая форма обучения по заданиям преподавателя, выполнение которых требует активной мыслительной деятельности. Методологическую основу самостоятельной работы студентов составляет деятельностный подход, когда цели обучения ориентированы на формирование умений решать типовые и нетиповые задачи, т.е. на реальные ситуации, где студентам необходимо обнаруживать знания по конкретной дисциплине.

Рекомендуемые этапы и приемы самостоятельной работы студентов:

1. Подбор рекомендуемой литературы.
2. Знакомство с вопросами плана семинарского занятия и методической разработки по данной теме.
3. Определение вопросов, по которым нужно прочитать или конспектировать литературу.
4. Составление схем, таблиц на основе текста лекций, учебника, монографии.

**Рекомендуемый комплекс средств обучения при самостоятельной работе студентов:**

- Методические разработки для студентов с основным содержанием курса.
- Матрица внутрипредметных связей.
- Дидактический раздаточный материал.
- Обзорный конспект лекций, вопросы лекции.
- Диа-, кино-, видеофильмы.
- Сборник задач, тесты (контрольные задания) и др.

**Рекомендуемые виды самостоятельной работы студентов:**

1. Конспектирование.
2. Реферирование литературы.
3. Аннотирование книг, статей.
4. Выполнение заданий поисково-исследовательского характера.
5. Углубленный анализ научно-методической литературы, проведение эксперимента.
6. Работа на лекции: составление или слежение за планом чтения лекции, проработка конспекта лекции, дополнение конспекта рекомендованной литературой.
7. Участие в работе семинара: подготовка конспектов выступлений на семинаре, рефератов, выполнение заданий.

8. Лабораторно-практические занятия: действие в соответствии с инструкциями и методическими указаниями, получение результата.
9. УИРС и НИРС при выполнении самостоятельной, контрольной, курсовой и дипломной работ.
10. Контрольная работа – письменное выполнение.
11. Выполнение заданий по наблюдению и сбору материалов в процессе практики.

### Глоссарий

**Альфа Кронбаха** в математическом смысле эквивалентна среднему от всех возможных расколотых оценок

**Валидность конструкторов** – степень, в которой операционализация теоретических конструкторов в исследовании позволяет делать обоснованные выводы

**Валидность транслирования** обращена к тому, насколько хорошо операционализация отражает конструктор. Данный подход по своей природе оказывается определительным – он предполагает, что исследователь обладает детальным определением конструктора и способен проверить операционализацию применительно к конструктору

**Вариационный ряд** – это выборка, упорядоченная по возрастанию значений

**Внутренне согласованная надежность** используется при оценке согласованности результатов по пунктам одного и того же тестового задания

**Выборочная ошибка** – стандартная ошибка в выборочном контексте

**Выборочное распределение** – распределение бесконечного количества выборок одного и того же объема называют

**Гетерогенная выборка** составляется в тех случаях, когда стремятся включить в выборку все возможные мнения и взгляды, а пропорциональность представления этих взглядов для нас не очень важна

**Дискриминантная валидность** направлена на определение того, в какой степени наша операционализация отличается (дивергентна) от других операционализаций, от которых она теоретически должна отличаться

**Дооперациональный** означает «до транслирования конструктора в измерения»

**Интервальное измерение** – дистанция между измерениями играет роль

**Интерквартильная широта** – одна из мер разброса или рассеяния данных. Равняется разности между верхним и нижним квартилями.

**Квантиль** – число  $x_p$ ,  $p$  в интервале  $[0; 1]$ , удовлетворяющее соотношению  $\Pr(X < x_p) \leq p \leq \Pr(X \leq x_p)$  называется  $p$ -квантилем. Процентиль определяется соответственно при значении  $p$ , выраженном в процентах.

**Квартили** – это набор квантилей для  $p = 0,25; 0,5; 0,75$ . Их оценками (квартилями эмпирического распределения) являются величины, делящие выборку данных на четыре группы, содержащие (по возможности) одинаковые количества наблюдений. Когда говорят о квартилях, то обычно

имеют в виду верхний квартиль Q3 и нижний квартиль Q1. Второй квартиль Q2 равен медиане. Нижний квартиль Q1 – это значение, ниже которого в упорядоченном множестве данных находится четверть данных, а верхний квартиль Q3 – это значение, выше которого в упорядоченном множестве данных находится четверть значений. Разность Q3 – Q1 называется интерквартильной шириной и является робастной мерой рассеяния.

**Конвергентная валидность** направлена на определение того, в какой степени данная операционализация сходна (конвергентна) с другими операционализациями, с которыми она теоретически должна быть сходна

**Критериальная валидность** обращена к тому, функционирует ли операционализация таким образом, который предусмотрен теорией конструкта

**Лицевая валидность** рассматривает операционализацию с «лицевой стороны», чтобы определить, насколько хорошо удалось транслирование конструкта в операционализацию

**Медиана** – это точка, по обе стороны которой располагаются одинаковое количество элементов выборки. Медиана является робастной оценкой центральной тенденции

**Медиана выборки** является оценкой медианы распределения

**Медиана распределения** – это точка  $m$ , определяемая условием: вероятность того, что случайная величина примет значение, не превосходящее  $m$ , равна  $\frac{1}{2}$ . Другими словами, медиана – это квантиль уровня  $p = 0,5$ .

**Многоэтапная выборка** – комбинирование выборочных методов

**Наблюдаемый паттерн** состоит из данных, применяемых для исследования теоретической модели

**Надежность** означает «допускающий возможность повторения» или «согласованный». Измерение считается надежным, если оно снова и снова приводит к одному и тому же результату

**Надежность внешних наблюдений** используется при оценке степени, в которой согласуются между собой оценки одного и того же явления различными наблюдателями

**Надежность параллельных форм** используется при оценке согласованности результатов двух тестов, сконструированных одинаковым образом исходя из одной и той же содержательной области

**Номинальное измерение** – количественное обозначение оказывается всего лишь уникальным «наименованием» свойства. Не предполагается какого-либо упорядочения рассматриваемых случаев

**Одновременная валидность** направлена на оценку способности нашей операционализации проводить различие между теми группами, между которыми она должна теоретически быть способна проводить различие

**Ответ испытуемого** – это конкретное измеряемое значение, получаемое нами от единицы (респондента), входящей в выборку

**Паттерн** – любое упорядочение объектов или единиц.

**Прогностическая валидность** направлена на оценивание способности операционализации прогнозировать то, что она теоретически может прогнозировать

**Пропорциональная квотная выборка** – отражает основные характеристики генеральной совокупности, благодаря тому, что каждая из них представлена в пропорциональном объеме

**Простая случайная выборка** – простейшая форма случайной выборки

**Процентили** – это величины, делящие выборку данных на сто групп, содержащих (по возможности) равное количество наблюдений. Например, 30% данных имеют значение, меньше 30-го перцентиля.

**Расколотая пополам надежность** – все пункты методики делят случайным образом с целью измерить один и тот же конструкт в каждой полученной группе

**Реверсивные шкалы (деления)** – в некоторых разновидностях шкалы Лайкерта, деления на шкале направлены в обратную сторону от общего направления шкалы. Оценочные значения по делениям таких шкал следует обращать, прежде чем складывать общую сумму. Это означает, что, если испытуемый выбрал оценку 1, то 1 следует считать 5, если испытуемый выбрал оценку 2, то 2 следует считать 4,  $3 = 3$ ,  $4 = 2$ , а  $5 = 1$ .

**Робастностью** статистического метода называется его свойство сохранять работоспособность за пределами тех предположений, для которых он выведен.

**Систематическая ошибка** вызывается любыми факторами, которые систематическим образом воздействуют на измерение переменной в выборке. Систематические ошибки называются **смещением** измерения

**Случайная ошибка** вызывается любыми факторами, которые случайным образом воздействуют на измерение переменной в выборке. Случайная ошибка иногда считается **шумом**

**Содержательная валидность** направлена на проверку операционализации по отношению к релевантной содержательной области конструкта

**Стандартное отклонение** – это распределение значений вокруг среднего в единственной выборке.

**Стандартная ошибка** – это распределение средних значений вокруг среднего от средних в выборочном распределении.

**Статистика** – это функция элементов выборки. Существует большое многообразие статистик, которыми мы можем воспользоваться – среднее, медиана, мода и т. п.

**Теоретический паттерн** – это гипотеза о том, чего следует ждать от полученных данных

**Теория истинного значения** – это теория измерения. Теория истинного значения выступает основой теории надежности

**Тест-ретестовая надежность** используется при оценке согласованности измерения в различные моменты времени

**Уровень измерения** соотносится с взаимосвязью между теми величинами, которые приписываются свойствам переменной

**Шкалирование** – область измерения, связанная с конструированием измерительного инструментария, которая сопрягает качественные конструкты с количественными метрическими единицами

# МОДУЛЬ 1

## ТЕМА 1. ТЕРМИНОЛОГИЯ СТАТИСТИЧЕСКОЙ ВЫБОРКИ

Начнем с определения некоторых весьма простых терминов, имеющих отношение к статистической выборке. Когда проводится выборочное исследование, то единицы, из которых состоит выборка – обычно это люди – дают нам один или несколько ответов. В этом смысле, **ответ** испытуемого – это конкретное измеряемое значение, получаемое нами от единицы (респондента), входящей в выборку. Например, респондент, заполняя опросный инструментарий, выбрал вариант ответа «4». Когда мы рассматриваем ответы по всей выборке, мы прибегаем к **статистике**. Существует большое многообразие статистик, которыми мы можем воспользоваться – среднее, медиана, мода и т. п. Допустим, что среднее значение в выборке составляет 3,72. Причина, по которой мы обращаемся к выборке, заключается в том, что мы стремимся оценить популяцию, из которой взята выборка. Если бы это было в наших силах, то мы бы предпочли оценивать непосредственно популяцию. Если мы измеряем всю популяцию целиком и рассчитываем среднее значение, то оно уже называется не статистикой, а **параметром** популяции.

<b>Переменная</b>	ответ (реакция)	1 2 3 ④ 5
<b>Статистика</b>	выборка	среднее = 3,75
<b>Параметр</b>	популяция	среднее = 3,72

### Выборочное распределение

Каким образом при помощи выборочной статистики оценивается параметр популяции? Для того, чтобы это понять, нам потребуется понятие о выборочном **распределении**. Проведем мысленный эксперимент. Представим себе, что, вместо того, чтобы использовать единую выборку, как это обычно делается в типовых исследованиях, мы взяли три выборки из одной и той же популяции. Более того, представим себе, что по всем трем выборкам нами получены единые ответы, обобщив которые, мы получили единую статистику, скажем, средний ответ.

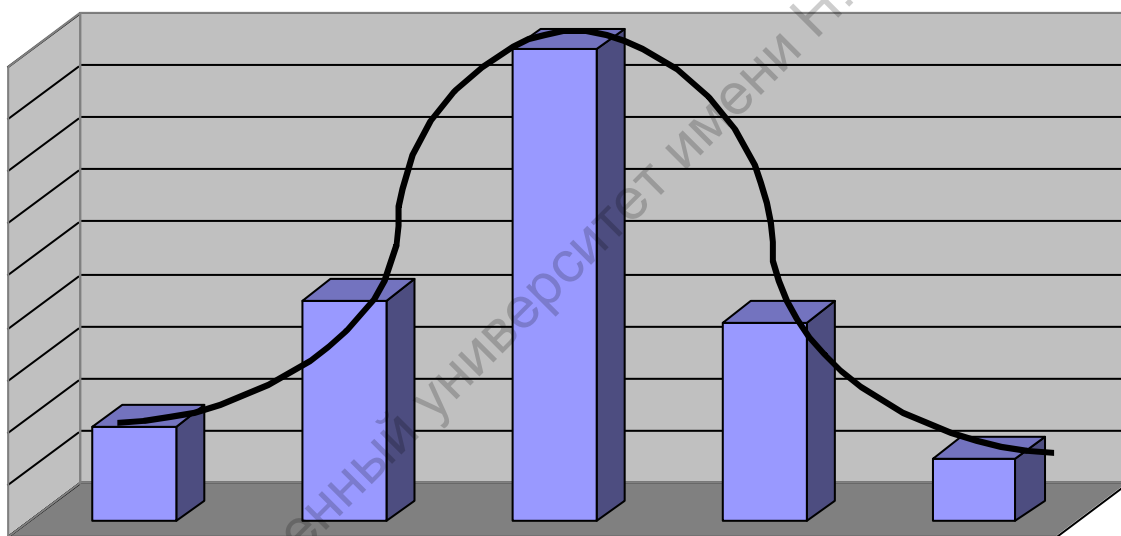
Даже при том, что все три выборки извлечены из одной и той же популяции, не следует ожидать, что получится одна и та же статистика для каждой выборки. Они будут слегка различаться либо по причине произвольности их составления, либо в силу естественных флуктуаций, либо в силу причудливости самих выборок. Однако, следует рассчитывать на то, что во всех трех выборках окажется сходная статистическая оценка, поскольку выборки извлечены из одной популяции.



Теперь представим себе, что мы располагаем *бесконечным* количеством выборок из одной и той же популяции, и что мы вычислили среднее значение для каждой выборки. Если изобразить их в виде гистограммы, то окажется, что в основной массе они приближаются к одному и тому же центральному значению, а таких выборок, в которых среднее значение заметно удалено от центрального значения, весьма мало. Иными словами, гистограмму вполне можно представить в виде *колоколообразной* кривой, которая в статистике служит показателем «нормального» распределения. Распределение бесконечного количества выборок одного и того же объема называют **выборочным распределением**.

## Выборочное распределение..

Гистограмма



## ...это распределение статистики по бесконечному множеству выборок

Рис. 1. Гистограмма выборочного распределения

Выборочное распределение рассматривают в редких случаях. Почему? Вы невнимательны! Потому что в этом случае нам бы потребовалось *бесконечное* количество выборок, а на нашей планете неизвестно ничего бесконечного. Тогда к чему мы вообще заговорили о выборочном распределении? Хороший вопрос! Потому что нам требуется понять, что наша выборка – всего лишь одна из потенциально бесконечного множества выборок, которые мы могли бы рассмотреть. Когда мы знаем о том, что такое выборочное распределение, то мы понимаем, что, несмотря на то, что полученная нами статистика, вероятно, приближается к центру выборочного распределения (ибо

большинство выборок к нему приближаются), однако, могло получиться так, что у нас оказалась какая-то крайняя выборка – просто нам (не) повезло.

Если взять среднее значение выборочного распределения – т. е. среднее значение от средних значений бесконечного количества выборок – то мы намного приблизимся к истинному значению популяции – т. е. к интересующему нас параметру. Итак, среднее выборочного распределения в действительности эквивалентно параметру. Тогда что такое стандартное отклонение выборочного распределения? Стандартное отклонение выборочного распределения показывает нам нечто о том, каким образом распределены различные выборки. В статистике это называется **стандартной ошибкой**, и нам следует отличать её от стандартного отклонения. Чтобы не создавать путаницу, повторим ещё раз:

- ❖ **стандартное отклонение** – это распределение значений вокруг среднего в единственной выборке;
- ❖ **стандартная ошибка** – это распределение средних значений вокруг среднего от средних в выборочном распределении.

Уловили?

## Выборочная ошибка

В выборочном контексте, стандартная ошибка называется **выборочной ошибкой**. Выборочная ошибка дает нам некоторое представление о точности нашей статистической оценки. Низкая выборочная ошибка означает, что получили относительно меньше вариабельности или размаха в выборочном распределении. Но опять-таки – в действительности мы никогда не видим выборочного распределения! Тогда как же нам вычислить выборочную ошибку? Наши расчеты основаны на *стандартном отклонении нашей выборки*. Чем больше выборочное стандартное отклонение, тем больше стандартная ошибка (и выборочная ошибка). Стандартная ошибка также соотносится с объемом выборки. Чем больше объем выборки, тем меньше стандартная ошибка. Почему? Потому что чем больше объем выборки, тем ближе выборка к реальной популяции. Если взять выборку, состоящую из целой популяции, то фактически не получим выборочной ошибки, потому что у нас нет выборки, а есть целая популяция. В этом случае значение среднего будет являться параметром.

## Правило 65%, 95% и 99%

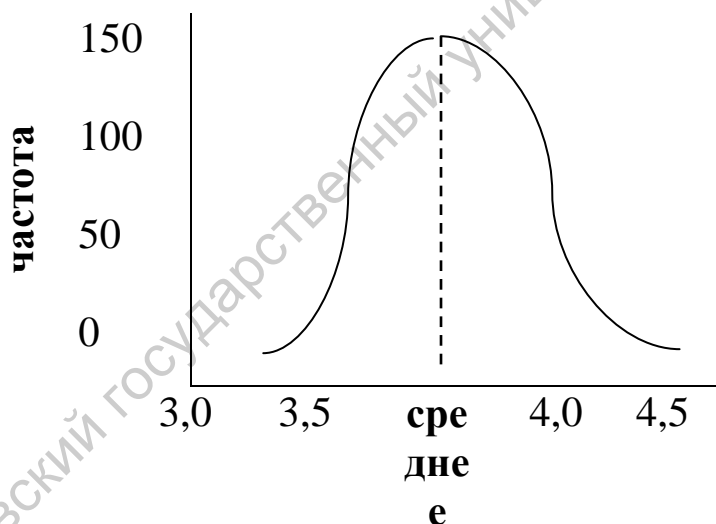
Возможно, вы уже слышали об этом прежде, но это настолько важно, что никогда не будет лишним это повторить... Существует общее правило, применимое к тем случаям, когда у нас имеется нормальное или колоколообразное распределение. Начнем со среднего – с центра распределения. Если отклониться от среднего вверх и вниз (т. е. влево и вправо) на одно стандартное значение, тогда мы охватим 65% всех случаев данного распределения (т. е. 65% площади под кривой). Если отклониться вверх и вниз на два стандартных значения, то мы охватим приблизительно 95% всех случаев. А если до-

пустить плюс-минус три стандартных величины, то мы охватим 99% всех случаев. Обратите внимание на то, что я самого начала не уточнил, о каком стандартном значении я веду речь – о стандартном отклонении либо о стандартной ошибке. Я не уточнил, потому что одно и то же правило верно для обоих типов распределения (т. е. и для распределения сырых данных, и для выборочного распределения).

около 99% случаев оказываются между 3,0 и 4,5

около 95% случаев оказываются  
между  
3,25 и 4,25

около 65% случаев ока-  
зываются между  
3,5 и 4,0



Выборочное распределение характеризуется средним, равным 3,75, и стандартным отклонением, равным 0,25

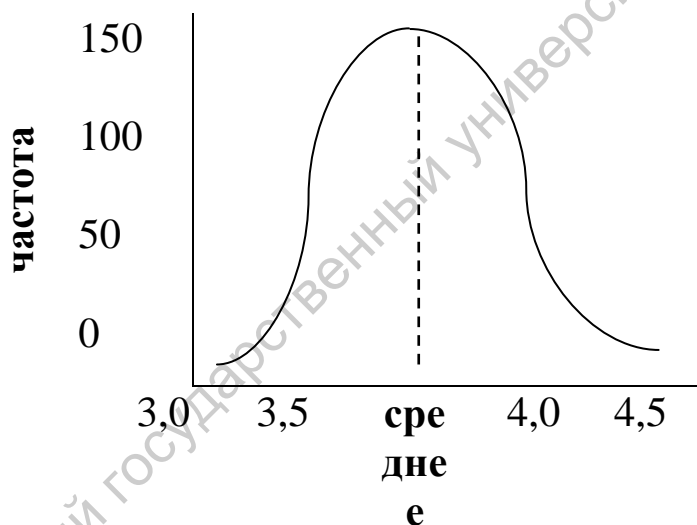
Например, допустим, то среднее значение для распределения равно 3,75, а стандартное значение равно 0,25. (Если бы это было распределение сырых данных, то мы бы брали величину стандартного отклонения. Если бы это было выборочное распределение, то мы бы брали величину стандартной ошибки). При отклонении вверх и вниз от среднего значения на одно стандартное значение, мы переместимся от среднего, равного 3,75 вверх на 0,25 и вниз на 0,25. Внутри этих границ – от 3,5 до 4,0 – мы встретим приблизительно 65% случаев. Вы можете спросить: что это значит? Если мы работаем с сырыми данными, и нам известны величин среднего значения и стандарт-

ного отклонения, то мы можем *прогнозировать* интервалы, в которых окажутся 65%, 95% и 99% рассматриваемых случаев. Эти интервалы называются – угадайте, как? – интервал 65%-ной доверительности, интервал 95%-ной доверительности и интервал 99%-ной доверительности.

около 99% случаев оказываются между 3,675 и 3,825

около 95% случаев оказываются между 3,70 и 3,80

около 65% случаев оказываются между 3,725 и 3,775



Выборочное распределение характеризуется средним, равным 3,75, и стандартной ошибкой, равной 0,025

Вот теперь все становится понятно! Если нам известно *выборочное распределение*, то мы в состоянии прогнозировать 65%-ный, 95%-ный и 99%-ный доверительные интервалы того, какими должны быть параметры популяции! А разве не для этих целей мы пользуемся выборкой? Чтобы быть в состоянии прогнозировать, какой окажется популяция по данной переменной? Всего один момент. В действительности, мы не располагаем выборочным распределением (говорю об этом уже в третий раз)! Однако, мы располагаем распределением для самой выборки. И мы можем на основе этого распределения оценить стандартную ошибку (и выборочную ошибку), потому что ошибка основана на стандартном отклонении, а оно нам известно. Ра-

зумеется, мы не знаем действительного значения параметра популяции – мы пытаемся его определить – однако, мы можем приложить к этому усилия и найти статистику выборки. Теперь, если мы знаем среднее значения выборочного распределения (либо принимаем его равным среднему значению нашей выборки), и нам известна оценка стандартной ошибки (мы рассчитали её по нашей выборке), то у нас есть два главных ингредиента, необходимых для нашего выборочного распределения, для того, чтобы определить доверительные интервалы параметра популяции.

Поясним на примере. Предположим, что мы проводим исследование, и что мы составили выборку из популяции. Предположим далее, что среднее по выборке равно 3,75, а стандартное отклонение равно 0,25. Это распределение сырых данных, представленное ранее. Спрашивается, каково в данном случае выборочное распределение? Мы не можем сконструировать выборочное распределение, но мы в состоянии его оценить. Для начала, предположим, что среднее выборочного распределения соответствует среднему в нашей выборке и равно 3,75. Далее, рассчитаем стандартную ошибку. Для этого, нам потребуются величина стандартного отклонения в нашей выборке и объем выборки (допустим, что  $N = 100$ ). Допустим, мы получаем стандартную ошибку, равную 0,025 (поверьте мне). Теперь у нас есть все, что нам нужно, для того, чтобы оценить доверительный интервал параметра популяции. По нашей оценке получается, что с вероятностью 65% истинное значение параметра окажется в интервале между 3,725 и 3,775 (т. е., 3,75 плюс/минус 0,025); что 95%-ный доверительный интервал имеет границы от 3,7 до 3,8; и с 99%-ной уверенностью можно сказать, что значение параметра популяции окажется в интервале от 3,675 до 3,825. Истинное значение в нашем придуманном примере составляет 3,72, значит, мы правильно рассчитали это значение для нашей выборки.

## ВЕРОЯТНОСТНАЯ ВЫБОРКА

Вероятностным выборочным методом является любой метод составления выборки, в котором применяется какая-то форма *случайного отбора*. Для того, чтобы метод составления выборки являлся случайным, необходимо обратиться к такому процессу либо процедуре, которые обеспечивают различным единицам популяции равные вероятности быть выбранными. Человечество с давних пор применяет различные формы случайного отбора, например, вытягивание жребия из шапки, или вытягивание короткой соломинки. В наши дни в качестве механизма генерирования случайных чисел, на основе которых производится случайный выбор, применяют компьютеры.

### Некоторые определения

Прежде, чем приступить к объяснению различных вероятностных методов, сформулируем некоторые основные понятия:

- ❖  $N$  = количество случаев, из которых формируется выборка
- ❖  $n$  = количество случаев в выборке
- ❖  ${}_N C_n$  = количество комбинаций  $n$  из  $N$

❖  $f = n/N$  = выборочная фракция

Итак, определив эти основные понятия, можно приступить к формулированию различных вероятностных выборочных методов.

### Простая случайная выборка

Простейшая форма случайной выборки называется **простая случайная выборка**. Характеристики простой случайной выборки:

❖ **Цель:** выбрать  $n$  единиц из  $N$  таким образом, чтобы каждая величина  $nC_n$  обладала равными шансами быть выбранной.

❖ **Процедура:** для формирования выборки применяется таблица случайных чисел, компьютерный генератор случайных чисел или механическое устройство.

Как же составляется простая случайная выборка? Предположим, что мы проводим какое-то исследование в туристическом агентстве и стремимся оценить представления клиентов о качестве сервисных услуг за прошлый год.

Сначала следует организовать выборочную массу. Для этого просматривают записи компании и выявляют всех клиентов за последние 12 месяцев. Если нам повезет, то в компьютере компании мы найдем нужную информацию и быстро оставим список клиентов.

Далее, переходим к составлению собственно выборки. Решаем, сколько человек мы включим в окончательную выборку. Для примера предположим, что нам нужно отобрать 100 человек из списка в 1000 клиентов. Тогда выборочная фракция  $f = n/N = 100/1000 = 0,1$  или 10%.

Теперь, есть несколько путей для составления выборки:

- 1) можно напечатать список 1000 клиентов на принтере, разрезать его на полоски, сложить полоски в шляпу, хорошенько их перемешать, закрыть глаза и вытянуть первые 100 полосок. Эта механическая процедура утомительна, качество выборки будет зависеть от того, насколько хорошо перемешали и насколько случайно вытягивали.
- 2) можно использовать машину, которая перемешивает шарики в лотереях. Потребуется три комплекта шариков с номерами от 0 до 9, чтобы получать трехзначные числа от 000 до 999 (номер 000 будет считаться номером 1000). Пронумеруем имена в списке от 1 до 1000, а далее при помощи машины с шариками выберем трехзначные номера. Неудобство данного метода заключается в том, что необходима машина с шариками.
- 3) Третий путь заключается в применении компьютера. Эта простая процедура особенно удобна в том случае, если в компьютере уже имеются имена клиентов. Многие компьютерные программы способны генерировать серии случайных чисел. Допустим, мы скопировали список клиентов и вставили его в столбец в программу EXCEL. Далее, в следующий столбец справа от списка вставляем функцию =СЛЧИС(), при помощи которой в программе EXCEL

ячейкам присваиваются случайные между 0 и 1. Далее, сортируем оба столбика – и список имен, и список случайных чисел – при помощи случайных чисел. Список перестраивается в случайном порядке от наименьшего до наибольшего случайного числа. Теперь остается только взять первые сто имен из отсортированного списка. Очень просто. На всю процедуру уходит минута.

Простую случайную выборку легко составить и легко объяснить другим. Поскольку простая случайная выборка – это прямой способ составления выборки, то целесообразно обобщить результаты такой выборки обратно применительно к популяции. Простая случайная выборка – не самый эффективный метод формирования выборки, в силу случайного характера она может оказаться не репрезентативной в отношении подгрупп популяции. Для решения этих проблем придется обратиться к другим методам составления выборки.

### Стратифицированная случайная выборка

Стратифицированную случайную выборку также называют пропорциональной или квотной случайной выборкой. Она предполагает деление популяции на гомогенные подгруппы, а затем составление простых случайных выборок для каждой подгруппы. Говоря более формально:

❖ **Цель:** Разделить популяцию на непересекающиеся подгруппы (т. е. на страты)  $N_1, N_2, \dots, N_i$ , таким образом, чтобы  $N_1 + N_2 + \dots + N_i = N$ . Затем для каждой страты составляется простая случайная выборка из  $f = n/N$ .

Существуют несколько важных причин, по которым стратифицированная выборка более предпочтительна перед простой случайной выборкой.

- 1) Стратифицированная выборка обеспечивает репрезентативность не только популяции в целом, но также важнейших подгрупп популяции, особенно небольших групп меньшинств. Если мы хотим иметь возможность рассматривать подгруппы, то формирование стратифицированной выборки – это единственный способ эффективного их рассмотрения. Если подгруппа крайне мала, то могут применяться различные выборочные фракции ( $f$ ) внутри отдельных страт с целью формирования случайной выборки малой группы (однако, для последующей оценки целой популяции придется рассчитывать веса внутригрупповых значений с применением выборочных фракций). Когда внутри страт у нас имеются одинаковые выборочные фракции, то мы проводим *пропорциональное* стратифицированное случайное выборочное исследование. Когда внутри страт у нас имеются различные выборочные фракции, то мы проводим *непропорциональное* стратифицированное случайное выборочное исследование.
- 2) Стратифицированная случайная выборка обычно обладает более высокой статистической точностью, по сравнению с простой слу-

чайной выборкой. Но только том случае, когда страты или групп гомогенные. Если они гомогенные, то следует рассчитывать на то, что вариабельность внутри групп окажется ниже, чем вариабельность в популяции в целом.

К примеру, допустим, популяцию клиентов нашего турагентства можно разделить на три группы: студенты, пожарные и футболисты. Предположим также, что пожарных и футболистов сравнительно мало (соответственно 10% и 5%). Если бы мы составляли простую случайную выборку  $n = 100$  при выборочной фракции 10%, то мы бы совершенно случайным образом получили бы 10 пожарников и 5 футболистов. Но в силу случайности выбора, мы бы получили их еще меньше!

Во-первых, следует решить – сколько человек нам нужно в каждой группе. Допустим, мы по-прежнему стремимся оставить выборку из 100 человек, отобранных из популяции в 1000 клиентов. Однако, для того, чтобы мы оказались в состоянии что-либо сказать по поводу наших подгрупп, каждая подгруппа должна включать не менее 25 случаев. Итак, пусть будут 50 студентов, 25 пожарных и 25 футболистов. Нам известно, что 10% популяции, или 100 клиентов, были пожарными. Если мы составим из них случайную выборку в количестве 25 респондентов, тогда внутрискратовая выборочная фракция окажется  $25/100 = 25\%$ . Аналогичным образом, нам известно, что 5% или 50 клиентов были футболистами. Итак, наша внутрискратовая выборочная фракция составляет  $25/50 = 50\%$ . Наконец, путем вычитания мы узнаем, что студентов было 850. Внутрискратовая выборочная фракция студентов составляет  $50/850 =$  примерно 5,88%. Поскольку внутрискратовая гомогенность превышает гомогенность внутри популяции, то следует ожидать большой статистической точности (меньшей дисперсии). А поскольку мы произвели стратифицирование, то мы получаем достаточное количество случаев в каждой группе для того, чтобы прийти к значимым выводам в отношении подгрупп.

### **Систематическая случайная выборка**

Для составления систематической случайной выборки следует выполнить следующий алгоритм:

- ❖ количество единиц в популяции от 1 до  $N$
- ❖ определить  $n$  (объем выборки), который желателен или необходим
- ❖  $k = N/n =$  размер интервала
- ❖ произвольно выбрать единицу измерения (целое число) от 1 до  $k$
- ❖ затем взять каждый  $k$ -тый член

Всё это намного понятнее на примере. Допустим, что у нас есть популяция, в которой всего  $N = 100$  человек, а нам требуется сформировать выборку  $n = 20$ . Для того, чтобы получить систематическую выборку, популяцию следует расположить в случайном порядке. Выборочная фракция составит  $f = 20/100 = 20\%$ . В данном случае размер интервала  $k$  равен  $N/n = 100/20 = 5$ . Теперь выберем произвольное целое число от 1 до 5. Допустим, мы вы-



брали 4. теперь, чтобы сформировать выборку, начнем с четвертого члена в списке, и отберем каждый k-тый член (каждый пятый, потому что  $k = 5$ ). В выборку попадают номера 4, 9, 14, 19 и т. д. до 100, а всего в выборке окажутся 20 членов.

Для того, чтоб это получалось, важно, чтобы члены популяции имели случайное распределение, по крайней мере, в отношении тех характеристик, которые измеряются.

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

**N = 100**

**желательно n = 20**

**N/n = 5**

**выбирается произвольное число  
от 1 до 5:  
допустим, 4**

**начинаем с номера 4 и берем каж-  
дый пятый член**

По какой причине применяется систематическая случайная выборка? По одной – её весьма просто сформировать. Всё, что нужно для формирования подобной выборки – выбрать одно случайное число, с которого начинается выборка. Кроме того, такая выборка может оказаться более точной, нежели простая случайная выборка. Наконец, существуют ситуации, в которых нет более простого способа сформировать случайную выборку. К примеру,

нам требуется провести исследование, предполагающее формирование выборки из книг, собранных в библиотеке. Мы подходим к полке, выбираем книгу, и отмечаем, когда её брали читатели в последний раз. Нас имеется хороший массив для формирования выборки в виде каталога из карточек. Для того, чтобы составить простую случайную выборку, мы определяем общее количество книг и генерируем случайные номера книг, которые составят на выборку. Но если мы выбрали книгу номер 74 329, легко ли нам будет её найти? Неужели придется пересчитывать все карточки в каталоге, пока не дойдем до номера 74 329? Стратифицирование также не решит нашу проблему в данном случае. Например, нельзя же стратифицировать весь каталог на отдельные ящики, и из каждого ящика вытянуть по одной карточке. Но мы упорно продолжаем считать все карточки в каталоге и в конце концов мы составили систематическую случайную выборку. Мы поступили следующим образом. Определили, что всего в библиотеке 100 000 книг. Для того, чтобы оставить выборку из 1000 книг, мы определяем, что выборочная фракция  $1000/100000 = 1\%$ . Для того, чтобы получить выборочный интервал  $k$ , мы делим  $N/n = 100000/100 = 1000$ . далее выбираем случайное целое число между 1 и 1000. Допустим, это число 257. Далее, определяем толщину стопки карточек в каталоге, допустим, что в среднем толщина стопки из 1000 карточек 6,74 см. Теперь мы знаем все, что нам нужно, чтобы сформировать выборку. Мы отсчитываем вручную 257-ую карточку и выписываем информацию о выбранной книге. Далее, берем циркуль, раскрываем его на ширину 6,74 см, устанавливаем на 257-ую карточку, а другим концом находим следующую карточку. Таким образом, мы отбираем 257-ую, 1257-ую, 2257-ую книгу и т. д. За короткое время мы отберем все нужные нам для выборки книги и формируем систематическую случайную выборку. Любой другой метод составления случайной выборки потребует значительно больше времени.

### **Кластерная (территориальная) случайная выборка**

Проблема, связанная со случайным выборочным методом, когда нам приходится формировать выборку на основе популяции, распределенной по широкому географическому региону, заключается именно в географической удаленности от компонентов, образующих популяцию. Допустим, вы составляете простую случайную выборку из всех жителей Саратовской области, для того, чтобы провести личные собеседования. В выборку войдут люди, живущие в разных районах области. Интервьюерам придется много разъезжать. Именно для этих целей была придумана **кластерная или территориальная случайная выборка**.

При составлении кластерной выборки выполняется следующая последовательность:

- ❖ популяцию делят на кластеры
- ❖ из кластеров составляют случайные выборки
- ❖ обследуют все компоненты выбранных кластеров

Допустим, мы проводим исследование по муниципальным образованиям, и для этого нам потребуется лично приезжать в муниципальные образования. Если составлять простую случайную выборку для области, то потребуется географически охватить всю область. Вместо этого, мы составляем кластерную выборку для пяти муниципальных образований. После этого мы посещаем каждое из пяти муниципальных образований. Ясно, что подобная стратегия позволяет уменьшить разъезды, кроме того, данный метод применяется при почтовых и телефонных опросах.

### **Многоэтапная выборка**

Четыре метода, которые мы уже рассмотрели – простой, стратифицированный, систематический и кластерный – это простейшие стратегии составления случайных выборок. В большинстве психологических исследований, применяются более сложные методы формирования выборки. Важнейший принцип состоит в том, что имеется возможность комбинировать описанные ранее простейшие методы множеством полезных способов, что позволяет сделать исследование наиболее результативным. Комбинирование выборочных методов называется **многоэтапной выборкой**.

Например, нам необходимо сформировать выборку по Саратовской области для проведения индивидуального психологического тестирования. На первом этапе процесса мы составим кластерную выборку. В выборку можно включить муниципальные образования. Но в случае кластерной выборки, нам придется тестировать каждого человека в составленном кластере. Поэтому внутри кластеров составляются стратифицированные случайные выборки. В этом случае мы имеем дело с двухэтапным процессом формирования выборки: внутри кластерных выборок имеются стратифицированные выборки. Либо, обратимся к проблеме формирования выборки из учащихся старших классов средних школ. Сначала возьмем список школ области, и составим из них простую случайную выборку. Внутри школ составим простую случайную выборку классов. А внутри классов составим простую случайную выборку учеников. Процесс составления выборки осуществляется в три этапа.

### **НЕВЕРОЯТНОСТНАЯ ВЫБОРКА**

Различие между вероятностной выборкой и невероятностной выборкой заключается в том, что для невероятностной выборки не требуется случайный отбор, а для вероятностной выборки случайный отбор необходим. Означает ли это, что невероятностные выборки не репрезентативны? Вовсе нет. Однако, это означает, что невероятностные выборки не описываются положениями теории вероятности. В отношении вероятностных выборок нам известны шансы на то, что популяция окажется представлена репрезентативно. Мы можем рассчитать интервалы вероятности для статистик. В отношении невероятностных выборок, мы можем отразить популяцию репрезентативно, а можем и нерепрезентативно, и каждый раз нам трудно будет узнавать, на-

сколько наша выборка репрезентативна. В общем, исследователи предпочитают вероятностные или случайные выборочные методы перед невероятностными, считая их более точными и надежными. Тем не менее, в прикладных психологических исследованиях возникают такие обстоятельства, когда случайная выборка неосуществима на практике или не имеет смысла теоретически. Далее мы рассмотрим широкий спектр невероятностных альтернатив.

Невероятностные выборочные методы могут быть разделены на два широких типа: *произвольные* и *преднамеренные*. Большинство выборочных методов имеют преднамеренную природу, потому что при составлении выборки у нас обычно имеется конкретный план. Важнейшее различие между этими типами выборочных методов заключается в различных типах преднамеренности при формировании выборок.

### **Произвольные выборки**

Один из широко распространенных методов формирования выборки указан в названии этого раздела. Сюда включается традиционное интервью с «человеком на улице», которое часто проводят новостные программы телевидения с целью быстрого (но нерепрезентативного) изучения общественного мнения. Сюда относится использование студентов в психологических исследованиях (нельзя ведь утверждать, что студенты репрезентативным образом представляют всю популяцию). В клинической практике в этой роли выступают пациенты. Во многих случаях выборку составляют из тех, кто изъявил добровольное согласие участвовать в исследовании. Ясно, что проблема, связанная с выборками данного типа, заключается в том, то у нас нет подтверждений тому, что данная выборка репрезентативна в отношении той популяции, на которую мы хотели бы распространить результаты нашего исследования.

### **Преднамеренные (целенаправленные) выборки**

В случае целенаправленных выборок, у нас есть определенная цель. Обычно нам необходимы одна или несколько групп с определенными характеристиками. Например, возле универмага может стоять человек с плакатом, который обращается к различными людям и просит разрешения их интервьюировать. Обычно так проводятся маркетинговые исследования. Допустим, необходимо провести интервью с женщиной 30 – 40 лет. Тогда высматривают всех, кто проходит мимо, и выбирают ту, которая внешне соответствует требуемым критериям. Интервьюер начинает с того, что проверяет, соответствует ли в действительности выбранный кандидат предъявляемым критериям. Целенаправленная выборка очень полезна в тех случаях, когда вам необходимо быстро получить необходимый контингент респондентов, а вопрос пропорционального состава выборки не очень важен. При опросе целенаправленной выборки высока вероятность узнать мнение именно интересующей нас популяции, однако, доминировать в этом мнении будут те подгруппы, с которыми легче наладить общение.

Все методы, о которых будет сказано далее, могут рассматриваться в качестве подкатегорий целенаправленных выборочных методов. Можно составлять выборку из конкретных групп или типов людей, как в случае модальной выборки, экспертной выборки или квотной выборки. Выборка может быть неоднородной – такая выборка называется гетерогенная выборка. Выборка может быть основана на неформальных социальных сетях, что позволяет выявлять конкретных респондентов, которые недоступны другими выборочными методами – такая выборка называется «снежный ком». В каждом из этих случаев мы знаем, чего мы хотим – мы составляем выборку с определенной целью.

### **Модальная выборка**

В статистике модой называется наиболее часто встречающееся значение выборки. При составлении модальной выборки, в выборку включается наиболее типичный случай. Например, в большинстве неформальных опросов общественного мнения опрашивают «среднего» избирателя. Данный подход чреват несколькими проблемами.

Во-первых, как мы узнаем, каков он – типичный или модальный случай? Можно сказать, что типичный избиратель – это человек среднего возраста, среднего образования, средних доходов с точки зрения популяции. Однако, средние показатели могут оказаться отнюдь не средними (к примеру, возможно смещенное распределение доходов). И почему мы решили, что эти три переменные – возраст, образование и доход – единственные или наиболее релевантные характеристики типичного избирателя? А что если национальность или религия играют более существенную роль? Ясно, что модальная выборка имеет смысл лишь в неформальном выборочном контексте.

### **Экспертная выборка**

Экспертная выборка предполагает составление выборки из тех людей, о которых известно, что они обладают специальными познаниями. Такие исследования еще называют панельными. Есть, по крайней мере, две причины для предпочтения экспертной выборки.

Во-первых, это наилучший способ узнать мнение людей, обладающих специальными познаниями. С этой точки зрения экспертная выборка оказывается не более, чем разновидностью целенаправленной выборки.

Во-вторых, экспертная выборка позволяет обосновать валидность других подходов к составлению выборки. К примеру, допустим, мы составляем модальную выборку и определяем критерии для определения модального (типичного) случая. Тогда собирают панель экспертов, имеющих опыт в данной проблеме и просят их прокомментировать адекватность и валидность тех критериев, которые вы сформулировали самостоятельно. Недостаток экспертного метода состоит в том, что эксперты, увы, могут ошибаться.

## Квотная выборка

При составлении квотной выборки, люди отбираются неслучайным образом в соответствии с установленными квотами. Существуют два типа квотных выборок: *пропорциональные* и *непропорциональные*. В **пропорциональной квотной выборке** отражаются основные характеристики популяции, благодаря тому, что каждая из них представлена в пропорциональном объеме. К примеру, допустим, известно, что популяция состоит на 60% из женщин и на 40% из мужчин, а нам нужно составить выборку из 100 респондентов, следовательно, в выборе должно быть такое же процентное соотношение женщин и мужчин. Проблема в связи с квотной выборкой (так же как и в связи с большинством целенаправленных выборок) состоит в необходимости сформулировать специфические характеристики, на которых будет основана квота: пол, возраст, образование, национальность, религия и т. п.

**Непропорциональная квотная выборка** менее строга. Определяется минимальное количество компонентов выборки по каждой категории, но при этом не стремятся к пропорциональному отражению популяции в выборке. Вместо пропорционального отражения популяции, стремятся по возможности отразить все возможные группы, существующие в данной популяции. Данный метод представляет собой невероятный аналог стратифицированной случайной выборки, и в общем случае применяется с целью обеспечить адекватное представительство в выборке самых маленьких групп популяции.

## Гетерогенная выборка

Гетерогенная выборка составляется в тех случаях, когда стремятся включить в выборку все возможные мнения и взгляды, а пропорциональность представления этих взглядов для нас не очень важна. Отличительной особенностью такого рода выборок является *разнообразие*. При проведении мозговых штурмов применяется аналог гетерогенных выборок, потому что главной целью в этих случаях оказывается получение широкого спектра идей, а не установление «среднего» или «типичного (модального) случая». Фактически, компонентами выборки оказываются не люди, а идеи. Мы представляем себе, что существует универсум всех возможных идей, релевантных некоторой теме, и стремимся составить выборку из идей, а не из людей. Понятно, что для того, чтобы представить все идеи, а особенно идеи необычные, необходимо собрать широкий контингент участников. В этом смысле,

## Вопросы для самоконтроля

### по теме 1 «Терминология статистической выборки»:

1. Сформулируйте определения понятий «ответ», «статистика», «параметр».
2. Что называется «выборочным распределением»?
3. Почему выборочное распределение рассматривают в редких случаях?

4. Объясните различие между стандартной ошибкой и стандартным отклонением.
5. Почему чем больше объем выборки, тем меньше стандартная ошибка?
6. Когда значение среднего будет являться параметром?
7. Сформулируйте правило 65%, 95% и 99%.
8. Как называются интервалы 65%, 95% и 99%?
9. Что такое вероятностный выборочный метод? Для чего он применяется?
10. Перечислите характеристики простой случайной выборки.
11. Как составляется простая случайная выборка?
12. По каким причинам стратифицированная выборка более предпочтительна, чем простая случайная выборка?
13. Каков алгоритм составления систематической случайной выборки?
14. Как называется выборка, составленная на основе популяции, распределенной по широкому географическому региону?
15. Какова последовательность составления кластерной выборки?
16. Как называется комбинирование выборочных методов?
17. В чем заключается различие между вероятностной выборкой и невероятностной выборкой? Означает ли это различие, что невероятностные выборки не репрезентативны?
18. Почему исследователи предпочитают вероятностные или случайные выборочные методы перед невероятностными?
19. Приведите примеры произвольных выборок.
20. В чем состоит особенность целенаправленных выборок?
21. Как называется неоднородная выборка?
22. Что включают в модальную выборку?
23. Кто участвует в панельных исследованиях?
24. В чем состоят преимущества экспертной выборки?
25. Какие типы квотных выборок вы знаете? В чем состоит различие между ними?
26. В чем заключается отличительная особенность гетерогенных выборок?
27. Как составляется выборка «снежный ком»?

## Тестовые задания к теме 1

1. Конкретное измеряемое значение, получаемое нами от единицы (респондента), входящей в выборку, называется:
  - а) характеристика
  - б) ответ
  - в) статистика

- г) параметр  
д) габарит
2. Причина, по которой мы обращаемся к выборке, заключается в том, что мы стремимся оценить
- а) испытуемого
  - б) выборку
  - в) результат
  - г) генеральную совокупность
3. Если мы измеряем всю генеральную совокупность и рассчитываем среднее значение, то оно называется
- а) характеристика
  - б) ответ
  - в) статистика
  - г) параметр
  - д) габарит
4. Если сформировать несколько выборок из одной и той же популяции, то во всех выборках статистическая оценка окажется
- а) различной
  - б) сходной
  - в) близкой к нулю
  - г) равной нулю
  - д) не поддающейся расчетам
5. Если среднее значение для каждой выборки изобразить их в виде гистограммы, то окажется, что в основной массе
- а) они равны нулю
  - б) они взаимно удалены
  - в) они приближаются к одному и тому же центральному значению
  - г) они кратны 3,14
  - д) они целые числа
6. Распределение бесконечного количества выборок одного и того же объема называют
- а) матрица
  - б) частное
  - в) многочлен
  - г) константа
  - д) выборочное распределение
7. Среднее значение от средних значений бесконечного количества выборок называется
- а) среднее значение суммы
  - б) среднее значение произведения
  - в) среднее значение выборочного распределения
  - г) среднее значение матрицы
  - д) среднее значение многочлена
8. Среднее выборочного распределения в действительности эквивалентно



- а) 3,14
  - б) нулю
  - в) сумме квадратов катетов
  - г) статистике
  - д) параметру
9. Распределение средних значений вокруг среднего от средних в выборочном распределении, называется
- а) выборочным средним
  - б) средним арифметическим
  - в) стандартным средним
  - г) стандартной ошибкой
  - д) стандартным отклонением
10. Чем больше выборочное стандартное отклонение, тем больше
- а) выборочное среднее
  - б) среднее арифметическое
  - в) стандартное среднее
  - г) стандартная ошибка
  - д) стандартное отклонение
11. Чем больше объем выборки, тем меньше
- а) выборочное среднее
  - б) среднее арифметическое
  - в) стандартное среднее
  - г) стандартная ошибка
  - д) стандартное отклонение
12. Чем больше объем выборки, тем ближе выборка
- а) к нулю
  - б) к выборочному среднему
  - в) к стандартной ошибке
  - г) к стандартному отклонению
  - д) к генеральной совокупности
13. Когда вместо выборки исследуется генеральная совокупность, то среднее значение является
- а) характеристикой
  - б) ответом
  - в) статистикой
  - г) параметром
  - д) габаритом
14. Если отклониться от среднего вверх и вниз (т. е. влево и вправо) на одно стандартное значение, тогда мы охватим
- а) 65% всех случаев данного распределения
  - б) 25% всех случаев данного распределения
  - в) 95% всех случаев данного распределения
  - г) 99% всех случаев данного распределения
  - д) 50% всех случаев данного распределения

15. Если отклониться от среднего вверх и вниз (т. е. влево и вправо) на три стандартных значения, тогда мы охватим
- а) 65% всех случаев данного распределения
  - б) 25% всех случаев данного распределения
  - в) 95% всех случаев данного распределения
  - г) 99% всех случаев данного распределения
  - д) 50% всех случаев данного распределения
16. Действительные значения параметров генеральной совокупности
- а) нам известны
  - б) нам неизвестны
  - в) приведены в справочнике
  - г) имеются в Интернете
  - д) затрудняюсь ответить
17. Если мы знаем средние значения выборочного распределения, и нам известна оценка стандартной ошибки, то мы можем определить
- а) ширину и высоту
  - б) угол наклона
  - в) стандартное отклонение
  - г) стандартную ошибку
  - д) доверительный интервал
18. Вероятностным выборочным методом является любой метод составления выборки, в котором применяется какая-то форма
- а) случайного отбора
  - б) естественного отбора
  - в) мысленного эксперимента
  - г) нормального распределения
  - д) алгебраической модели
19. Таблица случайных чисел, компьютерный генератор случайных чисел или механическое устройство используются при составлении
- а) простой случайной выборки
  - б) произвольной выборки
  - в) кластерной выборки
  - г) экспертной выборки
  - д) выборки «снежный ком»
20. Стратифицированную случайную выборку также называют
- а) кластерной выборкой
  - б) пропорциональной или квотной случайной выборкой
  - в) экспертной выборкой
  - г) простой случайной выборкой
  - д) выборкой «снежный ком»
21. Большая статистическая точность выборки означает, что
- а) дисперсия велика
  - б) дисперсия мала
  - в) дисперсия равна десятичному логарифму
  - г) дисперсия стремится к целому числу

- д) дисперсия стремится к бесконечности
22. Комбинирование выборочных методов называется
- а) кластерной выборкой
  - б) многоэтапной выборкой
  - в) экспертной выборкой
  - г) простой случайной выборкой
  - д) выборкой «снежный ком»
23. Различие между вероятностной выборкой и невероятностной выборкой заключается в том, что для вероятностной выборки необходим
- а) случайный отбор
  - б) естественный отбор
  - в) мысленный эксперимент
  - г) нормальное распределение
  - д) алгебраическая модель
24. Когда нам известны шансы на то, что генеральная совокупность окажется представлена репрезентативно, такая выборка называется
- а) кластерной выборкой
  - б) многоэтапной выборкой
  - в) экспертной выборкой
  - г) простой случайной выборкой
  - д) выборкой «снежный ком»
25. Исследователи предпочитают вероятностные или случайные выборочные методы перед невероятностными, считая их
- а) более современными
  - б) более распространенными
  - в) более точными и надежными
  - г) более понятными
  - д) более близкими
26. Выборка, составленная из тех людей, о которых известно, что они обладают специальными познаниями, называется
- а) кластерной выборкой
  - б) многоэтапной выборкой
  - в) экспертной выборкой
  - г) простой случайной выборкой
  - д) выборкой «снежный ком»

## ТЕМА 2. ВАЛИДНОСТЬ КОНСТРУКТОВ

Валидностью конструктов называют степень, в которой возможны обоснованные теоретические выводы на основе применяющейся в исследовании операционализации. Подобно внешней валидности, валидность конструктов имеет отношение к обобщениям. Однако, при том, что внешняя валидность предполагает обобщения, переносимые из контекста исследования на остальных людей, валидность конструктов предполагает обобщения, пе-

реносимые из экспериментов или измерений на *концепции* экспериментов или измерений. Например, когда вы измеряете то, что называется «самоуважением», то, что на самом деле вы измеряете?

В этой связи нужно рассказать случай. Во время второй мировой войны, правительство США привлекло несколько тысяч психологов для разработки различных методов измерения. Для отбора пилотов требовались личностные тесты, нужны были методики отбора персонала, психофизические методы определения времени реакции и т. п. После войны эти психологи стали искать работу вне военного сектора, и со своими разработками они пришли на гражданские предприятия. В начале пятидесятых годов Американская психологическая ассоциация стала придавать большое внимание проблеме качества или валидности новых методов измерения и стала вводить стандарты для психологических методов измерения. Эти усилия привели к формулированию представления о валидности конструкторов, которая, в свою очередь, привела к более грандиозной идее номологических сетей.

Номологическая сеть выступает основой идеи о валидности конструкторов, но так и не позволила исследователям определять, обладают ли применяемые методы измерения конструктивной валидностью. В 1959 году была предпринята попытка разработать метод оценивания валидности конструкторов с применением так называемой матрицы мультисвойств-мультиметодов (a multitrait-multimethod matrix, МТММ). В соответствии с подходом, использующим матрицу мультисвойств-мультиметодов, для того, чтобы утверждать, что измерение обладает конструктивной валидностью, необходимо продемонстрировать, что в измерении присутствуют и конвергентная валидность, и дискриминантная валидность. **Конвергентная** валидность считается продемонстрированной, если удастся показать, что измерения, которые теоретически считаются сильно взаимосвязанными, на практике сильно взаимосвязаны. **Дискриминантная** валидность считается установленной в том случае, когда удастся продемонстрировать, что измерения, которые не должны быть друг с другом взаимосвязаны, в действительности не взаимосвязаны.

Это трудно осуществить в реальности, особенно в контексте психологических исследований, поэтому мало попыток было предпринято в данном направлении. Если внимательно подойти к тем представлениям о валидности конструкторов, которые лежат в основе номологических сетей и матрицы мультисвойств-мультиметодов, то удастся выявить одну ведущую тему – тему идеи или «паттерна». Когда мы утверждаем, что наши методы измерения обладают конструктивной валидностью, то мы, по сути, утверждаем, что мы как исследователи понимаем то, каким образом наши экспериментальные конструкторы или теории измерения функционируют теоретически, и мы утверждаем, что можем представить доказательства того, что на практике они функционируют именно так, как мы себе представляем. У исследователя обязательно имеется теория относительно того, каким образом методы измерения в эксперименте соотносятся друг с другом (и с другими теоретическими компонентами), если угодно – имеется *теоретический паттерн*. Исследователь получает доказательства того, что данный метод измерения действи-

тельно функционирует именно таким образом, благодаря наблюдениям, или *наблюдаемому паттерну*.

Когда мы утверждаем валидность конструкторов, мы, по сути, утверждаем, что наш наблюдаемый паттерн – то, как вещи функционируют в реальности – согласуется с нашим теоретическим паттерном – тем, что мы думаем о реальности. Этот процесс может быть назван *паттерном соответствия*, он и есть сердцевина валидности конструкторов. Именно этот процесс заложен в основу номологической сети и в основу матрицы мультисвойств-мультиметодов. В принципе, возможно разработать конкретные и осуществимые методы, позволяющие практическим исследователям оценивать паттерны соответствия, т. е. оценивать конструктивную валидность своих исследований.

## **ИЗМЕРЕНИЕ РАЗЛИЧНЫХ ТИПОВ ВАЛИДНОСТИ**

В методологической литературе существует большая неразбериха, вызванная употреблением различных терминов для обозначения измерений валидности. Следует учитывать две вещи.

Во-первых, бессмысленно ограничивать наше рассмотрение лишь валидностью измерений. Необходимо рассматривать валидность любой операционализации. Иными словами, каждый раз, когда концепция или понятие транслируется в реальное функционирование (это и есть **операционализация**), следует задать себе вопрос о том, насколько хорошо выполнено транслирование. Эта проблема имеет отношение не только к экспериментам, но и к измерениям. Она же имеет отношение к составлению выборок. Интересующая исследователя популяция – это и есть «конструкт», а выборка – это операционализация. С этих позиций, уместно говорить о валидности конструкта при составлении выборки.

Во-вторых, мы будем употреблять термин «валидность конструкта» применительно к общему случаю транслирования любого конструкта в операционализацию. Для того, чтобы выразить различные аспекты конструктивной валидности, мы будем употреблять иные термины.

Имея в виду все сказанное выше, перечислим типы валидности, которые встречаются в исследовательских работах, когда идет речь о качестве измерения:

### **❖ Валидность конструкторов**

- **Валидность транслирования**
  - ✓ **Лицевая валидность**
  - ✓ **Содержательная валидность**
- **Критериальная валидность**
  - ✓ **Прогностическая валидность**
  - ✓ **Одновременная валидность**

- ✓ **Конвергентная валидность**
- ✓ **Дискриминантная валидность**

Начнем с того, что термин «валидность конструктора» оказывается наиболее широким. Валидностью конструктора называется приблизительная истинность вывода о том, что наша операционализация точным образом отражает данный конструктор. Все остальные термины различными способами обращены к этой общей проблеме.

Во-вторых, нужно проводить различие между двумя широкими типами валидности: валидностью транслирования и критериальной валидностью. Каждый из них соотносится с двумя главными способами оценивания валидности операционализации. **Валидность транслирования** обращена к тому, насколько хорошо операционализация отражает конструктор. Данный подход по своей природе оказывается определительным – он предполагает, что исследователь обладает детальным определением конструктора и способен проверить операционализацию применительно к конструктору. **Критериальная валидность** обращена к тому, функционирует ли операционализация таким образом, который предусмотрен теорией конструктора. Это более реляционный подход к валидности конструктора. Он подразумевает, что наша операционализация должна функционировать предсказуемым образом по отношению к другим операционализациям, основанным на нашей теории конструктора. (Если это трудно понять, пропустите эту часть, потом вернетесь к ней и снова перечитаете).

Рассмотрим конкретные типы валидности.

## **Валидность транслирования**

### **Лицевая валидность**

В случае лицевой валидности мы рассматриваем операционализацию с «лицевой стороны», чтобы определить, насколько хорошо удалось транслирование конструктора в операционализацию. Это, вероятно, самый слабый способ доказательства валидности конструктора. Например, можно взять методику измерения математических способностей, прочитать вопросы в ней и сказать: «Ис, это хорошая методика измерения математических способностей (т. е. эта методика заслуживает наименования «математические способности»)). Либо, можно ознакомиться с программой профилактики подростковых беременностей и заключить, что «Ис, это действительно программа профилактики подростковых беременностей». Разумеется, если больше ничего не делать для оценки лицевой валидности, тогда доказательства валидности окажутся очень слабыми, потому что они основаны на субъективном суждении. (Обратите внимание на то, что «слабое» доказательство не означает «ошибочного» доказательства. Исследовательский процесс предполагает опору на субъективные суждения. Но в такой форме суждения мало убедительны для других людей). Качество оценки лицевой валидности можно существенно улучшить, если сделать оценку более систематической. Например, если мы оцениваем

лицевую валидность методики измерения математических способностей, то для большей убедительности следует разослать тест тщательно сформированной экспертной выборке специалистов по тестированию математических способностей, а они выразят свои суждения о том, что ваша методика хороша для измерения математических способностей.

### **Содержательная валидность**

**Содержательная валидность** направлена на проверку операционализации по отношению к релевантной содержательной области конструкта. Данный подход подразумевает, что у исследователя имеется хорошее подробное описание содержательной области, что на самом деле не всегда так. К примеру, мы можем изложить все критерии, которым должна соответствовать программа, стремящаяся называться «программой профилактики беременностей среди подростков». Эта содержательная область, вероятно, может содержать

- ❖ определение адресной группы,
- ❖ критерии, в соответствии с которыми программу надлежит считать именно профилактической, а не терапевтической,
- ❖ базовую информацию о беременности,
- ❖ о воздержании,
- ❖ о методах контроля рождаемости, и т. п.

Далее, вооружившись этими критериями, можно приступить к исследованию собственно программы. Только те программы, которые отвечают перечисленным критериям, могут с полным правом называться «программами профилактики подростковых беременностей». Это представляется довольно прямолинейным, однако, в отношении других конструктов, таких как самоуважение или интеллект нелегко сформулировать критерии, образующие конкретную содержательную область.

### **Критериальная валидность**

Критериальная валидность позволяет проверить функционирование операционализации в отношении какого-то критерия. В чем отличие критериальной валидности от содержательной валидности? В случае содержательной валидности, критерии представляют собой определение самого конструкта. В случае критериальной валидности, обычно создается прогноз о том, каким образом будет функционировать операционализация с точки зрения нашей теории конструкта. Различия между отдельными типами критериальной валидности связаны с тем, какие критерии применяются в качестве стандарта для суждения.

### **Прогностическая валидность**

**Прогностическая валидность** направлена на оценивание способности операционализации прогнозировать то, что она теоретически может прогнозировать. Например, мы теоретически считаем, что методика измерения

математических способностей должна быть способна прогнозировать то, насколько человек подходит для инженерных профессий. Нашу методику можно дать опытным инженерам и посмотреть, имеется ли высокая корреляция между баллами по данной методике их зарплатами в качестве инженеров. Высокая корреляция будет служить подтверждением прогностической валидности – она будет показывать, что наша методика способна точно прогнозировать то, что мы теоретически считаем она должна быть способна прогнозировать.

### **Одновременная валидность**

**Одновременная валидность** направлена на оценку способности нашей операционализации проводить различие между теми группами, между которыми она должна теоретически быть способна проводить различие. Например, если мы создали методику для оценки маниакально-депрессивных состояний, то наша методика должна быть в состоянии проводить различие между теми людьми, которым ставится диагноз «маниакальная депрессия», и теми людьми, которым ставится диагноз «параноидальная шизофрения».

### **Конвергентная валидность**

**Конвергентная валидность** направлена на определение того, в какой степени данная операционализация сходна (конвергентна) с другими операционализациями, с которыми она теоретически должна быть сходна. Например, для доказательства конвергентной валидности теста математических способностей, можно рассчитать корреляцию между баллами по нашему тесту и баллами по другим тестам, которые направлены на измерение математических способностей. Высокие корреляции будут доказательством конвергентной валидности.

### **Дискриминантная валидность**

**Дискриминантная валидность** направлена на определение того, в какой степени наша операционализация отличается (дивергентна) от других операционализаций, от которых она теоретически должна отличаться. Например, для доказательства дискриминантной валидности теста математических способностей, рассчитывается корреляция между баллами по математическому тесту и баллами по тесту вербальных способностей. Низкие значения корреляции будут доказательством дискриминантной валидности.

## **ПОНЯТИЕ О ВАЛИДНОСТИ КОНСТРУКТОВ**

**Валидностью конструктов** обозначают степень, в которой операционализация теоретических конструктов в исследовании позволяет делать обоснованные выводы. Рассматриваемую проблему удобно разделить на две большие сферы – сферу теории и сферу наблюдения.

Сфера теории находится у вас в голове, и вы пытаетесь объяснить её другим людям. В неё включаются все идеи, теории и гипотезы относительно



окружающего мира, которыми вы обладаете. В сфере теории лежат ваши представления об эксперименте, о конструктах, об измерениях.

Сфера наблюдения состоит из того, что происходит в окружающем мире и из того, в чем происходящее проявляется. В сфере наблюдения находятся действительные эксперименты и измерения.



Валидность конструкта представляет собой оценку того, насколько хорошо вы транслировали свои идеи и теории в действительные эксперименты и измерения. Почему это важно? Потому что, когда вы думаете об окружающем мире или обсуждаете его с другими людьми (сфера теории), вы употребляете слова, которые представляют понятия. Если вы говорите кому-то, что есть такой учебник математики или репетитор, который поможет ребенку лучше усвоить математику, то вы общаетесь на уровне понятий или конструктов. Вы не описываете в операциональных подробностях того, что репетитор будет делать с ребенком. Вы не называете конкретных вопросов математического теста, с которыми ребенок станет лучше справляться. Вы говорите в общем смысле, применяя конструкты.

Если ваша рекомендация основывается на исследовании, которое свидетельствует об особом типе репетиторства, улучшаем ли баллы ребенка по математическому тесту, то вам необходимо быть уверенными в том, что тот тип репетиторства, который вы рекомендуете, полностью аналогичен проведенному исследованию, и что тот результат, который должен получиться, соответствует тому, который был измерен в эксперименте. В этом смысле ва-

лидность конструкта можно рассматривать в качестве «истинности наименования» разновидности случая.

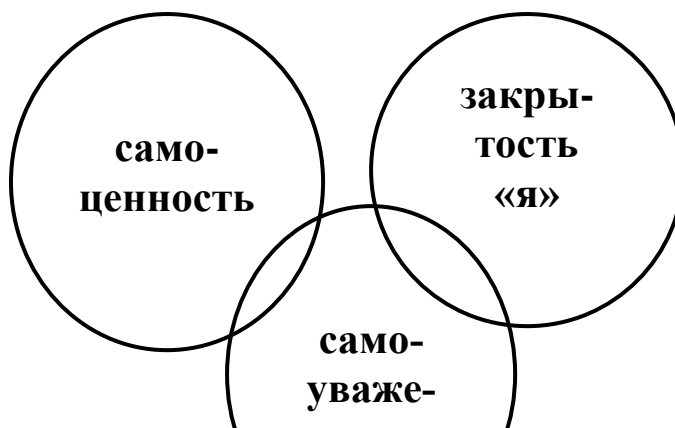
В действительности, существуют два широких образа представления понятия конструктивной валидности. Первый называется «дефиниционалистским» направлением, потому что с этой точки зрения подтверждение валидности конструкта заключается в определении конструкта настолько точным образом, чтобы его можно было бы непосредственно операционализировать. С дефиниционалистской точки зрения, либо конструкт операционализируют правильно, либо конструкт операционализируют неправильно – это тип мышления «или/или». Или вы измеряете самоуважение, или вы измеряете не самоуважение.

Второе направление называется «реляционалистское». С точки зрения реляционалиста, вещи не являются или черными, или белыми – все понятия более или менее соотносятся между собой. Смыслы терминов или конструктов различаются относительно, а не абсолютно. Ваша методика измерения, возможно, во многом измеряет самоуважение, но, возможно, она не измеряет самоуважения вообще. Возможно, существует другая методика, которая ближе к конструкту самоуважения, чем ваша методика. Реляционализм подразумевает, что смысл изменяется постепенно. Реляционализм отвергает идею о том, что в качестве основы для определения конструктов мы можем полагаться на операциональные формулировки.

Для того, чтобы лучше представить себе различие между дефиниционализмом и реляционализмом, давайте представим себе то, каким образом закон соотносится с истиной. Например, свидетель в суде дает присягу говорить «правду, только правду и ничего кроме правды». Что это означает? Если свидетеля заставляют клясться только в том, что он будет говорить правду, то свидетель может воспринимать это как обязанность не говорить неправды. Но это не будет гарантией того, что свидетель скажет все, что считает правдивым. О некоторых важных вещах свидетель может умолчать. Однако, тем не менее, свидетель говорит правду. Просто он не говорит всей правды. С другой стороны, свидетель обязан не говорить «ничего кроме правды». Это означает, что можно просто утверждать, что утверждение  $x$  – правда, а утверждение  $y$  – неправда.

Теперь давайте посмотрим, каким образом эта клятва транслируется в контекст измерения и валидности конструктов. Допустим, мы хотим, чтобы наше измерение отражало «конструкт, весь конструкт и ничего, кроме конструкта». Что это означает?

### ... и ничего кроме самоуважения





Допустим, что у нас имеются пять понятий, которые концептуальным образом соотносятся друг с другом: самоуважение, самооценность, закрытость «я», самоуверенность и открытость. Большинство людей скажут, что эти понятия аналогичные, хотя их можно отличить друг от друга. Если мы пытаемся разработать методику измерения самооценки, тогда что будет значить измерение «самоуважения, только самоуважения, ничего кроме самоуважения»? Если понятие самоуважения пересекается с другими понятиями, то как нам измерить всё самоуважение (которое, предположительно, включает пересекающиеся части других понятий) и ничего кроме самоуважения? Это немыслимо! Если считать, что смысл имеет реляционистскую природу – что некоторые понятия «ближе» по смыслу, чем другие понятия – тогда рассматриваемая здесь модель не может быть моделью валидности конструкторов.

Фактически, мы увидим, что большинство методологов психологических исследований (нарочно или непреднамеренно) отвергают дефицицистское направление и отдают предпочтение реляционистскому направлению. Для того, чтобы обеспечить валидность конструкторов, необходимо соблюдать следующие требования:

- ❖ Необходимо включить конструктор, который подлежит операционализации (например, самоуважение) в **семантическую сеть** (или смысловую сеть). Это означает, что нужно решить, на какое другое понятие данный конструктор более-менее похож по смыслу.
- ❖ Необходимо найти непосредственное доказательство того, что мы контролируем операционализацию конструктора, т. е. что наши операционализации похожи на то, на что они должны теоретически быть похожими. Если мы хотим измерить самоуважение, то мы должны суметь объяснить почему мы операционализировали вопрос именно таким образом. Нужно уметь доказать, что опросник измеряет самоуважение, не включая в измерение способностей и т. п.
- ❖ Нужно привести подтверждения того, что наши **данные подкрепляют наше теоретическое видение** взаимосвязей между конструкторами. Если мы считаем, что самоуважение ближе по смыслу к самооценности, чем к тревожности, то нужно суметь показать, что измерение самоуважения больше коррелирует с измерением самооценности, чем с измерением тревожности.

## КОНВЕРГЕНТНАЯ И ДИСКРИМИНАНТНАЯ ВАЛИДНОСТЬ

Конвергентную и дискриминантную валидность относят к подкатегориям или подтипам валидности конструкторов. Важно обратить внимание на то, что они действуют совместно – если удастся доказать, что есть подтверждения для конвергентной валидности и для дискриминантной валидности, значит, по определению, удалось доказать, что есть подтверждения валидности конструкторов. Но по одиночке ни одной из них недостаточно для обоснования валидности конструкторов.

Конвергентная валидность и дискриминантная валидность могут быть представлены в виде взаимно пересекающихся утверждений:

измерения конструкторов, которые теоретически *ДОЛЖНЫ* соотноситься друг с другом, фактически, наблюдаются в качестве взаимно соотнесенных

(т. е. необходимо суметь показать соответствие или *КОНВЕРГЕНЦИЮ* между сходными конструкторами)

и

измерения конструкторов, которые теоретически *НЕ ДОЛЖНЫ* соотноситься друг с другом, фактически, наблюдаются в качестве взаимно не соотнесенных

(т. е. необходимо суметь оказать несоответствие между несходными конструкторами)

Для того, чтобы оценить степень, в которой любые два измерения соотносятся друг с другом, обычно применяется коэффициент корреляции. Иными словами, рассматривается взаимосвязь между измерениями. Корреляции между теоретически сходными измерениями должны оказаться большими, а корреляции между теоретически несходными измерениями должны оказаться маленькими.

Главная проблема в связи с идеей конвергентности/дискриминантности заключается в употреблении слов «большая» и «маленькая» в вышеприведенном предложении. Вопрос прост: насколько велика должна быть корреляция, чтобы служить подтверждением конвергентности, и насколько мала она должна быть, чтобы служить подтверждением дискриминантности? Ответ такой – мы не знаем! Вообще, требуется, чтобы конвергентная корреляция была как можно больше, а дискриминантная корреляция – как можно меньше, но это не всеобщее правило. Пусть нас это не останавливает. В действительности можно сказать, что конвергентная корреляция должна быть больше, чем дискриминантная корреляция. Это хотя бы кое-что проясняет.

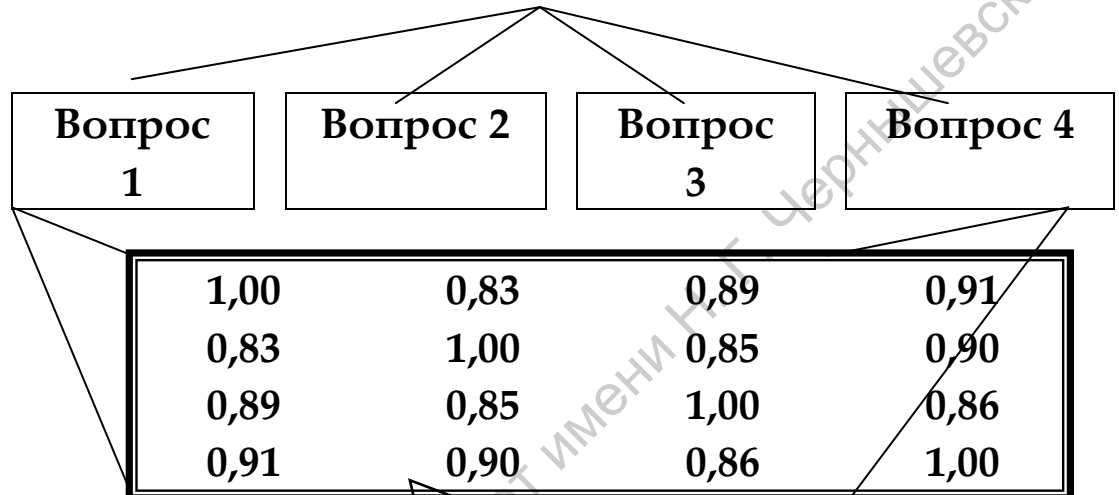
Прежде, чем углубиться в идею конвергенции/дискриминации, рассмотрим каждую из них на примере.

## Конвергентная валидность

Для подтверждения конвергентной валидности, необходимо показать, что те измерения, которые должны быть взаимосвязаны, реально взаимосвязаны.

### Теория

## Конструкт самоуважения



### Наблюдение

Корреляции служат подтверждением того, что все пункты опросника относятся к одному и тому же конструкту

Мы видим, что все четыре измерения (каждое из которых выступает отдельной меткой на шкале) отражают конструкт «самоуважение». Например, вопрос 1 может быть сформулирован таким образом: «Я хорошо к себе отношусь» с применением формата 1-5 шкалы Лайкерта. Теоретически, мы полагаем, что все четыре вопроса отражают идею самоуважения (по этой причине верхняя часть схемы названа «Теория»). В нижней части схемы («Наблюдение») представлены интеркорреляции по четырем вопросам. Допустим, мы предложили нашу шкалу выборке из респондентов, получили ответы и рассчитали корреляции. Мы видим, что интеркорреляции для всех пар вопросов очень большие (корреляция бывает в пределах от  $-1,0$  до  $+1,0$ ). Тем самым, мы получили подтверждение нашей теории о том, что все четыре вопроса связаны с одним и тем же конструктом.

Однако, следует отметить, что из того, что интеркорреляции показывают связь всех вопросов с одним и тем же конструктом, отнюдь не следует, что этим конструктом автоматически оказывается конструкт «самоуважение». Возможно, что все четыре вопроса связаны с каким-либо другим кон-

структом. Все, что мы можем предполагать на основании паттерна корреляций – четыре вопроса конвергируют с чем-то единым, как бы оно ни называлось.

### Дискриминантная валидность

Для подтверждения дискриминантной валидности, необходимо показать, что те измерения, которые не должны быть взаимосвязаны, реально не взаимосвязаны. На схеме изображены четыре измерения (по два для каждой шкалы). Однако, на сей раз два вопроса, как мы полагаем, отражают конструкт «самоуважение», тогда как другие два, как мы полагаем, отражают конструкт «локус контроля». В верхней части схемы представлены теоретически ожидаемые нами взаимосвязи между четырьмя вопросами. Если дискриминантная валидность присутствует, то взаимосвязь между измерениями двух различных конструктов должна быть очень малой (мы не знаем, насколько мала должна быть «малая» взаимосвязь, но мы разберемся с этим позже). Имеются четыре корреляции между измерениями, которые отражают различные конструкты, их значения приведены в нижней части схемы («Наблюдение»). Видно, что эти четыре корреляции между конструктами очень малы (т. е. близки к нулю), что они определенно меньше конвергентных корреляций на предыдущей схеме.

#### Теория



#### Наблюдение

Как и в предшествующем случае, из того, что мы представили подтверждения того, что два комплекса измерений по два измерения в каждом представляются соотношенными с различными конструктами (по причине того, что их интеркорреляции малы), отнюдь не следует, что конструкты, с которыми они связаны – это именно «самоуважение» и «локус контроля». Однако, этим корреляциями в действительности подтверждается, что два комплекса измерений не соотносятся между собой.

### Обобщим вышеизложенное

Итак, к чему мы пришли? Мы увидели, как получают подтверждения конвергентной и дискриминантной валидности отдельно одна от другой. Для того, чтобы утверждать наличие валидности конструкторов, необходимо суметь показать, что оба эти типа валидности нашли свои подтверждения. Иными словами, необходимо объединить оба принципа воедино в единый анализ. Выглядит это так.

#### Теория

### Конструктор са- моуважения

### Конструктор локуса контроля

СУ <sub>1</sub>	СУ <sub>2</sub>	СУ <sub>3</sub>	ЛК <sub>1</sub>	ЛК <sub>2</sub>	ЛК <sub>3</sub>
-----------------	-----------------	-----------------	-----------------	-----------------	-----------------

	СУ <sub>1</sub>	СУ <sub>2</sub>	СУ <sub>3</sub>	ЛК <sub>1</sub>	ЛК <sub>2</sub>	ЛК <sub>3</sub>
СУ <sub>1</sub>	1,00	0,83	0,89	0,02	0,12	0,09
СУ <sub>2</sub>	0,83	1,00	0,85	0,05	0,11	0,03
СУ <sub>3</sub>	0,89	0,85	1,00	0,04	0,00	0,06
ЛК <sub>1</sub>	0,02	0,05	0,04	1,00	0,84	0,93
ЛК <sub>2</sub>	0,12	0,11	0,00	0,84	1,00	0,91
ЛК <sub>3</sub>	0,09	0,03	0,06	0,91	0,91	1,00

#### Наблюдение

Корреляциями подтверждается и конвергентная валидность, и дискриминантная валидность, а, следовательно, подтверждается валидность конструкторов

На схеме показаны шесть измерений, три из которых теоретически соотносятся с конструктором самоуважения, а другие три, как мы предполагаем, соотносятся с конструктором локуса контроля. Верхняя часть схемы показывает теоретическую выкладку. В нижней части схемы представлена корреляционная матрица, основанная на пилотной матрице. Для понимания этой таблицы необходимо сначала выявить конвергентные корреляции и дискрими-

нантные корреляции. Имеются два комплекса или два блока конвергентных коэффициентов: один блок 3 x 3 образован интеркорреляциями самоуважения, а другой блок 3 x 3 образован интеркорреляциями локуса контроля. Имеются еще два блока 3 x 3 дискриминантных коэффициентов, и если проявить смекалку, то понимаешь, что они являются теми же самыми значениями в зеркальном отражении.

Как мы можем интерпретировать паттерны корреляций? Уже говорилось о том, что неизвестно никаких строгих правил в отношении того, насколько велики или малы должны быть корреляции для того, чтобы служить подтверждением валидности того или иного типа. Однако, известно, что конвергентные корреляции должны всегда быть больше дискриминантных корреляций. Посмотрим получше на таблицу и увидим, что в нашем примере конвергентные корреляции *всегда* больше дискриминантных корреляций. Из этого можно сделать вывод о том, что данная корреляционная матрица содержит подтверждения и конвергентной валидности, и дискриминантной валидности – и всё в одном анализе!

Однако, при том, что данный паттерн подтверждает и конвергентную валидность, и дискриминантную валидность, свидетельствует ли он о том, что три измерения самоуважения действительно измеряют самоуважение, или что три измерения локуса контроля действительно измеряют локус контроля? Разумеется, нет. Иначе было бы слишком просто.

Итак, что пользы в нашем анализе? Он демонстрирует, что, как предполагалось, три измерения самоуважения отражают один и тот же конструкт (чем бы он ни оказался), три измерения локуса контроля также отражают один и тот же конструкт (чем бы он ни являлся), а два комплекса измерений представляются отражающими два других конструкта (чем бы они ни были). Для одного простого анализа не так плохо.

Как перейти к действительно интересным вопросам? Каким образом нам показать, что наши измерения действительно отражают самоуважение или локус контроля? Придется вас разочаровать, но на этот вопрос не существует простого ответа (готов поспорить, что вы знали, что так и будет). Существует множество вещей, которые мы можем сделать в связи с этим вопросом. Прежде всего, можно подойти к валидности конструктов другими способами, чтобы получить подтверждения того, что мы действительно измеряем то, что мы измеряем. Например, можно применить подходы с точки зрения лицевой валидности или содержательной валидности с целью показать, что наши измерения отражают те конструкты, о которых мы говорим.

Один из наиболее мощных подходов состоит в том, чтобы включить в измерение ещё больше конструктов. Чем более комплексной станет наша теоретическая модель (если мы отыщем подтверждения верного паттерна корреляций), тем больше доказательств мы получим тому, что мы знаем, о чем мы говорим (теоретически). Разумеется, добавление множества новых измерений затрудняет вычисление всех корреляций, необходимых для выявления истинного паттерна. Во многих исследованиях стоимость исследования из-за этого увеличивается. Однако, при всей непрактичности такого под-



хода, если удастся себе позволить добавление новых измерений и конструкторов, то увеличивается возможность оценивать валидность конструкторов с применением таких подходов, как матрица мультисвойств-мультиметодов и номологическая сеть.

Вероятно, наиболее интересный подход к установлению валидности конструкторов заключен в идее о соответствии паттернов. Вместо того, чтобы рассматривать конвергентную валидность и дискриминантную валидность как *качественно* различные, подход с точки зрения соответствия паттернов рассматривает их как различия в *степени*. Эта идея представляется более разумной, и помогает избежать проблемы определения того, насколько велики или малы должны быть корреляции, чтобы можно было утверждать существование конвергенции или дискриминации.

## УГРОЗЫ ВАЛИДНОСТИ КОНСТРУКТОВ

Прежде, чем перейти к рассмотрению наиболее важных угроз валидности конструкторов<sup>1</sup>, нужно определить – что понимается под угрозой валидности. В научном исследовании мы стремимся прийти к выводу о том, что наш конструктор хорошо операционализирован, а измерения конструктора отражают то, что мы хотим, чтобы они отражали. Будет ли такой вывод верным? На что может быть направлена критика подобных выводов? Каким образом обосновать подобные выводы?

### Неадекватное дооперациональное объяснение конструкторов

Здесь нет никакой тяжеловесности. Слово **дооперациональный** означает «до транслирования конструктора в измерения». Слово **объяснение** означает определение того смысла, который исследователь вкладывает в конструктор. В чем здесь угроза? Представьте, что вы предложили новый тип подхода к реабилитации. Появляется критик, который говорит, что ваша программа *не нова*, и она *не программа реабилитации*. Говорят, что вы плохо продумали свои конструкторы. Есть несколько выходов из ситуации:

- ❖ лучше продумать свои конструкторы
- ❖ применить методы разъяснения своих конструкторов (например, наглядные – диаграммы и т. п.)
- ❖ пригласить экспертов к обсуждению ваших операционализаций

### Монооперациональный уклон

Монооперациональный уклон возможен в отношении независимой переменной или причины в эксперименте – он не применим к измерениям и к результатам. Если эксперимент проводится только в одной версии в одном месте и один раз, то есть риск не охватить всего понятийного многообразия. Каждая операционализация дефицитарна относительно того конструктора, на котором она основана. Если вы полагаете, что ваш эксперимент отражает

<sup>1</sup> Cook T. D., Campbell D. T. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin, 1979.

созданный вами конструкт, то критики будут склонны утверждать, что он отражает лишь ту конкретную версию, которую вы реализовали, а вовсе не тот конструкт, который вы мысленно имели в виду. Выход: постарайтесь провести различные версии эксперимента.

### **Монометодический уклон**

Монометодический уклон относится к измерениям и к наблюдениям, но не относится к причинам. Иначе говоря, это в основном та же самая проблема, что и монооперациональный уклон. Имея только одну версию опросника для измерения самоуважения, невозможно предоставить достаточно подтверждений тому, что вы действительно измеряете самоуважение. Критики скажут, что вы не измеряете самоуважение – вы измеряете лишь какую-то его часть. Выход: попытаться создать множество средств измерения главных конструктов и попытаться доказать (на материале пилотного исследования), что применяемый вами измерительный инструментариум функционирует именно так, как теоретически и должен функционировать.

### **Взаимодействие различных экспериментальных воздействий**

Допустим, что вы разработали новую программу, стимулирующую подростков из группы риска ходить в школу и не беременеть. Результаты исследования показывают, что девушки в экспериментальной группе чаще посещают школу и реже беременеют. Вы довольны своей программой, однако, появляется критик, который указывает на то, что контингент в вашей экспериментальной группе одновременно принимает участие в некоторых других программах, рассчитанных на такой же эффект. Можно ли в действительности считать достигнутый эффект результатом вашей программы? Реальная программа, которой подвергаются подростки, на самом деле оказывается сочетанием различных программ, в которых они принимают участие.

### **Взаимодействие тестирования и воздействия**

Само по себе тестирование или измерение способно ли сделать группу более восприимчивой к экспериментальному воздействию? Если ответ да, то тестирование на самом деле оказывается частью экспериментального воздействия, оно неотделимо от эффекта воздействия. Эта проблема связана с валидностью конструктов, потому что то, что мы хотим называть словом «эксперимент», в реальности включает в себя тестирование.

### **Ограниченная обобщаемость конструктов**

Эту угрозу валидности конструктов можно назвать «непреднамеренными последствиями». Вы проводите исследование и делаете вывод, что терапия X обладает эффектом. Терапия X действительно вызывает уменьшение симптомов, однако, вы не сумели предвидеть ужасные негативные последствия побочных эффектов этой терапии. Когда вы говорите, что терапия X эффективна, то под эффективностью вы подразумеваете лишь те симптомы, на

которые терапия X непосредственно направлена. Следует иметь в виду, что к обобщению наблюдаемого эффекта (терапия X эффективна) в отношении других возможных результатов следует подходить осторожно.

### **Неразделенность конструкторов и уровней конструкторов**

Предположим, что испытывается новое лекарственное средство против рака. Экспериментальная группа, составленная случайным образом, получает определенную дозу лекарства, а другая группа получает плацебо. Возможно, что наблюдаемый результат окажется единственно верным для данного уровня дозировки. Небольшие увеличения или уменьшения дозы способны изменить результат радикальным образом. В данном контексте, несправедливо использовать название лекарства в качестве названия (или конструктора), описывающего проводимое лечение, потому что исследуется лишь узкий диапазон дозировки. Подобно другим угрозам валидности конструкторов – название не соответствует тому, что исследуется.

### **«Социальные» угрозы валидности конструкторов**

Существуют некоторые важные угрозы валидности конструкторов, о которых следует сказать особо, потому что причина этих угроз лежит в социальной и человеческой природе исследовательского поведения.

### **Формулирование гипотез**

Большинство людей не просто пассивно принимают участие в исследовательских проектах. Они пытаются угадать – о чем это исследование. Они «догадываются» об истинной цели исследования. И они ведут себя в соответствии с собственными догадками, а не в соответствии с воздействием, оказываемым на них экспериментатором. Если проводить психолого-педагогическое исследование в аудитории, то студенты могут предположить, что главная зависимая переменная – это уровень учебной активности. Если уровень учебной активности студентов повышается – не потому, что это требуется экспериментом, а потому что студенты решили, что именно это изучается – тогда полученный результат исследования нельзя будет считать эффектом, возникшим в результате эксперимента. Соответственно, валидность конструкторов окажется под угрозой.

### **Предвзятое оценивание**

Многие люди испытывают тревогу, когда их оценивают. У некоторых существует фобия в отношении ситуации, связанных с тестированием и с измерениями. Если их предубежденность оборачивается низкими баллами, то их результат исследования нельзя считать эффектом эксперимента. Другая разновидность предвзятого оценивания связана с человеческим желанием «хорошо выглядеть» или «казаться умным». Если, стремясь выглядеть умным, испытуемый получает высокие баллы, то такой результат исследования

нельзя считать эффектом эксперимента. В обоих случаях смешивают предубежденность с экспериментом.

### **Ожидания экспериментатора**

В наши дни, когда проводится множество прикладных исследований за стенами лабораторий, термин «экспериментатор» обычно не употребляется применительно к лицу, ответственному за исследование. Поэтому более уместно говорить об ожиданиях исследователя. Исследователь способен исказить результаты исследования множеством способов, намеренно и непреднамеренно. Иногда исследователь может сообщить испытуемому о том, какой результат желателен для исследования. К этому исследователя может вынудить испытуемый, который желает казаться умным и ведет себя соответствующим образом. К примеру, исследователь может выразить удовлетворение, когда испытуемый дает желательный ответ. Если реакция испытуемого вызывается подобным поведением со стороны исследователя, то результат исследования не может считаться экспериментальным эффектом.

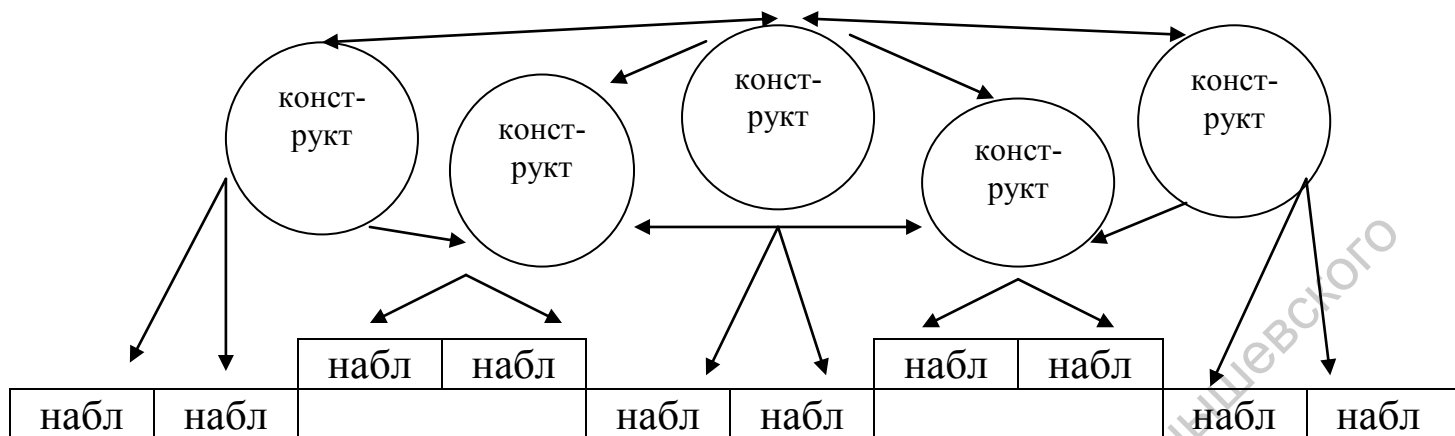
### **НОМОЛОГИЧЕСКАЯ СЕТЬ**

Идею номологической сети<sup>2</sup> разработали Ли Кронбах и Пол Мель в результате составной части усилий, предпринимавшихся Американской психологической ассоциацией с целью разработать стандарты психологического тестирования. Термин «номологический» заимствован из греческого языка и означает «законный», поэтому номологическую сеть можно себе представить как «сеть закона». Номологическая сеть является выражением взглядов Кронбаха и Меля на валидность конструкторов. То есть, для того, чтобы представить доказательства того, что ваш измерительный инструментариум обладает валидностью конструкторов, необходимо, как утверждали Кронбах и Мель, создать для вашего измерительного инструментария номологическую сеть. Эта сеть состоит из

- ❖ теоретической рамки для того, что вы собираетесь измерить,
- ❖ эмпирической рамки для того, каким образом вы обрелись измерять, и
- ❖ конкретизации связей между этими рамками.

## **НОМОЛОГИЧЕСКАЯ СЕТЬ выражает те понятия (конструкторы), которые представляют интерес в исследовании**

<sup>2</sup> Cronbach L., Meehl P. Construct Validity in Psychological Tests // Psychological Bulletin, 1955. No. 52 (4). P. 281 – 302.



## ...их наблюдаемые проявления и взаимосвязи между ними

Номологическая сеть основывается на множестве принципов, которыми руководствуется исследователь, когда занимается установлением валидности конструктов:

- ❖ В науке необходима ясность по поводу того, что чем является или какой смысл имеет, поэтому формулирование законов может предшествовать описанию того, что наблюдается.
- ❖ Законы в номологической сети могут отражать соотношения:
  - между наблюдаемыми свойствам или количествами
  - между теоретическими конструктами
  - между теоретическими конструктами и тем, что наблюдается
- ❖ По крайней мере некоторые законы в номологической сети должны относиться к тому, что наблюдается
- ❖ Разработка номологической сети зависит от того, насколько много нам известно о теоретическом конструкте, встречающемся в номологической сети, либо от увеличения определенности её компонентов
- ❖ Основное правило для включения в сеть нового конструкта состоит в необходимости создавать новые номологические законы, подтверждаемые наблюдениями, либо редуцировать количество номологических законов, необходимых для прогнозирования того, что наблюдается.
- ❖ Качественно отличные операционализации либо «накладываются» друг на друга, либо измеряют одно и то же, если их положение в номологической сети привязывает их к одной и той переменной конструкта.

Кронбах и Мель пытались связать сферу теоретического/концептуального со сферой наблюдаемого, ведь именно на это прежде

всего и направлена валидность конструкторов. Номологическая сеть способна выступить философским обоснованием валидности конструкторов, однако, она не образует практической и полезной методологии для действительного оценивания валидности конструкторов. Следующим шагом в эволюции идеи о валидности конструкторов стала разработка матрицы мультисвойств-мультиметодов.

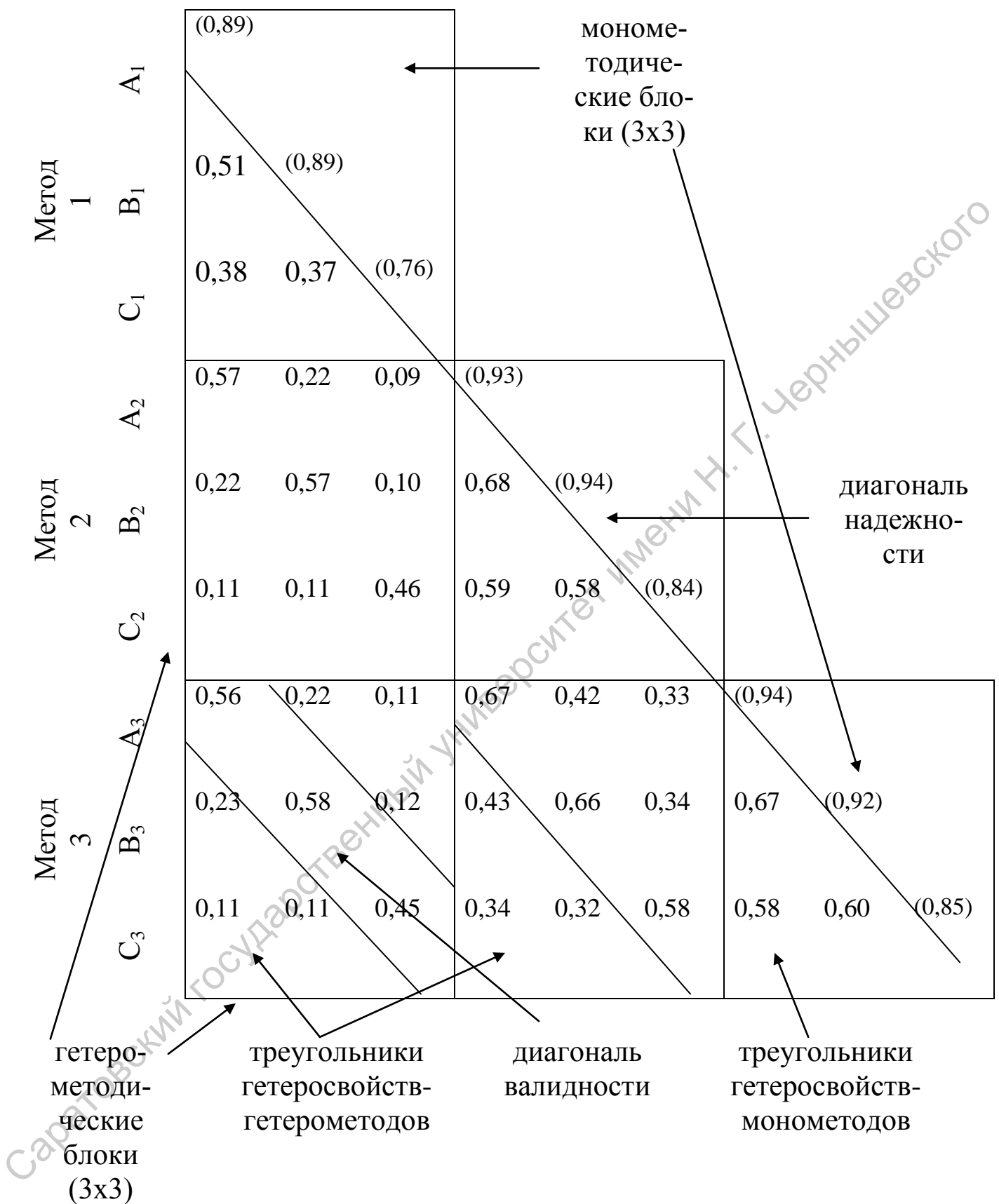
## МАТРИЦА МУЛЬТИСВОЙСТВ-МУЛЬТИМЕТОДОВ

Матрица мультисвойств-мультиметодов – это подход к оценке валидности конструкторов через комплекс наблюдений. Её разработали в 1959 году Д. Кэмпбелл и Д. Фиске<sup>3</sup>, которые пытались создать практическую методологию, пригодную для исследователей (в противовес идее о номологической сети, которая была теоретически полезна, но не предполагала методологии). Наряду с матрицей мультисвойств-мультиметодов, Д. Кэмпбелл и Д. Фиске предложили два новых типа валидности – конвергентную валидность и дискриминантную валидность – в качестве подкатегорий валидности конструкторов. **Конвергентная валидность** – это степень, в которой понятия, которые теоретически должны быть взаимосвязаны, взаимосвязаны в реальности. **Дискриминантная валидность** – это степень, в которой понятия, которые теоретически не должны быть взаимосвязаны, не взаимосвязаны в реальности. И конвергентную валидность, и дискриминантную валидность можно оценивать с применением матрицы мультисвойств-мультиметодов. Для того, чтобы иметь основания утверждать, что наши измерения обладают валидностью конструкторов, необходимо продемонстрировать подтверждения и конвергентной валидности, и дискриминантной валидности.

Матрица мультисвойств-мультиметодов представляет собой матрицу или таблицу корреляций, устроенную таким образом, чтоб облегчить интерпретацию оценок валидности конструкторов. Матрица мультисвойств-мультиметодов предполагает, что мы измеряем каждое из нескольких понятий (Д. Кэмпбелл и Д. Фиске называют их *свойствами*) поочередно при помощи различных методов (например, заполнение опросника, непосредственное наблюдение и т. п.) Матрица мультисвойств-мультиметодов представляет собой очень строгую методологию – в идеале, необходимо измерить *каждое* понятие *каждым* методом.

		Метод 1			Метод 2			Метод 3		
Свой-	ства	A <sub>1</sub>	B <sub>1</sub>	C <sub>1</sub>	A <sub>2</sub>	B <sub>2</sub>	C <sub>2</sub>	A <sub>3</sub>	B <sub>3</sub>	C <sub>3</sub>

<sup>3</sup> Campbell D., Fiske D. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. 1959. No. 56 (2). P. 81 – 105.



Для того, чтобы построить матрицу мультисвойств-мультиметодов, необходимо расположить корреляции внутри матрицы с точки зрения понятий и применявшихся методов. В нашем примере имеются три понятия (свойства 1, 2, 3), каждое из которых измеряется тремя различными методами (А, В, С). Обратите внимание на то, что каждый метод образует внутри матрицы от-

дельный блок. По своей сути, матрица мультисвойств-мультиметодов – это просто корреляционная матрица измерений, с одним исключением – вместо единиц по диагонали (как в типичной корреляционной матрице) по диагонали указаны значения надежности каждого метода измерения.

Прежде, чем приступать к интерпретации матрицы мультисвойств-мультиметодов, важно понять, каким образом выявляются различные части матрицы.

- ✓ Во-первых, следует обратить внимание на то, что в матрице нет ничего, кроме коэффициентов корреляции. Это квадратная, симметричная матрица, по этой причине мы рассматриваем лишь её половину – нижний треугольник.
- ✓ Во-вторых, эти корреляции можно группировать тремя способами: в диагонали, в треугольники и в блоки. Получается следующее:

- ❖ **Диагональ надежности (моносвойство-монометод).** Показатели надежности каждого измерения в матрице. Существуют различные способы оценки надежности (например, тест-ретест, внутренняя согласованность). На диагонали надежности представлено столько корреляций, сколько проведено измерений различными способами – в нашем примере девять измерений и девять показателей надежности. Первый показатель надежности в нашем примере – это корреляция свойства А, метода 1 со свойством А, методом 1 (сокращенно обозначим А1 – А1). Обратим внимание на то, что, по сути, это корреляция измерения с самим собой. Фактически, такая корреляция всегда оказывается совершенной (т. е.,  $r = 1,0$ ). Вместо этого мы вносим в матрицу величину надежности. Это значение также может считаться корреляцией моносвойства-монометода.

- ❖ **Диагонали валидности (моносвойство-гетерометод).** Корреляции между измерениями одного и того же свойства с применением различных методов. Поскольку матрица мультисвойств-мультиметодов организована методическими блоками, то внутри каждого методического блока существует одна диагональ валидности. Например, корреляция А1-А2 равна 0,57. Это корреляция между двумя измерениями одного и того же свойства (А), измеренного двумя различными методами (1 и 2). Поскольку оба измерения относятся к одному и тому же свойству или понятию, то следует ожидать, что они будут сильно коррелировать между собой. Такие значения можно считать корреляциями моносвойства-гетерометода.

- ❖ **Треугольники гетеросвойств-монометодов.** Это корреляции между измерениями, выполненные с применением одного и того же метода. Например, корреляция А1-В1 равна



0,51 верхнем левом треугольнике гетеросвойства-монометода. Обратим внимание на то, что у таких корреляций общим является метод измерения, но не свойство или понятие. Если такие корреляции велики, то по причине того, что измерение различных объектов с применением одного и того же метода приводит в результате к взаимосвязанным (коррелирующим) значениям. Либо, выражаясь более прямолинейно, мы получили сильный фактор метода.

- ❖ **Треугольники гетеросвойств-гетерометодов.** Такие корреляции различаются и своими свойствами, и своими методами. Например, корреляция А1-В2. Поскольку у этих корреляций отсутствуют как общие свойства, так и общие методы, то ожидается, что они окажутся одними из самых малых в матрице.
- ❖ **Блоки монометодов.** Они состоят из всех корреляций, для которых использованы одинаковые методы измерения. Блоков может быть столько, сколько применяли методов измерения.
- ❖ **Блоки гетерометодов.** Они состоят из всех корреляций, для которых *не были применены* одинаковые методы. Таких блоков может быть  $(K(K - 1))/2$ , где  $K$  = количеству методов. В нашем примере имеются 3 метода, поэтому есть 3 таких блока.

### Принципы интерпретации

Теперь можно приступить к изучению принципов интерпретации различных частей матрицы мультисвойств-мультиметодов. Необходимо усвоить, что интерпретация матрицы мультисвойств-мультиметодов требует от исследователя способности рассуждать. Даже если какие-то принципы матрицы мультисвойств-мультиметодов окажутся нарушенными, тем не менее, можно прийти к выводу о наличии прочной валидности конструкторов. Иными словами, *абсолютное* следование принципам в условиях прикладного исследования не строго обязательно. Интерпретацию матрицы мультисвойств-мультиметодов можно сравнить с тем, как врач-терапевт читает рентгеновский снимок. Наметанный глаз замечает такие вещи, которые новичок может пропустить! Исследователь, имеющий опыт интерпретации матрицы мультисвойств-мультиметодов, способен с её помощью выявить слабые места измерительного инструментария, а также оценить валидность конструкторов.

	Свойства	Опросник			Учитель			Родитель	
		СУ <sub>1</sub>	ЗЯ <sub>1</sub>	ЛК <sub>1</sub>	СУ <sub>2</sub>	ЗЯ <sub>2</sub>	ЛК <sub>2</sub>	СУ <sub>3</sub>	ЗЯ <sub>3</sub>
Оп-	СУ <sub>1</sub>	(0,89)							
	ЗЯ <sub>1</sub>	0,51	(0,89)						
	ЛК <sub>1</sub>	0,38	0,37	(0,76)					

Учи-	СУ <sub>2</sub>	0,57	0,22	0,09	(0,93)				
	ЗЯ <sub>2</sub>	0,22	0,57	0,10	0,68	(0,94)			
	ЛК <sub>2</sub>	0,11	0,11	0,46	0,59	0,58	(0,84)		
Роди-	СУ <sub>3</sub>	0,56	0,22	0,11	0,67	0,42	0,33	(0,94)	
	ЗЯ <sub>3</sub>	0,23	0,58	0,12	0,43	0,66	0,34	0,67	(0,92)
	ЛК <sub>3</sub>	0,11	0,11	0,45	0,34	0,32	0,58	0,58	0,60

С целью более конкретного прояснения принципов интерпретации, рассмотрим реальный пример. Представьте себе, что мы проводим исследование учащихся шестого класса и измеряем три черты или понятия: самоуважение (СУ), закрытость «я» (ЗЯ) и локус контроля (ЛК). Каждое из них мы измеряем тремя различными способами: с помощью опросника, с помощью оценок ребенка учителем, с помощью оценок ребенка родителем. Результаты представляем в виде матрицы мультисвойств-мультиметодов. Наша задача заключается том, чтобы выявить адекватные коэффициенты матрицы и прийти к решению относительно обоснованности утверждения о наличии валидности конструкторов.

Основные принципы или правила в отношении матрицы мультисвойств-мультиметодов таковы:

- ❖ Коэффициенты на диагонали надежности должны быть самыми большими в матрице. То есть свойство должно сильнее коррелировать с самим собой, нежели с чем-либо ещё! В нашем примере так и получается.
- ❖ Коэффициенты на диагоналях валидности должны значимым образом отличаться от нуля и быть достаточно большими. Это важное подтверждение конвергентной валидности. Все корреляции в нашем примере соответствуют указанному критерию.
- ❖ Коэффициент валидности должен быть больше, чем значения в одном с ним столбце и в одной с ним строке треугольника гетеросвойств-гетерометодов. Иными словами, корреляция СУ<sub>1</sub>-СУ<sub>2</sub> должна быть больше, чем корреляции СУ<sub>1</sub>-ЗЯ<sub>2</sub>, СУ<sub>1</sub>-ЛК<sub>2</sub>, СУ<sub>2</sub>-ЗЯ<sub>1</sub>, СУ<sub>2</sub>-ЛК<sub>1</sub>. В нашем примере это верно в всех случаях.
- ❖ Коэффициент валидности должен быть больше всех коэффициентов треугольника гетеросвойств-монометода. Это означает, что факторы свойств должны быть сильнее, чем факторы методов. Обратите внимание на то, что в нашем примере это справедливо *не во всех случаях*. К примеру, корреляция ЛК<sub>1</sub>-ЛК<sub>2</sub> составляет 0,46, что меньше, чем корреляции СУ<sub>2</sub>-ЗЯ<sub>2</sub>, СУ<sub>2</sub>-ЛК<sub>2</sub>, ЗЯ<sub>2</sub>-ЛК<sub>2</sub>, что свидетельствует о том, что здесь возможен фактор метода, особенно в отношении метода наблюдения за ребенком со стороны учителя.
- ❖ Следует рассмотреть взаимосвязи одного и того же паттерна свойств во всех треугольниках. Наш пример очевидно соответствует указанному критерию. Обратим внимание на то, что во всех треугольниках взаимоотношение СУ-ЗЯ почти в два раза превышает все взаимоотношения с участием ЛК.

## **Преимущества и недостатки матрицы мульти свойств-мульти методов**

Идея матрицы мульти свойств-мульти методов предоставила операциональную методологию оценивания валидности конструкторов. Оказалось возможным в одной матрице исследовать одновременно и конвергентную валидность, и дискриминантную валидность. За счет уравнивания методов со свойствами, Д. Кэмпбелл и Д. Фиске выделили важность рассмотрения наряду с тем, что мы измеряем, также то, каким образом проводится измерение.

Несмотря на очевидные достоинства, матрица мульти свойств-мульти методов не получила широкого распространения за прошедшие полвека после 1959 года. Этому есть ряд причин.

- ❖ Во-первых, в своей самой кристальной форме, матрица мульти свойств-мульти методов требует, чтобы каждое свойство измерялось каждым методом. Д. Кэмпбелл и Д. Фиске особенно на этом настаивали. Однако, в некоторых прикладных исследованиях бывает невозможно измерить все свойства всеми желательными методами (следует ли применять метод наблюдения при взвешивании?). Во многих прикладных психологических исследованиях бывает невозможно даже выделить методы из общего плана исследования.
- ❖ Во-вторых, матрица мульти свойств-мульти методов предполагает рассуждения исследователя, что не способствовало её широкому распространению (хотя следует признать, что в этом её сильная сторона). Многие исследователи желали бы получить тест валидности конструкторов, выражающийся единственным статистически коэффициентом, т. е. желали получить эквивалент коэффициента надежности. Матрица мульти свойств-мульти методов не рассчитана на количественное выражение степени валидности конструкторов.
- ❖ В-третьих, нацеленность матрицы мульти свойств-мульти методов на рассуждения означает, что различные исследователи вполне обоснованно могут прийти к различным выводам.

### **Модифицированная матрица мульти свойств-мульти методов - исключается фактор методов**

Как уже отмечалось, одним из наиболее трудных аспектов матрицы мульти свойств-мульти методов с точки зрения её практического применения оказывается её устройство, предполагающее комбинирование свойств и методов. Однако, для определения конвергентной валидности и дискриминант-

ной валидности отнюдь не требуется фактор методов. Для понимания этого следует ещё раз вернуться о взглядам Д. Кэмпбелла и Д. Фиске на конвергентную валидность и дискриминатную валидность.

### Что такое конвергентная валидность?

Это принцип, в соответствии с которым *измерения теоретически сходных конструктов должны сильно коррелировать*. Эту мысль можно развить далее и представить себе измерительный инструментарий, включающий несколько вопросов (пунктов), например, шкала с четырьмя делениями, предназначенная для измерения самоуважения. Если каждое деление на шкале действительно отражает конструкт «самоуважение», тогда следует рассчитывать на то, что все четыре пункта (вопроса) опросника окажутся в сильной степени интеркоррелирующими. Эти большие коэффициенты корреляции служат подтверждениями конвергентной валидности.

**Конструкт самоуважения**

Вопрос 1	Вопрос 2	Вопрос 3	Вопрос 4
1,00	0,83	0,89	0,91
0,83	1,00	0,85	0,90
0,89	0,85	1,00	0,86
0,91	0,90	0,86	1,00

### Что такое дискриминатная валидность?

Это принцип, в соответствии с которым *измерения теоретически различных конструктов не должны сильно коррелировать друг с другом*. Этот принцип проиллюстрирован на примере двух конструктов – самоуважения и локуса контроля, каждый из которых измерен двумя методами. Следует ожидать, что, по причине того, что измеряются разные конструкты, кросскорреляции между конструктами должны быть малыми. Эти малые коэффициенты корреляции выступают подтверждениями дискриминантной валидности.

Теперь можно собрать все воедино и посмотреть, каким образом конвергентная валидность и дискриминантная валидность могут рассматриваться одновременно. Итак, у нас есть два конструкта: самоуважение и локус контроля, каждый из которых измерен тремя методами. Корреляции внутри каждого конструкта должны оказаться больше кросскорреляций между конструктами, т. е. конвергентные корреляции должны оказаться больше дискриминантных корреляций.

Следует отметить одну важную вещь в отношении рассматриваемой матрицы – она не содержит фактор методов столь же очевидно, как матрица

мультисвойств-мультиметодов. Подобно матрице мультисвойств-мультиметодов, рассматриваемая нами матрица позволяет определять конвергентную валидность и дискриминантную валидность, но в ней явно представлены лишь интравзаимосвязи и интервзаимосвязи. Из нашего примера следует, что в матрице мультисвойств-мультиметодов в действительно заложены две главные темы.

Первая идея об одновременном рассмотрении паттерна конвергенции и дискриминации. Эта идея аналогична по своей цели номологической сети – паттерн взаимосвязей рассматривается на основе теории номологической сети.

Вторая идея, заложенная в матрицу мультисвойств-мультиметодов – это обращение к методам как к **потенциально запутывающему фактору**.

Методы могут запутать результаты, однако они не в любом исследовании создают для нас трудности. Хотя и приходится исследовать результаты с целью обнаружения возможных факторов методов, тем не менее, сочетание оценки запутанности результатов с оценкой валидности конструкторов предельно осложняет применение методологии матрицы мультисвойств-мультиметодов.

## Конструкт са- моуважения

## Конструкт локуса контроля

	СУ <sub>1</sub>	СУ <sub>2</sub>	СУ <sub>3</sub>	ЛК <sub>1</sub>	ЛК <sub>2</sub>	ЛК <sub>3</sub>
СУ <sub>1</sub>	1,00	0,83	0,89	0,02	0,12	0,09
СУ <sub>2</sub>	0,83	1,00	0,85	0,05	0,11	0,03
СУ <sub>3</sub>	0,89	0,85	1,00	0,04	0,00	0,06
ЛК <sub>1</sub>	0,02	0,05	0,04	1,00	0,84	0,93
ЛК <sub>2</sub>	0,12	0,11	0,00	0,84	1,00	0,91
ЛК <sub>3</sub>	0,09	0,03	0,06	0,91	0,91	1,00

Возможно, что разделение этих двух тем увеличит нашу способность выявлять конвергентную валидность и дискриминантную валидность. Но тогда как поступать с факторами методов? Один из возможных выходов представляется в отказе от тестирования методов в пользу повторения исследовательского проекта. Тем самым, если какой-то результат установлен с применением нескольких измерений, то можно задуматься над тем, будет ли тот же самый результат возникать при повторении исследования с применением

другого измерительного инструментария и методов измерения тех же самых конструкторов. Вопрос о методах представляется скорее вопросом об обобщении результатов исследования (через методы измерения), чем вопросом о валидности конструкторов.

При такой постановке вопроса мы переходим от идеи матрицы мультисвойств-мультиметодов к идее матрицы мультисвойств, которая обеспечивает нам возможность исследовать конвергентную валидность и дискриминантную валидность, а, следовательно, валидность конструкторов. Мы убедимся в этом, когда увидим, что конвергентная валидность и дискриминантная валидность представляют собой различия в степени. Тогда у нас появится основание оценивать валидность через соответствие паттернов.

## СОГЛАСОВАНИЕ ПАТТЕРНОВ КОНСТРУКТИВНОЙ ВАЛИДНОСТИ

*Паттерном* называется любое упорядочение объектов или единиц. Слово «упорядочение» здесь означает, что паттерн по определению не является случайным (произвольным) и, по крайней мере потенциально, может быть описан. Любые теории предполагают какой-то паттерн, однако, теории и паттерны это не одно и то же. Вообще, теория постулирует структурные взаимосвязи между ключевыми конструктами. Теория служит основой для генерирования паттернов. Например,  $E = MC^2$  это теоретическая формулировка. На основе этой формулы можно сформировать паттерны, если рассчитывать значение для одной переменной, зная значения для других переменных. Не все теории выражаются в математическом виде, особенно в прикладных психологических исследованиях, но во всех теориях содержится информация, обеспечивающая возможность формулирования паттернов.

Соответствие паттернов всегда предполагает попытку связать два паттерна, один из которых – теоретический паттерн, а второй – наблюдаемый или операциональный паттерн. В верхней части рисунка представлена область теории. Теория может возникать в результате формального теоретизирования, может состоять из мыслей и интуитивных догадок исследователя, либо получаться в результате некой комбинации того и другого. *Задача концептуализации* предполагает транслирование этих идей в некий конкретизируемый теоретический паттерн, заданный верхней частью рисунка. В нижней части рисунка представлена область наблюдения. В широком понимании в неё включаются непосредственные наблюдения в форме впечатлений, полевых записей и т. п., а также более формальные объективные измерения. Собрание или упорядочение релевантных операционализаций (т. е. релевантных теоретическому паттерну) называется паттерном наблюдения, который проиллюстрирован в нижней части нашего рисунка. Формулирование выводов предполагает *соотнесение, связь или соответствие этих двух паттернов*. В той степени, в какой паттерны согласуются<sup>4</sup> друг с другом, можно за-

<sup>4</sup> Trochim W. Pattern Matching, Validity, and Conceptualization in Program Evaluation. // Evaluation Review, 1985. No. 9 (5). P. 575 – 604.

ключить, что подтверждается данная теория и любые иные теории, предполагающие такой наблюдаемый паттерн.

*Область  
теории*



*Область  
наблюдения*

Важно показать, что не существует альтернативных достоверных теорий, которыми объясняется наблюдаемый паттерн. Эта задача намного облегчается, когда интересующий нас теоретический паттерн является уникальным<sup>5</sup>. На самом деле, усложненные теоретические паттерны подобны уникальным отпечаткам пальцев, поэтому затруднительно сконструировать внятные альтернативные паттерны, которые бы позволяли приходить к ана-

<sup>5</sup> Trochim W. Outcome Pattern Matching and Program Theory // Evaluation and Program Planning, 1989. No. 12. P. 355 – 366.

логичным результатам. В той степени, в какой теоретические и наблюдаемые паттерны не согласуются, теория может оказаться неверной или плохо сформулированной, наблюдения могут оказаться неадекватными или неточными, либо возможно сочетание того и другого.

В любых исследованиях применяются принципы согласования паттернов, хотя это редко делается осознанно. В контексте традиционного эксперимента с участием двух групп, к примеру, типичный теоретический паттерн результата представляет собой гипотезу о том, что существуют значимые различия между экспериментальной группой и контрольной группой. Паттерн наблюдаемого результата может включать в себя средние значения для двух групп по одному или более измерениям. Согласованность паттернов устанавливается путем проверки значимости с помощью t-критерия или дисперсионного анализа (ANOVA). При проведении опросного исследования (анкетирования), разновидности согласования паттернов выступают основанием для обобщения в отношении различных понятий или подгрупп популяции. При качественных исследованиях согласованность паттернов лежит в центре любых усилий по осуществлению тематического анализа.

Современные исследования можно охарактеризовать с точки зрения согласованности паттернов, однако, идея согласованности паттернов подразумевает намного большее. Она предлагает способы улучшения современных методов исследования. Говоря конкретно, согласованность паттернов подразумевает, что чем *сложнее паттерн, то, если он согласован, тем валиднее оказывается теория*. Согласованность паттернов не отличается фундаментальным образом от традиционных подходов, связанных с проверкой гипотез и построением моделей. ***Теоретический паттерн – это гипотеза о том, чего следует ждать от полученных данных. Наблюдаемый паттерн состоит из данных, применяемых для исследования теоретической модели.*** Главное отличие согласования паттернов от более традиционного подхода, направленного на проверку гипотез, заключается в том, что согласование паттернов стимулирует использование более сложных и подробных гипотез, а к наблюдениям подходит с многомерной перспективы, а не с одномерной.

### **Согласованность паттернов и валидность конструкторов**

Согласование паттернов находит применение в различных проблемных областях психологических исследований. Мы обратимся к одной из них – к проблемной области, связанной с валидностью конструкторов.

На рисунке представлена структура согласования паттерна в примере с пятью измеряемыми конструкторами: арифметика, алгебра, геометрия, правописание и чтение. С целью разработки теоретического паттерна на основе этих конструкторов мы прибегаем к конкретизации концепта. При конкретизации концепта мы придумываем множество вопросов, имеющих отношение к арифметике, алгебре, геометрии, правописанию и чтению. Мы сортируем вопросы, объединяя подобные вопросы с подобными. Нам становится очевидна связь каждого вопроса с остальными вопросами: более подобные вопросы мы



располагаем ближе друг к другу, а менее подобные вопросы отдаляем друг от друга. Если представить все вопросы в виде точек на графике, то между точками можно провести прямую линию. Мы получим матрицу расстояний между точками. Используя упорядоченные подобным образом вопросы, можно (1) разрабатывать измерительный инструментарий; (2) формировать выборку.

**Область теории**

- ❖ теории
- ❖ идеи
- ❖ интуитивные догадки

**Задача концептуализации**

**Теоретический паттерн**

**Согласование паттернов**

**Наблюдаемый паттерн**

**Организация данных**

**Область наблюдения**

- ❖ наблюдения
- ❖ данные
- ❖ измерения

- ❖ арифметика
- ❖ алгебра
- ❖ геометрия
- ❖ правописание
- ❖ чтение

**Конкретизация концепта**

**Конкретизированный концепт**

**Согласованность паттернов г**

**Корреляции**

**Анализ данных**

- ❖ арифметика ?
- ❖ алгебра ?
- ❖ геометрия ?
- ❖ правописание ?
- ❖ чтение ?

Со стороны области наблюдения, мы располагаем одним или несколькими тестовыми заданиями, включающими вопросы по арифметике, алгебре, геометрии, правописанию и чтению. Мы анализируем данные по этим зада-

ниям и конструируем матрицу корреляций между пунктами тестовых заданий.

**Область  
теории**

**Матрица расстояний между точками**

<u>Пара</u>	<u>Расстояние</u>
1,2	0,342
1,3	0,475
...	...
n-1, n	3,421

**Согласованность паттернов  $\Gamma$**

**Матрица корреляций между вопросами тестового задания**

<u>Вопросы</u>	<u>Корреляции</u>
1,2	0,92
1,3	0,99
...	...
n-1, n	0,03

**Область  
наблюдения**

Мы хотим сравнить матрицу расстояний между точками, в которой отражена конкретизация нашего концепта (т. е. теоретический паттерн) с корреляционной матрицей вопросов тестовых заданий (т. е. наблюдаемый паттерн). Каким образом мы это реализуем? Предположим, что для конкретизации конструкта мы использовали 100 вопросов, по 20 вопросов для каждого конструкта. Соответственно, тестовые задания включают 100 вопросов, по 20 в каждой учебной дисциплине. Таким образом, матрица принимает вид 100 x 100. Поскольку обе матрицы симметричны, то реально получаем  $(N(N - 1))/2 = (100(99))/2 = 9900/2 = 4,950$  пар корреляций (исключая диагональ). При линейном расположении значений матрицы корреляций, мы получаем вектор-

столбец, состоящий из 4950 номеров. Первый номер – значение корреляции в паре (1,2), следующий номер – пара (1,3) и т. д. до (N-1, N) или (99, 100). Теперь можно рассчитать общую корреляцию между двумя столбцами, которая будет являться корреляцией между нашим теоретическим паттерном и нашим наблюдаемым паттерном. Мы получим *корреляцию паттерна согласования*. Допустим, что в нашем примере она равна  $-0,93$ . Почему мы получили *отрицательную* корреляцию? Потому что мы коррелируем *расстояния* в концептуализации со *сходствами* в матрице корреляций. Мы можем рассчитывать на то, что чем *больше* дистанции в концептуализации, тем *меньше* корреляции между вопросами тестового задания, а чем *меньше* дистанции, тем *больше* корреляции.

Корреляция паттерна согласования – это итоговая оценка степени валидности конструкторов в рассматриваемом примере, потому что она выражает степень, в которой операциональные измерения отражают наши теоретические предположения.

### **Преимущества и недостатки согласования паттернов**

У подхода к оцениванию валидности конструкторов с позиций согласования паттернов имеются несколько недостатков. Наиболее очевидный недостаток состоит в том, что согласование паттернов требует точной теоретической конкретизации конструкторов. Обычно в прикладных психологических исследованиях этого не делают, по крайней мере, не на том уровне, который требуется в данном случае. Вероятно, этим следует заниматься. Вероятно, необходимы строгие формулировки, позволяющие структурировать теоретические и наблюдаемые паттерны одинаковым образом, для того, чтобы напрямую их коррелировать. Оба паттерна необходимо представлять в количественном виде, а, в конечном счете, представлять их в виде матриц, имеющих одинаковые размеры. В большинстве исследований сравнительно легко построить матрицу корреляций между вопросами тестовых заданий. Однако, редко прибегают к конкретизации концептов, позволяющей рассчитывать теоретические паттерны таким образом, чтобы связать их с наблюдаемыми паттернами. Вероятно, эти тоже следует заниматься.

У подхода, основанного на согласовании паттернов, имеется ряд преимуществ, особенно при сравнении с матрицей мультисвойств-мультиметодов:

- ❖ Согласование паттернов – это более общий и более гибкий подход, чем матрица мультисвойств-мультиметодов. Для согласования паттернов не требуется измерять каждый конструктор множеством методов.
- ❖ Согласование паттернов позволяет рассматривать конвергенцию и дискриминацию в виде *континуума*. Концепты более-менее подобны, поэтому их взаимосвязи более-менее конвергентны либо дискриминантны. Благодаря этому, мы можем провести различие между упрощенными дихотомическими категориальными поня-

тиями и постпозитивистским континуальным понятием конвергенции/дискриминации.

- ❖ Подход, основанный на согласовании паттернов, действительно позволяет оценивать итоговую валидность конструкторов данного комплекса измерений в данном конкретном контексте. Обратите внимание на то, что не определяют валидности конструкторов одного измерения. Это объясняется тем, что валидность конструкторов, как и конвергенция/дискриминация, это всегда относительный показатель. Точно так же, как для того, чтобы отличить что-либо от чего-либо требуется то, от чего можно отличать, валидность конструкторов может быть охарактеризована только на языке теоретической семантики или номологической сети – того концептуального контекста, внутри которого валидность конструкторов существует. Величина корреляции согласованности паттернов указывает в нашем конкретном исследовании – удалось ли продемонстрировать взаимоотношение между тем, что мы теоретически полагаем взаимосвязанным с нашим измерением, и тем, с чем наше измерение действительно взаимосвязано.
- ❖ Поскольку для согласования паттернов необходим более конкретный теоретический паттерн, нежели тот, с которым мы обычно имеем дело, то необходимо конкретизировать то, что мы думаем о конструкторах нашего исследования. Психологические исследования нередко критикуют за концептуальную расплывчатость, за то, что старые конструкторы облачаются в новый вид, за неспособность осуществить исследовательскую эволюцию в отношении главных теоретических конструкторов. Вероятно, делаемый при согласовании паттернов акцент на теории заставит исследователей подходить более скрупулезно к концептуальным основаниям эмпирической деятельности. В конечном счете, разве не к этому обращена валидность конструкторов?

### **Вопросы для самоконтроля по теме «Валидность конструкторов»**

1. Что такое теоретический паттерн и наблюдаемый паттерн?
2. Что такое операционализация?
3. Какие типы валидности встречаются в исследовательских работах, когда идет речь о качестве измерения?
4. Чему служит лицевая валидность?
5. На что направлена содержательная валидность?
6. Что позволяет проверить критериальная валидность?
7. Что оценивает прогностическая валидность?
8. Что оценивает одновременная валидность?
9. Что называют валидностью конструкторов?

10. В чем состоит «дефиниционалистское» направление представления понятия конструктивной валидности?
11. В чем состоит «реляционалистское» направление представления понятия конструктивной валидности?
12. Перечислите требования, которые необходимо соблюдать для того, чтобы обеспечить валидность конструкторов.
13. Сформулируйте определения конвергентной и дискриминантной валидности.
14. При помощи чего обычно оценивается та степень, в которой любые два измерения соотносятся друг с другом?
15. В чем состоит главная проблема в связи с идеей конвергентности/дискриминантности?
16. Что необходимо показать для подтверждения конвергентной валидности?
17. Что необходимо показать для подтверждения дискриминантной валидности?
18. Как преодолевается неадекватное дооперациональное объяснение конструкторов?
19. В чем состоит монооперациональный уклон и как его преодолеть?
20. В чем состоит нометодический уклон и как его преодолеть?
21. Охарактеризуйте «социальные» угрозы валидности конструкторов.
22. Из чего состоит номологическая сеть?
23. На каких принципах основывается номологическая сеть?
24. Каким образом выявляются различные части матрицы мультисвойств-мультиметодов?
25. Каковы основные принципы интерпретации мультисвойств-мультиметодов?
26. Каковы основные преимущества и недостатки матрицы мультисвойств-мультиметодов?
27. Благодаря чему можно улучшить способность выявлять конвергентную валидность и дискриминантную валидность?
28. Что называют словом «паттерн»?
29. Что служит основой для генерирования паттернов?
30. Что всегда предполагает соответствие паттернов?
31. В чем заключается задача концептуализации?
32. Как называется собрание или упорядочение операционализаций, релевантных теоретическому паттерну?
33. О чем можно сделать вывод на основании того, в какой степени паттерны согласуются друг с другом?
34. Когда применяются принципы согласования паттернов?
35. Что может включать в себя паттерн наблюдаемого результата?
36. Что такое корреляция паттерна согласования?
37. Как получить корреляцию паттерна согласования?
38. В чем заключаются преимущества и недостатки согласования паттернов?

## Тестовые задания к теме 2

1. Степень, в которой возможны обоснованные теоретические выводы на основе применяющейся в исследовании операционализации, называется
  - а) мысленный эксперимент
  - б) алгебраическая модель
  - в) внутренняя валидность
  - г) валидность выводов
  - д) валидность конструкторов
2. Валидность конструкторов имеет отношение
  - а) к распределению
  - б) к произведению
  - в) к отклонению
  - г) к параметру
  - д) к обобщению
3. Если удастся показать, что измерения, которые теоретически считаются сильно взаимосвязанными, на практике сильно взаимосвязаны, то такая валидность называется
  - а) внутренняя валидность
  - б) валидность выводов
  - в) валидность конструкторов
  - г) конвергентная валидность
  - д) дискриминантная валидность
4. Теория относительно того, каким образом методы измерения в эксперименте соотносятся друг с другом и с другими теоретическими компонентами, называется
  - а) теоретическая валидность
  - б) наблюдаемая валидность
  - в) теоретический паттерн
  - г) наблюдаемый паттерн
  - д) выборочное распределение
5. Разработать конкретные и осуществимые методы, позволяющие практикующим исследователям оценивать паттерны соответствия, т. е. оценивать конструктивную валидность своих исследований
  - а) в принципе невозможно
  - б) в принципе возможно
  - в) в принципе согласен
  - г) в принципе не важно
  - д) затрудняюсь ответить
6. Когда концепция или понятие транслируется в реальное функционирование, это называется
  - а) рандомизация
  - б) триангуляция
  - в) схематизация

- г) популяция  
д) операционализация
7. Какой термин характеризуется наиболее широким значением:  
а) прогностическая валидность  
б) конвергентная валидность  
в) дискриминантная валидность  
г) одновременная валидность  
д) валидность конструкта
8. То, насколько хорошо операционализация отражает конструкт, определяется  
а) прогностической валидностью  
б) конвергентной валидностью  
в) дискриминантной валидностью  
г) одновременной валидностью  
д) валидностью транслирования
9. То, насколько хорошо удалось транслирование конструкта в операционализацию, определяется  
а) лицевой валидностью  
б) конвергентной валидностью  
в) дискриминантной валидностью  
г) прогностической валидностью  
д) валидностью транслирования
10. На проверку операционализации по отношению к релевантной содержательной области конструкта направлена  
а) содержательная валидность  
б) валидность выводов  
в) валидность конструктов  
г) конвергентная валидность  
д) дискриминантная валидность
11. Проверить функционирование операционализации в отношении какого-то критерия позволяет  
а) содержательная валидность  
б) валидность выводов  
в) валидность конструктов  
г) лицевая валидность  
д) критериальная валидность
12. На оценку способности нашей операционализации проводить различие между теми группами, между которыми она должна теоретически быть способна проводить различие, направлена  
а) прогностическая валидность  
б) конвергентная валидность

- в) дискриминантная валидность
- г) одновременная валидность
- д) валидность конструкта

13. Вещи не являются или черными, или белыми – все понятия более или менее соотносятся между собой, с точки зрения

- а) фрейдизма
- б) универсализма
- в) дефиниционализма
- г) реляционализм
- д) операционализма

14. Отвергает идею о том, что в качестве основы для определения конструкторов мы можем полагаться на операциональные формулировки

- а) фрейдизм
- б) универсализм
- в) дефиниционализм
- г) реляционализм
- д) операционализм

15. Для того, чтобы оценить степень, в которой любые два измерения соотносятся друг с другом, обычно применяется

- а) стандартная ошибка
- б) стандартное отклонение
- в) теоретический паттерн
- г) коэффициент дисперсии
- д) коэффициент корреляции

16. Насколько велика должна быть корреляция, чтобы служить подтверждением конвергентности, и насколько мала она должна быть, чтобы служить подтверждением дискриминантности?

- а) дискриминантная корреляция должна быть больше, чем конвергентная корреляция
- б) конвергентная корреляция должна быть больше, чем дискриминантная корреляция
- в) дискриминантная корреляция должна быть равна конвергентной корреляции
- г) в принципе неважно
- д) затрудняюсь ответить

17. Для подтверждения конвергентной валидности, необходимо показать, что те измерения, которые должны быть взаимосвязаны

- а) реально надежны
- б) реально не взаимосвязаны
- в) реально взаимосвязаны



- г) в принципе неважно  
д) затрудняюсь ответить
18. Для подтверждения дискриминантной валидности, необходимо показать, что те измерения, которые не должны быть взаимосвязаны
- а) реально надежны  
б) реально не взаимосвязаны  
в) реально взаимосвязаны  
г) в принципе неважно  
д) затрудняюсь ответить
19. Если дискриминантная валидность присутствует, то взаимосвязь между измерениями двух различных конструктов должна быть
- а) кратной 3,14  
б) кратной целому числу  
в) надежной  
г) очень большой  
д) очень малой
20. Свидетельствует ли паттерн, подтверждающий и конвергентную валидность, и дискриминантную валидность, о том, что измерения данного конструкта действительно измеряют именно этот конструкт:
- а) разумеется, да  
б) разумеется, нет  
в) строгие правила отсутствуют  
г) в принципе неважно  
д) затрудняюсь ответить
21. Вместо того, чтобы рассматривать конвергентную валидность и дискриминантную валидность как качественно различные, подход с точки зрения соответствия паттернов рассматривает их
- а) как различия в форме  
б) как различия в содержании  
в) как различия в степени  
г) как различия в принципе  
д) как различия в методе
22. Слово дооперациональный означает
- а) до транслирования измерения в конструкт  
б) после транслирования измерения в конструкт  
в) до транслирования конструкта в измерения  
г) после транслирования конструкта в измерения  
д) поддающийся операционализации
23. В отношении независимой переменной или причины в эксперименте возможен
- а) монооперациональный уклон  
б) мультиоперациональный уклон  
в) монометодический уклон

- г) мультиметодический уклон  
д) сочетание уклонов
24. Относится к измерениям и к наблюдениям, но не относится к причинам
- а) монооперациональный уклон  
б) мультиоперациональный уклон  
в) монометодический уклон  
г) мультиметодический уклон  
д) сочетание обоих уклонов
25. Ограниченную обобщаемость конструктов можно назвать
- а) преднамеренными последствиями  
б) непреднамеренными последствиями  
в) неизбежными последствиями  
г) необходимыми причинами  
д) недостаточными причинами
26. Матрица или таблица корреляций, устроенная таким образом, чтобы облегчить интерпретацию оценок валидности конструктов, называется
- а) номологическая сеть  
б) линейная матрица  
в) симметричная матрица  
г) выборочное распределение  
д) матрица мультисвойств-мультиметодов
27. Корреляции внутри матрицы мультисвойств-мультиметодов располагаются
- а) с точки зрения применявшихся методов  
б) с точки зрения понятий  
в) с точки зрения понятий и применявшихся методов  
г) в принципе неважно  
д) затрудняюсь ответить
28. По своей сути, матрица мультисвойств-мультиметодов – это
- а) транспонированная матрица  
б) обратная матрица  
в) единичная матрица  
г) корреляционная матрица  
д) выборочное распределение
29. В матрице мультисвойств-мультиметодов все корреляции, для которых использованы одинаковые методы измерения, образуют
- а) диагонали валидности  
б) треугольники гетеросвойств-монометодов  
в) треугольники гетеросвойств-гетерометодов  
г) блоки монометодов  
д) блоки гетерометодов
30. Интерпретация матрицы мультисвойств-мультиметодов требует от исследователя способности
- а) способности фантазировать  
б) способности измерять

- в) способности рассуждать
- г) способности вычислять
- д) способности взаимодействовать

31. В матрице мультисвойств-мультиметодов коэффициенты на диагонали надежности должны быть

- а) самыми маленькими
- б) самыми большими
- в) не самыми маленькими
- г) не самыми большими
- д) в принципе неважно

32. В матрице мультисвойств-мультиметодов коэффициенты на диагоналях валидности должны быть

- а) отличаться от нуля
- б) равны нулю
- в) равны между собой
- г) неравны между собой
- д) в принципе неважно

33. В матрице мультисвойств-мультиметодов коэффициент валидности должен быть

- а) равен значениям в одном с ним столбце
- б) меньше, чем значения в одном с ним столбце
- в) больше, чем значения в одном с ним столбце
- г) равен нулю
- д) в принципе неважно

34. Матрица мультисвойств-мультиметодов требует, чтобы каждое свойство измерялось

- а) любым методом
- б) каждым методом
- в) новым методом
- г) проверенным методом
- д) передовым методом

35. Для определения конвергентной валидности и дискриминантной валидности, фактор методов

- а) требуется
- б) не требуется
- в) отсутствует
- г) допускается
- д) в принципе неважно

36. Конвергентная валидность и дискриминантная валидность – это

- а) конструкты
- б) методы
- в) параметры
- г) принципы

- д) в принципе неважно
37. Когда измеряются разные конструкты, то кросскорреляции между конструктами должны быть
- а) равными
  - б) разными
  - в) большими
  - г) малыми
  - д) надежными
38. Идея об одновременном рассмотрении паттерна конвергенции и дискриминации аналогична по своей цели
- а) матрице корреляций
  - б) матрице мультисвойств-мультиметодов
  - в) доверительному интервалу
  - г) выборочному распределению
  - д) номологической сети
39. Любое упорядочение объектов или единиц называется
- а) параметром
  - б) статистикой
  - в) выборкой
  - г) генеральной совокупностью
  - д) паттерном
40. Основой для генерирования паттернов служит
- а) практика
  - б) наблюдение
  - в) эксперимент
  - г) теория
  - д) измерение
41. Попытка связать теоретический паттерн и наблюдаемый или операциональный паттерн предполагает
- а) комбинацию паттернов
  - б) конструирование паттернов
  - в) соответствие паттернов
  - г) исключение паттернов
  - д) дополнение паттернов
42. Транслирование идей в некий конкретизируемый теоретический паттерн называется
- а) конкретизацией
  - б) трансакцией
  - в) популяцией
  - г) рационализацией
  - д) концептуализацией
43. Гипотеза о том, чего следует ждать от полученных данных, называется
- а) наглядный паттерн
  - б) фактический паттерн
  - в) теоретический паттерн

- г) наблюдаемый паттерн  
д) психологический паттерн
44. Данные, применяемые для исследования теоретической модели, образуют
- а) наглядный паттерн  
б) фактический паттерн  
в) теоретический паттерн  
г) наблюдаемый паттерн  
д) психологический паттерн
45. Согласование паттернов позволяет рассматривать конвергенцию и дискриминацию
- а) в виде матрицы корреляций  
б) в виде колоколообразной кривой  
в) в виде гистограммы  
г) в виде доверительного интервала  
д) у виде континуума
46. На вопрос о том, удалось ли продемонстрировать взаимоотношение между тем, что мы теоретически полагаем взаимосвязанным с нашим измерением, и тем, с чем наше измерение действительно взаимосвязано, позволяет ответить
- а) величина корреляции согласованности ответов  
б) величина корреляции согласованности параметров  
в) величина корреляции согласованности переменных  
г) величина корреляции согласованности измерений  
д) величина корреляции согласованности паттернов

### ТЕМА 3. НАДЁЖНОСТЬ

Надежность предполагает качество измерения. В обиходном смысле, надежность представляет собой «согласованность» или «повторимость» измерений. Прежде, чем сформулировать точное определение надежности, необходимо проделать предварительную работу.

- ❖ Прежде всего, необходимо узнать об основаниях надежности, или о *теории измерения истинного значения*.
- ❖ Необходимо понять различные *типы ошибок измерения*, потому что ошибки в измерениях играют важную роль в уменьшении надежности.
- ❖ После этого можно рассматривать основы *теории надежности*, включая точное определение того, что такое надежность. оказывается, что надежность невозможно вычислить – её можно только оценивать. По этой причине, существуют множество различных *типов надёжности*, в каждом из которых предлагаются множество способов оценки надежности.
- ❖ В завершение, важно интегрировать идею надежности с другим главным критерием качества измерения – валидностью, и прийти

к пониманию взаимосвязей между надежностью и валидностью в измерении.

## ТЕОРИЯ ИСТИННОГО ЗНАЧЕНИЯ

**Теория истинного значения** – это теория измерения. Подобно любым теориям, следует признать, что она не доказана – она сформулирована в качестве модели того, каким образом функционирует окружающий мир. Подобно очень многим влиятельным моделям, теория истинного значения весьма проста. Действительно, в теории истинного значения утверждается, что каждое измерение представляет собой сумму двух слагаемых:

$$\begin{array}{ccccccc} X & = & T & + & e \\ \text{Наблюдаемое} & & \text{Истинная} & & \text{Случайная} \\ \text{число} & & \text{способность} & & \text{ошибка} \end{array}$$

**Истинная способность** (или истинный уровень) испытуемого, которая измеряется;

**случайная ошибка.**

Измерение можно наблюдать – баллы по тесту, значение по методике измерения самоуважения, шкальные значения веса человека. Того, что находится в правой части данного уравнения, мы наблюдать не можем. Мы предполагаем, что в правой части уравнения есть два слагаемых.

Простое уравнение

$$X = T + e_x$$

имеет параллельное уравнение на уровне дисперсии или вариабельности измерения:

$$\text{var} (X) = \text{var} (T) + \text{var} (e_x)$$

Говоря более человеческим языком, это означает, что вариабельность измерения представляет собой сумму вариабельности истинного значения и вариабельности случайной ошибки. Понимание этого имеет важное значение для рассмотрения более комплексных моделей коррекции ошибок измерений.

Почему важна теория истинного значения?

- ❖ Во-первых, это простая, но убедительная модель измерения. Она напоминает нам о том, что в большинстве измерений имеется компонент ошибки.

- ❖ Во-вторых, теория истинного значения выступает фундаментом теории надежности.
  - измерение, лишенное случайной ошибки (т. е. только истинное значение) абсолютно надежно;
  - измерение, лишенное истинного значения (т. е. только случайная ошибка) имеет нулевую надежность.

- ❖ В-третьих, теорию истинного значения можно применять в компьютерных имитациях в качестве основы для генерирования «наблюдаемых» значений, обладающих некоторыми известными свойствами.

Следует знать, что модель истинного значения – это не единственная существующая модель измерения. Теоретики измерений продолжают создавать всё более и более сложные модели, которые, по их мнению, лучше отражают реальность. Однако, эти модели настолько усложнены, что мы не будем их рассматривать. В любом случае, теория истинного значения должна дать вам представление о том, почему вообще важны модели измерений и каким образом их можно применять в качестве основания для формулирования главных идей научного исследования.

## ОШИБКА ИЗМЕРЕНИЯ

Теория истинного значения служит хорошей простой моделью измерения, но она не всегда точным образом отражает действительность. Конкретно, она предполагает, что любое наблюдение складывается из истинного значения плюс величина случайной ошибки. Но так ли это? Что, если любые ошибки не случайны? Возможно ли, что некоторые ошибки оказываются систематическими, что они присущи всем или почти всем испытуемым в группе? Один из способов решить этот вопрос состоит в пересмотре простой модели истинного значения путем разделения компонента ошибки на два подкомпонента: на **случайную ошибку** и на **систематическую ошибку**. Рассмотрим различие между этими двумя типами ошибки и попытаемся диагностировать их воздействие на научное исследование.

$$X = T + e$$

*Два компонента:*

$e_r$  \* *случайная ошибка*

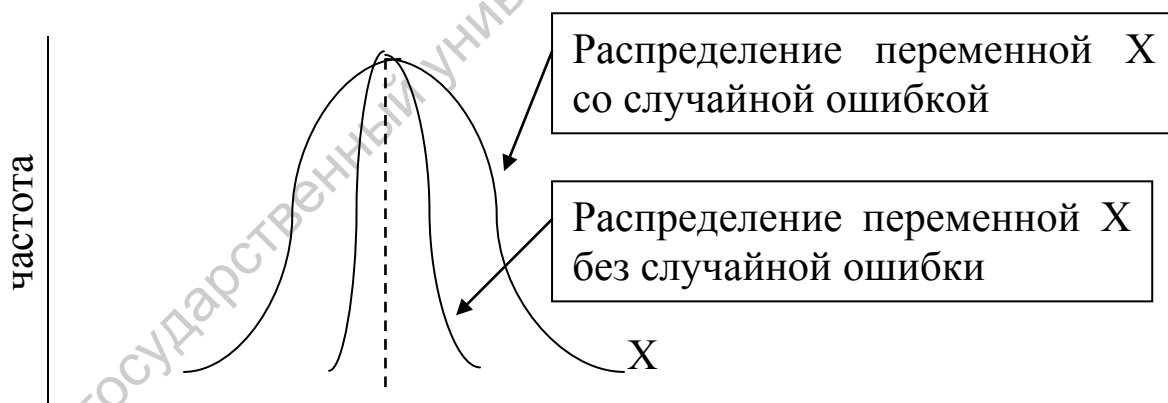
$e_s$  \* *систематическая ошибка*

$$X = T + e_r + e_s$$

### Что такое случайная ошибка?

Случайная ошибка вызывается любыми факторами, которые случайным образом воздействуют на измерение переменной в выборке. К примеру, настроение испытуемого может улучшать или ухудшать выполнение тестовых заданий. У одних испытуемых может быть хорошее настроение, а у других испытуемых возможна депрессия. Если настроение влияет на проведение измерения, то оно способно искусственным образом понижать баллы, в которых фиксируется выполнение тестовых заданий, либо повышать эти баллы у других испытуемых. Важной особенностью случайной ошибки оказывается то, что она не обладает каким-либо устойчивым воздействием на всю выборку. Вместо этого, она случайным образом выталкивает наблюдаемые значения то вверх, то вниз. Это означает, что если бы мы увидели все случайные ошибки в распределении, то их сумма оказалась бы равной нулю – положительных ошибок оказалось бы столько же, сколько отрицательных.

Важная особенность случайной ошибки состоит в том, что она придает данным вариабельность, но на среднем функционировании группы испытуемых она не отражается. По этой причине, случайная ошибка иногда считается шумом.



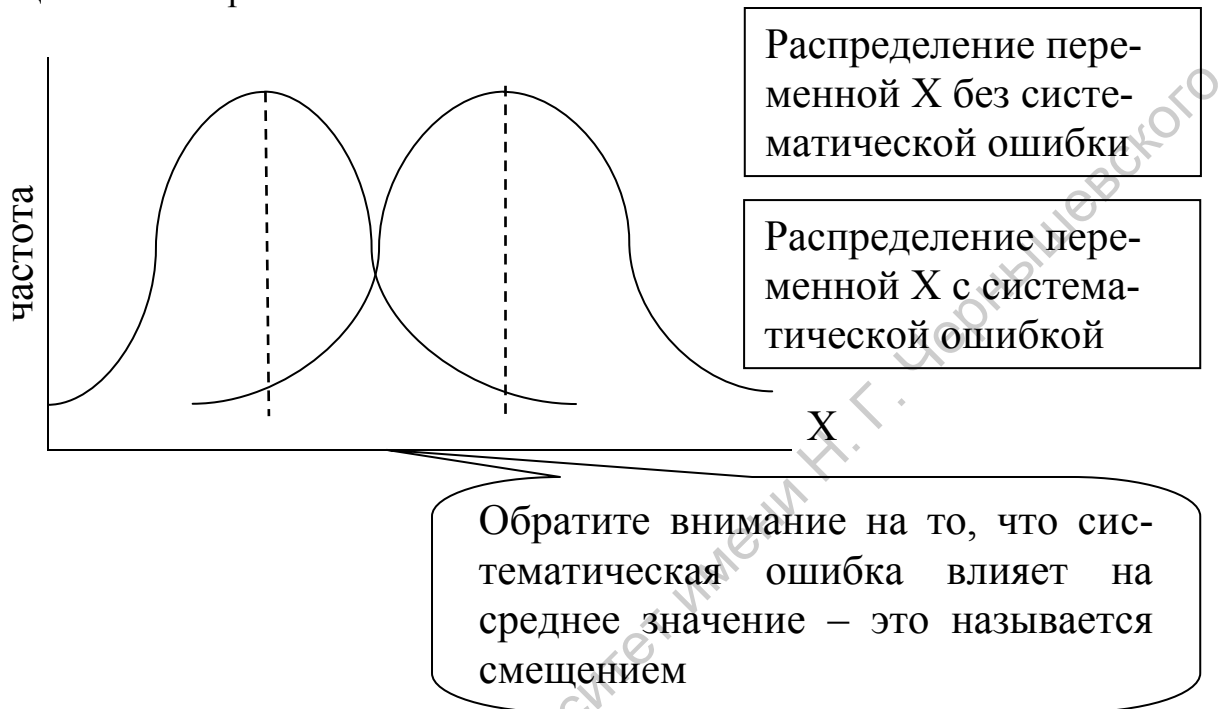
Обратите внимание на то, что случайная ошибка не влияет на среднее значение, а только на вариабельность вокруг среднего значения

### Что такое систематическая ошибка?

Систематическая ошибка вызывается любыми факторами, которые систематическим образом воздействуют на измерение переменной в выборке. К примеру, если за окнами аудитории, в которой студенты проходят тестирование, совершается громкое уличное движение, то этот шум повлияет на бал-



лы всех испытуемых – в этом случае, шум оказывает систематическое воздействие. В отличие от случайной ошибки, систематические ошибки имеют выраженную тенденцию – или положительную тенденцию, или отрицательную тенденцию. По этой причине, систематические ошибки называются **смещением измерения**.



### Уменьшение ошибки измерения

Итак, каким образом мы можем уменьшить ошибки измерения, и случайные, и систематические?

- ❖ Один из выходов заключается в проведении пилотного тестирования измерительного инструментария, получении обратной связи от испытуемых в отношении того, насколько легко или насколько трудно прошло измерение, а также информации о том, каким образом обстановка тестирования повлияла на результаты.
- ❖ Второй выход представляется в том направлении, что, если проводится сбор данных измерения с привлечением специальных людей (интервьюеров или наблюдателей), то нужно позаботиться о проведении тренинга для таких людей с тем, чтобы они нечаянно не внесли в данные ошибок.
- ❖ Третий выход состоит в том, что сбор данных должен подвергаться тщательному двойному контролю. Ввод данных в компьютер должен быть проверен дважды. Это означает, что данные вводятся два раза, во второй раз машина ввода данных проверяет, ввели ли вы точно такие же данные, которые вы вводили первый раз.
- ❖ Четвертый выход состоит в применении статистической процедуры коррекции ошибки измерения. Эти процедуры могут быть довольно простые формулы, применяемые непосредственно к водимым данным,

а могут быть очень сложными процедурами моделирования тех эффектов, которые оказывают ошибки.

- ❖ Пятый выход – самый лучший вариант заключается в том, чтобы применить несколько рассмотренных вариантов. В особенности, если различные измерения характеризуются не одинаковыми систематическими ошибками, то нужно уметь проводить **триангуляцию** различных измерений.

## ТЕОРИЯ НАДЕЖНОСТИ

Что такое **надежность**? Мы слышим этот термин применительно к различным исследовательским контекстам, но что он означает в действительности? Если задуматься над употреблением слова «надежный» в повседневной речи, то можно найти подсказку. Например, о машине часто говорят, что она надежная: «У меня надежная машина». Или журналисты говорят об «информации из надежного источника». В обоих случаях слово «надежный» значит «заслуживающий доверия». В научном исследовании термин «надежность» тоже в общем смысле означает нечто, заслуживающее доверия, но такое определение не совсем точное. Что значит, что в научном контексте измерение или наблюдение заслуживает доверия? Выражение «заслуживает доверия» не совсем точное, потому что его легко спутать с идеей о валидном измерении. Разумеется, когда мы говорим о заслуживающем доверия измерении, мы имеем в виду, что это измерение и надежно, и валидно. Поэтому при формулировании определения надежности следует проявить точность.

**В научном исследовании термин «надежность» означает «допускающий возможность повторения» или «согласованный». Измерение считается надежным, если оно снова и снова приводит к одному и тому же результату (если то, что мы измеряем, остается без изменений!)**

Рассмотрим более подробно, что означают слова о том, что измерение «допускает повторение» или «согласовано». Начнем с определения того измерения, которое мы произвольно обозначили  $X$ . Это может быть балл испытуемого по арифметическому тесту, либо это может быть показатель серьезности заболевания. Это то значение (количественное либо иное), которое мы наблюдаем в исследовании.

$X_1$   
↓

$X_2$   
↓

$$\begin{array}{ccc}
 T + e_1 & & T + e_2 \\
 \uparrow & & \uparrow \\
 \hline
 \end{array}$$

Для того, чтобы увидеть – насколько наблюдение допускает возможность повторения, либо насколько оно согласовано, его можно измерить дважды. Первое и второе наблюдения одного и того же измерения обозначим индексами. Если допустить, что то, что мы измеряем, не изменилось за период между первым и вторым измерениями, тогда начинаешь понимать, каким образом достигается надежность. Когда мы наблюдаем количественное значение того, что мы измеряем, мы обычно рассматриваем это значение состоящим из двух компонентов: «истинное» значение или действительный уровень для данного испытуемого по этому измерению, и «ошибка» при его измерении.

Важно иметь в виду, что мы наблюдаем значение  $X$  – мы никогда в действительности не видим ни истинного значения ( $T$ ), ни значения ошибки ( $e$ ). например, при выполнении математического теста, студент получил 85 баллов. Это значение, которое мы наблюдаем,  $X$  равен 85. Но в действительности может оказаться, что студент лучше знает математику, чем это показывает математический тест. Допустим, что истинные математические способности этого студента равны 89 (т. е.,  $T = 89$ ). Это означает, что величина ошибки для данного студента равна  $-4$ . Что это значит? Возможно, что у студента выдался трудный день, возможно, что студент не завтракал, возможно, студент поссорился, возможно, во время выполнения тестовых заданий студент отвлекся. Перечисленные и подобные им факторы способствуют ошибочности измерения, в результате наблюдаемые способности испытуемого представляются менее выраженными, по сравнению с истинными или действительными способностями.

Вернемся к надежности. Если наше измерение  $X$  надежно, то, если мы измерим или пронаблюдаем его дважды на одних и тех же испытуемых, то баллы в обоих случаях существенно совпадут. Но почему баллы должны совпадать? Из рисунка видно, что единственная вещь, которая является общей для двух наблюдений, это их истинные значения или  $T$ . Откуда это известно? Потому что значения ошибок ( $e_1$  и  $e_2$ ) имеют неодинаковые нижние индексы, указывая на то, что это различные значения. Однако, символ истинного значения  $T$  одинаков для обоих наблюдений. Что это означает? Это означает, что два наблюдаемых значения  $X_1$  и  $X_2$  связаны лишь в той степени, в которой в обоих наблюдениях присутствует истинное значение. Следует помнить о том, что величина ошибки считается случайной. Иногда ошибки приводят к тому, что результаты теста оказываются лучше истинных способностей испытуемого, а иногда баллы по тесту оказываются хуже истинных способностей. Однако, истинное значение – истинные способности испытуемого – остаются одними и теми же при обоих измерениях (разумеется, если

предположить, что за временной период между двумя измерениями способности испытуемого не изменились).

С учетом всего вышесказанного, мы можем сформулировать более точное определение надежности. Надежность представляет собой **соотношение** или дробь. Для непосвященных дробь может формулироваться таким образом:

### **истинный уровень надежности** **полная величина измерения**

Надежность можно себе представить в качестве пропорции «истины» в измерении. Мы не говорим сейчас о надежности измерения в отношении индивидуального испытуемого – надежность представляет собой характеристику измерения, применимую ко всем испытуемым. Итак, чтобы подобраться поближе к более формальному определению, следует переформулировать вышеизложенное определение на языке комплекса наблюдений. Простейший способ сделать это состоит в рассмотрении дисперсии баллов. Вспомним, что дисперсия представляет собой измерение растекания или распределения *комплекса* значений. Теперь мы можем дать следующее определение:

### **дисперсия истинного значения** **дисперсия измерения**

Слегка изменив техническую терминологию за счет введение алгебраических символов названий переменных и сокращенного обозначения дисперсии, получаем:

$$\frac{\text{var (T)}}{\text{var (X)}}$$

Настал решающий момент. Если взглянуть на вышеприведенное уравнение, то следует признать, что знаменатель дроби можно легко рассчитать – он есть дисперсия наблюдаемых нами значений. (Дисперсия вычисляется как сумма от квадрата отклонений значений от среднего, поделенная на количество баллов, помните?) Но как рассчитать дисперсию истинных значений? Истинных значений мы не видим – мы видим только значения X! А если мы не сможем рассчитать дисперсию истинных значений, то мы не сможем вычислить соотношение, это означает, что мы *не сможем рассчитать надежность!* Это всем ясно? Идея состоит в том, что...

**мы не можем рассчитать надежность,  
потому что мы не можем рассчитать**

## дисперсию истинных значений

Отлично. Так что же нам делать с этим? Если мы не можем вычислить надежность, то, вероятно, лучшее, что мы можем сделать – это *оценить* её. Возможно, нам удастся получить оценку вариабельности истинных значений. Как это делается? Помните про наши два измерения,  $X_1$  и  $X_2$ ? Исходя из теории истинных значений, предположим, что эти два наблюдения связаны друг другом в той степени, в какой истинное значение является общим для них обоих. Итак, рассчитаем корреляцию между  $X_1$  и  $X_2$ . Простая формула для расчета корреляции имеет следующий вид:

$$\frac{\text{ковариация } (X_1, X_2)}{\text{CO } (X_1) * \text{CO } (X_2)}$$

где CO означает «стандартное отклонение» (которое равно квадратному корню из дисперсии). Если приглядеться к этой записи повнимательнее, то можно увидеть, что ковариация, которая измеряет всего лишь «совместную» дисперсию для обоих измерений, должна служить показателем вариабельности истинных значений, потому что истинные значения  $X_1$  и  $X_2$  – это единственное, что есть общего у двух наблюдений! Итак, числитель – это, по сути, оценка дисперсии **var (T)** в нашем контексте. А поскольку в знаменателе находится произведение стандартных отклонений двух наблюдений одного и того же измерения в разные моменты времени, то следует ожидать, что эти два значения будут одинаковыми (измеряется одно и то же), и что их квадраты тоже будут одинаковыми. Однако, квадрат стандартного отклонения – это то же самое, что дисперсия измерения. Поэтому, в знаменателе дроби находится дисперсия измерения (или **var (X)**). Если вы внимательно читали этот абзац, то должны были заметить, что корреляция между двумя наблюдениями одного и того же измерения представляет собой оценку надежности.

Настала пора сделать некоторые выводы. Мы уже видели, что надежность рассчитать невозможно, потому что невозможно измерить компонент истинного значения наблюдения. Мы также видели, что компонент истинного значения можно *оценить* через ковариацию между двумя наблюдениями одного и того же измерения. Имея всё это в виду, надежность можно ценить через корреляцию между двумя наблюдениями одного и того же измерения. Оказывается, что существует несколько способов вычисления корреляции надежности. Они будут рассмотрены в следующем разделе.

Здесь следует рассмотреть ещё один вопрос. Насколько велика оценка надежности? Чтобы ответить на этот вопрос, вернемся к уже известной формуле:

$$\frac{\text{var } (T)}{\text{var } (X)}$$

и вспомним, что поскольку

$$X = T + e$$

то в знаменателе можно написать:

$$\frac{\text{var}(T)}{\text{var}(T) + \text{var}(e)}$$

Путем такой несложной подстановки можно легко определить интервал оценки надежности. Если измерение абсолютно надежно, то ошибка измерения отсутствует – всё, что наблюдается, оказывается истинным значением. Поэтому, в случае абсолютно надежного измерения, формула имеет следующий вид:

$$\frac{\text{var}(T)}{\text{var}(T)}$$

а надежность равна 1. Теперь, если измерение абсолютно ненадежно, то в нем отсутствует истинное значение – измерение сплошь ошибочно. В этом случае формула принимает следующий вид:

$$\frac{0}{\text{var}(e)}$$

а надежность равна 0. Значит, надежность всегда находится в интервале от 0 до 1. Величина оценки надежности указывает нам на пропорцию вариабельности измерения, применительно к истинному значению. Надежность, равная 0,5, означает, что около половины дисперсии наблюдаемого значения характеризуется истинностью, а половина характеризуется ошибочностью. Надежность, равная 0,8, означает, что вариабельность на 80% характеризуется истинностью, а на 20% характеризуется ошибочностью. И т. п.

## ТИПЫ НАДЕЖНОСТИ

Из теории надежности мы узнали, что невозможно точно вычислить надежность. Вместо этого, надежность приходится оценивать, а такие попытки всегда неабсолютны. Познакомимся с основными способами оценки надежности, с их сильными и слабыми сторонами.

Существуют четыре общих класса оценок надежности, все они оценивают надежность по-разному:

- ❖ **Надежность внешних наблюдений.** Используется при оценке степени, в которой согласуются между собой оценки одного и того же явления различными наблюдателями.
- ❖ **Тест-ретестовая надежность.** Используется при оценке согласованности измерения в различные моменты времени.
- ❖ **Надежность параллельных форм.** Используется при оценке согласованности результатов двух тестов, сконструированных одинаковым образом исходя из одной и той же содержательной области.
- ❖ **Внутренне согласованная надежность.** Используется при оценке согласованности результатов по пунктам одного и того же тестового задания.

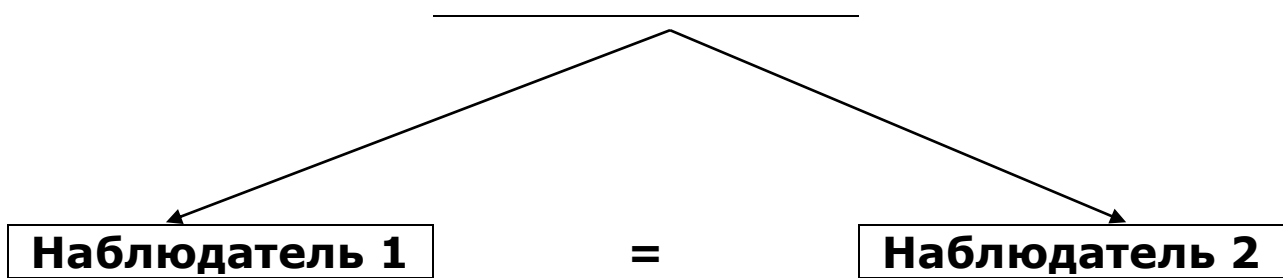
### **Надежность внешних наблюдений**

Каждый раз, когда частью процедуры измерения становятся люди, необходимо позаботиться о том, чтобы полученные результаты отличались надежностью или согласованностью. Людям свойственна несогласованность. Люди легко отвлекаются. Люди устают от повторного выполнения заданий. Люди мечтают. Люди не понимают.

Итак, каким образом определяется то, согласуются ли между собой наблюдения двух человек? Вероятно, за исследовательским контекстом измерения следует обеспечить надежность внешних наблюдений. В конечном счете, если пытаться обеспечивать надежность на основе данных проведенного наблюдения, а надежность оказывается низкой, то у исследователя возникнет шок. Вероятно, лучший способ обеспечения надежности внешних наблюдений заключается в проведении постороннего исследования либо пилотного исследования. А если исследование длится долгое время, то подтверждение время от времени надежности внешних наблюдений позволяет убедиться в том, что внешние наблюдатели не изменились.

Существуют два главных способа действительно образом обеспечить надежность внешних наблюдений. Если проводимое измерение предполагает категории – задача наблюдателей заключается в проверке того, к какой категории будет отнесен какой объект – то можно рассчитать процент согласия между внешними наблюдателями. Например, допустим, что у нас имеются 100 наблюдений, которые внешние наблюдатели должны рассортировать по трем категориям. Допустим, что 86 наблюдений из 100 внешние наблюдатели распределили по категориям одинаковым образом. В этом случае процент согласия внешних наблюдателей составляет 86%. Согласен, что это грубый метод, однако, он действительно дает нам представление о том, насколько внешние наблюдатели согласны друг с другом, и является действенным при любом количестве категорий, используемых для наблюдения.

<p><b>Объект или явление</b></p>
--------------------------------------



Другой главный способ оценки надежности внешних наблюдений применим в тех случаях, когда измерение является непрерывным. В таких случаях требуется рассчитывать корреляции между оценками внешних наблюдателей. Например, наблюдатели оценивают аудиторную активность студентов о шкале от 1 до 7. Они могут сообщать о своих оценках через определенные промежутки времени (например, каждые 30 секунд). Корреляция между этими оценками представляет собой оценку надежности или согласованности внешних наблюдателей.

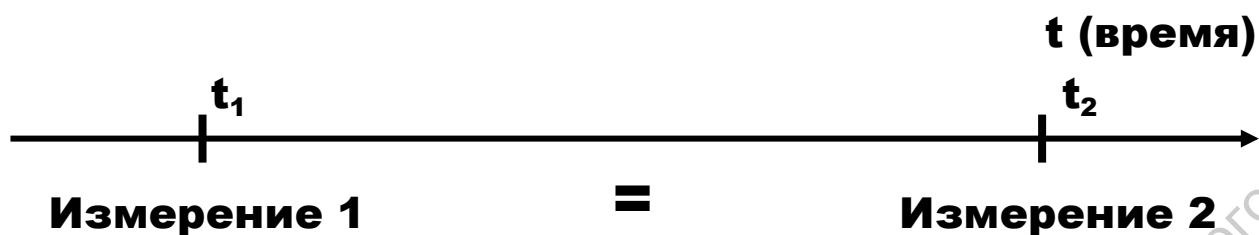
Такой тип надежности можно назвать «калибровкой» внешних наблюдателей. Существуют другие факторы обеспечения надежности внешних наблюдений, которые не обязательно требуется рассчитывать. Например, экспериментатор проводит исследование в психиатрическом отделении, где сиделка каждое утро производит оценивание состояние каждого пациента по десятибалльной шкале. Разумеется, медсестры каждый день меняются, поэтому необходимо найти способ обеспечить сравнимость оценок, которые выставляет каждая медсестра. С этой целью производится еженедельная «калибровка» в виде планерки медсестер, на котором рассматриваются оценки, выставленные пациентам каждой медсестрой и обсуждается – почему каждая медсестра выставила именно такие оценки. При возникновении разногласий, медсестры пытаются сформулировать правила, в соответствии с которыми должна выставляться оценка «3» или оценка «4». И хотя эта процедура не является собственно оцениванием надежности, она значительно повышает надежность оценок внешних наблюдателей.

### **Тест-ретестовая надежность**

Оценка тест-ретестовой надежности производится тогда, когда одна и та же тестовая методика применяется к одной и той же (либо к аналогичной) выборке не менее двух раз. Данный подход подразумевает, что измеряемый конструкт не претерпевает значительных изменений в перерыве между двумя измерениями. Величина временного интервала между измерениями играет решающую роль. Известно, что при измерении одного и того же объекта дважды, корреляция между двумя наблюдениями отчасти будет зависеть от того, сколько времени прошло в промежутке между измерениями. Чем короче временной интервал, тем больше корреляция, чем дольше временной интервал, тем меньше корреляция. Причина заключается в том, что два измерения соотносятся со временем – чем короче время, тем больше сходства между факторами, вызывающими ошибку измерения. Поскольку корреляция

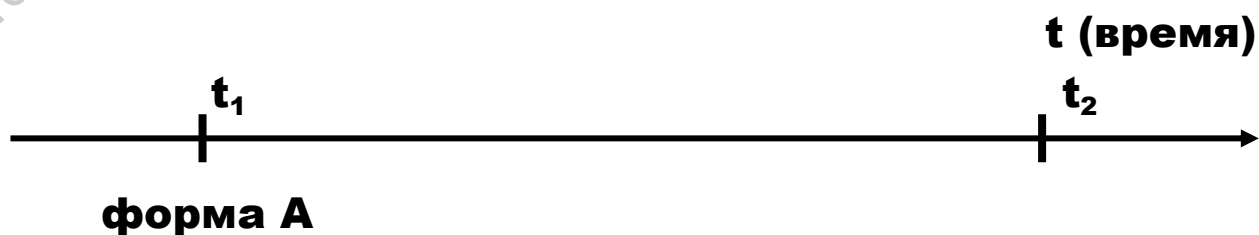


представляет собой тест-ретестовую оценку надежности, то в зависимости от временного интервала получаются существенно различающиеся оценки.



### Надежность параллельных форм

В случае надежности параллельных форм, прежде всего, необходимо создать две параллельные формы. Одним из способов создания параллельных форм является разработка большого комплекса вопросов, адресованных одному и тому же конструкту. Далее все вопросы делятся на две группы. Оба полученных опросника проводятся на одной и той же выборке испытуемых. Корреляция между двумя параллельными формами является оценкой надежности. Крупная проблема в связи с данным подходом состоит в необходимости генерирования большого количества пунктов опросника, относящихся к одному и тому же конструкту. Часто это довольно трудно. Более того, данный подход подразумевает, что случайное деление пополам образует параллельные или эквивалентные формы. Иногда результат бывает противоположным. Подход на основе параллельных форм весьма аналогичен расколотой пополам надежности, которая описана ниже. Главное отличие заключается в том, что параллельные формы конструируются таким образом, чтобы их можно было применять по отдельности и независимо друг от друга. Например, нас интересуют угрозы внутренней валидности, возникающие при экспериментальном тестировании. Если применить форму А перед экспериментальным тестированием, а форму Б после экспериментального тестирования, то проблема окажется сведенной к минимуму. Было бы еще лучше, если бы перед экспериментальным тестированием одни испытуемые в случайном порядке выполняли форму А, а другие испытуемые – форму Б, а после экспериментального тестирования менялись. При расколотой пополам надежности имеется некая тестовая методика, которая применяется целиком, а затем в случайном порядке создаются расколотые половины, по которым производится оценка надежности.





## форма Б

### Внутренне согласованная надежность

В случае оценки внутренне согласованной надежности, одна тестовая методика применяется к группе людей однократно. Далее выносится суждение относительно надежности тестовой методики на основании того, насколько одинаков результат по отдельным пунктам тестовой методики, отражающим один и тот же конструкт. Исследуется то, насколько согласованы результаты по различным пунктам одного и того же конструкта внутри тестовой методики.

Существуют несколько способов оценить внутреннюю согласованность измерения:

- ❖ Средняя корреляция между пунктами методики
- ❖ Средняя корреляция по общему баллу
- ❖ Расколота пополам надежность
- ❖ Альфа Кронбаха ( $\alpha$ )

### Средняя корреляция между пунктами методики

Средняя корреляция между пунктами методики рассчитывается для всех пунктов методики, которые сконструированы с целью измерения одного и того же конструкта. Сначала вычисляются корреляции между каждой парой пунктов опросника. Например, если у нас шесть пунктов опросника, тогда обчисляют 15 различных пар (т. е. 15 корреляций). Далее вычисляется среднее значение полученных коэффициентов корреляции. В приведенном примере значения коэффициентов корреляции располагаются в интервале от 0,84 до 0,95, а средняя корреляция составляет 0,90.

#### Средняя корреляция между пунктами методики

		Вопрос 1	Вопрос 2	Вопрос 3	Вопрос 4	Вопрос 5	Вопрос 6
И	Э	Вопрос 1	1,00				

<b>Вопрос 2</b>	0,89	1,00				
<b>Вопрос 3</b>	0,91	0,92	1,00			
<b>Вопрос 4</b>	0,88	0,93	0,95	1,00		
<b>Вопрос 5</b>	0,84	0,86	0,92	0,85	1,00	
<b>Вопрос 6</b>	0,88	0,91	0,95	0,87	0,85	1,00

↓  
0,90

### Средняя корреляция по общему баллу

В данном подходе также применяются корреляции между пунктами. В дополнение к этому, высчитывается общий балл по шести пунктам, который играет в анализе роль седьмой переменной. В нижней строке корреляционной матрице представлены шесть корреляций между значениями по каждому из шести пунктов с итоговым значением по всем шести пунктам. Значения коэффициентов корреляции находятся в интервале от 0,82 до 0,88, а среднее значение коэффициентов корреляции равно 0,85.

### Средняя корреляция по общему баллу

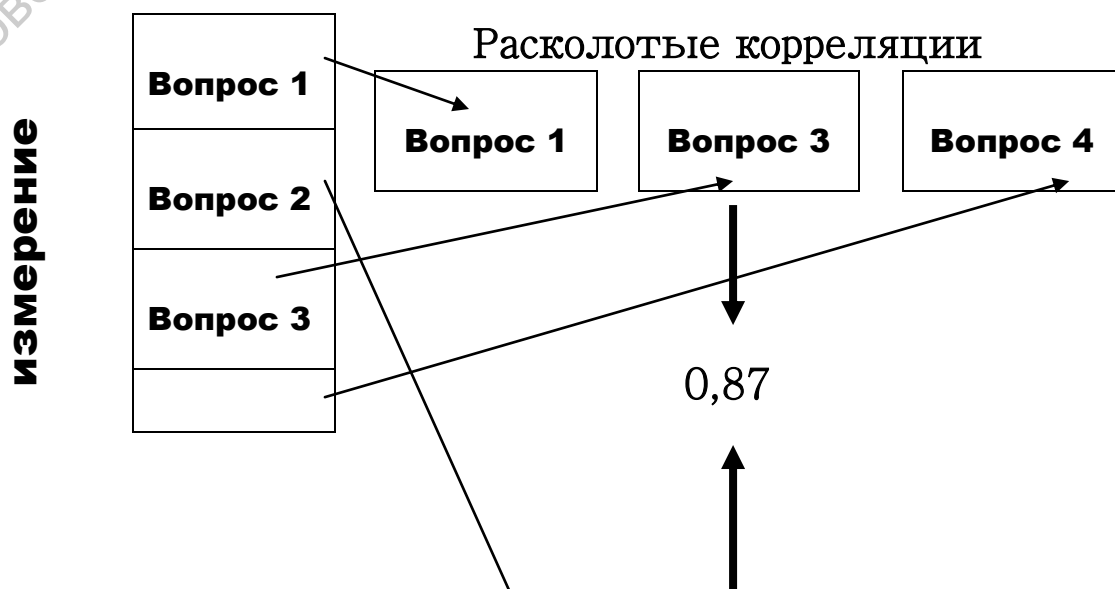
	<b>Вопрос 1</b>	<b>Вопрос 2</b>	<b>Вопрос 3</b>	<b>Вопрос 4</b>	<b>Вопрос 5</b>	<b>Вопрос 6</b>
<b>Вопрос 1</b>	1,00					
<b>Вопрос 2</b>	0,89	1,00				

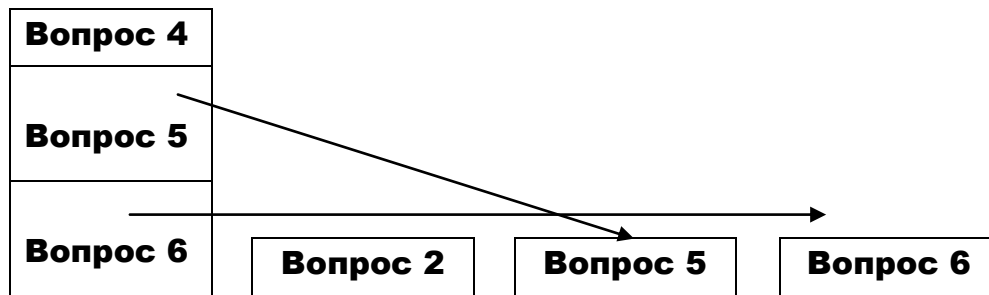
<b>Вопрос 3</b>	0,91	0,92	1,00				
<b>Вопрос 4</b>	0,88	0,93	0,95	1,00			
<b>Вопрос 5</b>	0,84	0,86	0,92	0,85	1,00		
<b>Вопрос 6</b>	0,88	0,91	0,95	0,87	0,85	1,00	
<b>Итого</b>	0,84	0,88	0,86	0,87	0,83	0,82	1,00

↓  
0,85

### Расколотая пополам надежность

В случае расколотой пополам надежности, все пункты методики делят случайным образом с целью измерить один и тот же конструкт в каждой полученной группе. Вся тестовая методика в полном объеме применяется к некой выборке испытуемых и рассчитывается итоговое значение для каждой случайным образом выделенной половины. Расколотая пополам надежность – это всего лишь корреляция между двумя итоговыми значениями. В нашем примере она равна 0,87.





### Альфа Кронбаха ( $\alpha$ )

Представьте себе, что мы вычисляем расколотую пополам надежность, а затем случайным образом перегруппировываем тестовые задания в другие расколотые половины и пересчитываем, и так продолжается до тех пор, пока мы не рассчитали все возможные при расколотых половинах оценки надежности. Альфа Кронбаха в математическом смысле эквивалентна среднему от всех возможных расколотых оценок, хотя в действительности она вычисляется иначе. Обратите внимание, что слова о вычислении «всех возможных расколотых оценок» не означают, что каждый раз измерение проводится на новой выборке! Для этого жизни не хватит. Напротив, высчитываются все расколотые оценки для той же самой выборки. Поскольку вся выборка выполнила каждое из шести тестовых заданий, то всё, что нам ещё остаётся сделать – выполнить компьютерный анализ случайных подкомплексов тестовых заданий и высчитать получающиеся в результате коэффициенты корреляции. На рисунке представлены несколько расколотых оценок для нашего примера, состоящего из шести тестовых заданий, которые обозначены РК (расколотые корреляции) с индексами. Следует иметь в виду, что, хотя альфа Кронбаха и эквивалентна среднему от всех возможных расколотых корреляций, но в действительности она никогда не вычисляется таким образом. Один умный математик (полагаю, это Кронбах!) предложил способ более быстрого получения математического эквивалента.

**измерение**

<b>Вопрос 1</b>
<b>Вопрос 2</b>
<b>Вопрос 3</b>

### Альфа Кронбаха ( $\alpha$ )

$$PK_1 = 0,87$$

$$PK_2 = 0,85$$

$$PK_3 = 0,91$$

$$\alpha = 0,85$$

<b>Вопрос 4</b>	$PK_4 = 0,83$
<b>Вопрос 5</b>	$PK_5 = 0,86$
	...
<b>Вопрос 6</b>	$PK_n = 0,85$

### Сравнение оценочных показателей надежности

У каждого из оценочных показателей надежности имеются собственные преимущества и недостатки. Надежность внешних наблюдений представляет один из наилучших способов оценки надежности, когда измерение проводится методом наблюдения. Однако, для такой оценки надежности требуется множество внешних наблюдателей. В качестве альтернативы можно рассмотреть корреляции оценок одного и того же единственного наблюдателя, повторенные дважды. Например, допустим, мы собрали видеозаписи взаимодействия матери и ребенка, и составили код оценивания видеозаписей с точки зрения того, как мать улыбается своему ребенку. Для определения надежности внешних наблюдений следует составить выборку из видеозаписей и предложить двоим внешним наблюдателям независимо друг от друга выставить кодированные оценки.

Для оценки тест-ретестовой надежности следует предложить одному и тому же внешнему наблюдателю выставить кодированные оценки дважды. Если у вас имеется команда внешних наблюдателей, и вы желаете убедиться в том, что они выставляют согласованные оценки, то в этом случае больше подходит оценка надежности внешних наблюдений. Если показатель оценки надежности внешних наблюдений окажется достаточно большим, тогда следует доверять независимым оценкам внешних наблюдателей.

Когда имеется лишь один внешний наблюдатель, и вы не желаете специально готовить других наблюдателей, то лучше применять тест-ретестовую оценку надежности. С другой стороны, в некоторых исследованиях имеет смысл сочетать оценку надежности внешних наблюдений с тест-ретестовую оценку надежности внешних наблюдателей.

Оценка надежности методом параллельных форм обычно применяется лишь в тех ситуациях, когда исследователь сознательно намеревается использовать обе параллельные формы в качестве альтернативных инструментов измерения одного и того же свойства или объекта. И оценка методом параллельных форм, и все методы оценки внутренней согласованности обладают одним главным ограничением – им требуются множество тестовых заданий, предназначенных для измерения одного и того же конструкта. В некоторых исследовательских контекстах добиться этого сравнительно легко (например, при конструировании математического теста), но в отношении некоторых субъективных конструктов это создает непреодолимую проблему. Если тестовых заданий действительно много, то чаще всего для оценки внутренней согласованности используется альфа Кронбаха.

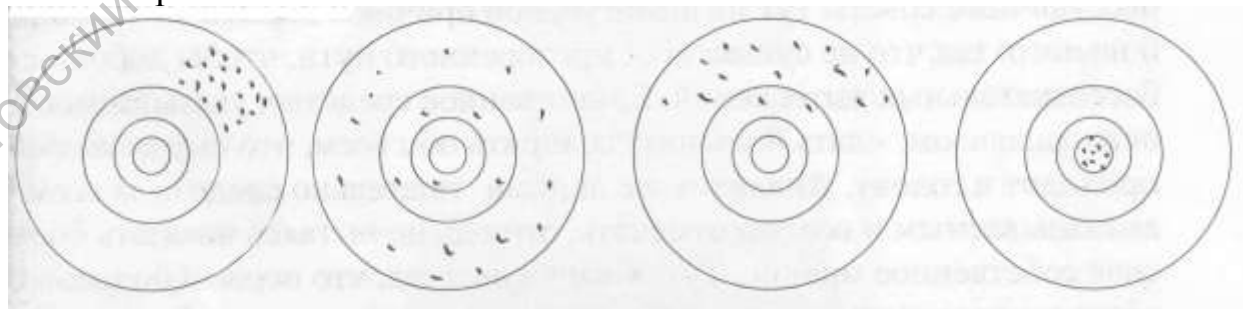
Тест-ретестовый оценочный показатель в особенности целесообразен в большинстве экспериментальных и квазиэкспериментальных планах исследований, где имеется контрольная группа, не испытывающая экспериментального воздействия. При проведении исследований по таким планам контрольная группа всегда измеряется дважды – до эксперимента и после эксперимента. Главная проблема применительно к данному подходу состоит в том, что у нас отсутствует какая-либо информация относительно надежности, пока мы не получим послеэкспериментальные данные и, если надежность оказывается низкой, исследователь испытывает большое разочарование.

Каждый оценочный показатель надежности показывает различную величину надежности. В общем, значение оценки надежности методом внешних наблюдений и значение тест-ретестовой оценки надежности оказываются ниже, чем значение оценки надежности методом параллельных форм и значения оценок внутренней согласованности, потому что первые предполагают измерение через некоторый временной интервал либо измерение с участием различных внешних наблюдателей. Поскольку оценки надежности часто используются при проведении статистического анализа по квазиэкспериментальному плану (например, анализ по плану неэквивалентных групп), тот факт, что различные оценки способны значительно расходиться ещё сильнее осложняет анализ.

## НАДЕЖНОСТЬ И ВАЛИДНОСТЬ

Надежность и валидность часто мыслятся в качестве двух отдельных идей, но в действительности они друг с другом взаимосвязаны. Эту взаимосвязь можно мыслить двояко.

Одна из метафор, выражающих соотношение между надежностью и валидностью – это мишень. В центре мишени находится конструкт, который мы стараемся измерить. Представим, что каждый испытуемый, на котором мы измеряем интересующий нас конструкт – это выстрел по мишени. Если для данного испытуемого конструкт измерен абсолютно, значит, вы попали в центр мишени. В противном случае вы не попали в центр мишени. Чем менее точно измерение по данному испытуемому, тем дальше от центра мишени ваш выстрел.



**Надежно, но не валидно**

**Валидно, но не надежно**

**Ни валидно, ни надежно**

**И валидно, и надежно**

На рисунках изображены четыре возможных ситуации:

- ❖ На первом рисунке все попадания согласованы между собой, но мимо центра мишени. Это означает, что мы последовательно и систематическим образом измеряем неверное значение для всех испытуемых. Такое измерение надежно, но не валидно (т. е. оно согласовано, но ошибочно).
- ❖ На втором рисунке попадания случайным образом распределились по всей мишени. Вы редко попадаете в центр мишени, но, в среднем, вы получаете верный результат по группе испытуемых (хотя и не совсем верный по отдельным испытуемым). В данном случае, вы располагаете валидной оценкой группы, но эта оценка несогласованная. Из второго рисунка ясно видно, что надежность непосредственно связана с вариабельностью проводимого измерения.
- ❖ Третий рисунок иллюстрирует случай, при котором попадания распределяются по мишени, но согласованным образом не попадают в яблочко. В данном случае, измерение не является ни валидным, ни надежным.
- ❖ На четвертом рисунке – вариант Робина Гуда, согласованные попадания в яблочко. Измерение и валидно, и надежно.

Второй способ осмысления взаимосвязи между надежностью и валидностью представлен в виде матрицы 2 x 2.

**ОДИНАКОВЫЙ**

**КОНСТРУКТ**

**ОДИНАКОВЫЙ  
надежность**

**РАЗЛИЧНЫЙ  
дискриминантная  
валидность**

вер-
бальный
пись-
менный

=

вер-
бальный
пись-
менный

=

вербаль-
ный
письмен-
ный

математи-
ческий
письмен-
ный



## Метод



В столбцах матрицы указано, какой конструкт вы измеряете – один и тот же, или различный. В строках показано, какой метод измерения вы применяете – один и тот же, или различный. Допустим, существуют два конструкта, которые мы бы хотели измерить: вербальные способности и математические способности студента. Допустим, что каждый конструкт можно измерить двумя способами. Первый способ – выполнение письменного теста (типа ЕГЭ). Второй способ - опросить преподавателя, обучающего данного студента, оценить способности на основе непосредственных наблюдений за студентом в классе.

- ❖ В *левой верхней* клетке показано сравнение баллов по вербальному письменному тесту с балами по вербальному письменному тесту. Но каким образом возможно сравнение измерения с самим собой? Оно возможно благодаря оценке надежности письменного теста через тест-ретестовую корреляцию, через параллельные формы, либо через измерение внутренней согласованности. В этой ячейке оценивается надежность измерения.
- ❖ В *нижней левой* ячейке показано сравнение баллов по вербальному письменному тесту с рейтингом, определенным преподавателем в балльной форме. Поскольку мы стараемся измерить тот же самый конструкт, то мы ищем конвергентную валидность.
- ❖ В ячейке в *верхнем правом* углу показано сравнение баллов по вербальному письменному тесту с баллами по математическому письменному тесту. Здесь сравниваются два различных конструкта (вербальный с математическим), поэтому следует ожидать, что взаимосвязь окажется слабее, чем при сравнении конструкта с самим собой (например, вербального конструкта с вербальным, или математического конструкта с математическим). Таким образом, создается дискриминация между двумя конструктами, и нам следует рассматривать дискриминантную валидность.
- ❖ Наконец, в *нижней правой* ячейке сравниваются баллы по вербальному письменному тесту с выраженной в баллах наблюдаемой оценкой преподавателем математических способностей студента. Подобно ячейке в

верхнем правом углу, сравниваются два различных конструкта (вербальный с математическим), а это то, что оценивает дискриминантная валидность. Однако, в нашем случае сравниваются также два различных метода измерения (письменный экзамен и наблюдение преподавателем). Поэтому данный случай представляет собой очень большую дискриминантность, от которой следует ожидать, что взаимосвязь окажется ещё слабее, чем в верхней правой ячейке.

В четырех ячейках матрицы представлены четыре различные значения, которые исследуются методом мультисвойств-мультиметодов в целях оценки валидности конструктов.

Проиллюстрированный подход к надежности и валидности позволяет рассматривать их не как обособленные понятия, а в качестве континуума. На одном полюсе континуума расположена ситуация, в которой конструкты и методы их измерения одинаковы (надежность), а на другом полюсе оказывается ситуация, в которой концепты и методы их измерения различны (очень большая дискриминантная валидность).

### **Вопросы для самоконтроля по теме «Надёжность»**

1. Какое качество предполагает надежность?
2. Из каких компонентов складывается любое измерение?
3. Из чего образуется вариабельность измерения?
4. Почему важна теория истинного значения?
5. Что такое случайная ошибка?
6. Что такое систематическая ошибка?
7. Каким образом могут быть уменьшены случайные и систематические ошибки измерения?
8. Что такое надежность в научном исследовании?
9. Каким образом можно убедиться в том, что наблюдение согласовано (или допускает возможность повторения)?
10. Что общего имеется у двух наблюдений?
11. В каком виде можно себе представить надежность?
12. Можно ли рассчитать дисперсию истинных значений?
13. Можно ли вычислить надежность?
14. Что служит показателем совместной дисперсии двух измерений?
15. В виде чего выражается оценка надежности?
16. Насколько велика оценка надежности?
17. Как определить интервал оценки надежности?
18. Чему равна ошибка измерения, если измерение абсолютно надежно?
19. В каком интервале всегда находится надежность?
20. Что означает надежность, равная 0,5?
21. Сколько общих классов оценок надежности существует?
22. Каким образом можно обеспечить надежность внешних наблюдений?  
Сколько способов Вы знаете?

23. Каким образом оценить надежность внешних наблюдений в тех случаях, когда измерение является непрерывным?
24. Какой тип надежности можно назвать «калибровкой» внешних наблюдателей? Приведите пример.
25. Когда проводится оценка тест-ретестовой надежности?
26. Что необходимо для оценки надежности параллельных форм?
27. В чем заключается главное отличие подхода на основе параллельных форм от расколотой пополам надежности?
28. Как проводится оценка внутренне согласованной надежности?
29. Сколько существует способов оценки внутренней согласованности измерения?
30. Как рассчитать среднюю корреляцию между пунктами методики?
31. Как рассчитать среднюю корреляцию по общему баллу?
32. Что такое расколотая пополам надежность?
33. Чему в математическом смысле эквивалентна альфа Кронбаха?
34. Каковы преимущества и недостатки каждого из оценочных показателей надежности?
35. Обособлены ли друг от друга понятия надежности и валидности? В виде чего их можно рассматривать?

### Тестовые задания к теме 3.

1. Надежность предполагает качество
  - а) отношения
  - б) измерения
  - в) рассуждения
  - г) вычисления
  - д) применения
2. Теория истинного значения – это теория
  - а) отношения
  - б) измерения
  - в) рассуждения
  - г) вычисления
  - д) применения
3. В теории истинного значения утверждается, что каждое измерение представляет собой
  - а) произведение двух слагаемых
  - б) сумму двух слагаемых
  - в) разность двух слагаемых
  - г) квадрат двух слагаемых
  - д) матрицу двух слагаемых
4. Теория истинного значения выступает фундаментом

- а) теории эволюции
  - б) теории относительности
  - в) теории вероятности
  - г) теории надежности
  - д) теории личности
5. Влияние любых факторов (например, настроения испытуемого) на измерение переменной в выборке, следует считать
- а) случайной ошибкой
  - б) простительной ошибкой
  - в) систематической ошибкой
  - г) уважительной причиной
  - д) медвежьей услугой
6. Случайная ошибка выталкивает наблюдаемые значения
- а) куда подальше
  - б) куда поближе
  - в) куда поглубже
  - г) туда, где надежнее
  - д) то вверх, то вниз
7. Если бы мы увидели все случайные ошибки в распределении, то их сумма оказалась бы
- а) равной 3,14
  - б) равной нулю
  - в) четной
  - г) нечетной
  - д) надежной
8. На среднем функционировании группы испытуемых случайная ошибка
- а) отражается
  - б) преломляется
  - в) ускоряется
  - г) замедляется
  - д) не отражается
9. Случайная ошибка считается
- а) сигналом
  - б) кошмаром
  - в) шумом
  - г) гамом
  - д) равной нулю
10. В отличие от случайной ошибки, систематические ошибки имеют
- а) выраженную форму
  - б) выраженную погрешность
  - в) выраженную склонность
  - г) выраженную тенденцию
  - д) выраженный контур
11. Когда мы говорим о заслуживающем доверия измерении, мы имеем в виду, что это измерение

- а) не надежно, но валидно
- б) и надежно, и валидно
- в) надежно, но не валидно
- г) не надежно, и не валидно
- д) в принципе не важно

12. В научном исследовании термин «надежность» означает допускающий возможность

- а) измерения
- б) вычисления
- в) сомнения
- г) отрицания
- д) повторения

13. Для того, чтобы увидеть – насколько наблюдение согласовано, его можно

- а) измерить правильно
- б) измерить внимательно
- в) измерить тщательно
- г) измерить точно
- д) измерить дважды

14. Когда мы наблюдаем количественное значение того, что мы измеряем, мы обычно рассматриваем это значение состоящим

- а) из одного компонента
- б) из двух компонентов
- в) из многих компонентов
- г) из сходных компонентов
- д) из теоретических компонентов

15. Если наше измерение  $X$  надежно, то, если мы измерим или пронаблюдаем его дважды на одних и тех же испытуемых, то баллы в обоих случаях

- а) существенно искажутся
- б) существенно совпадут
- в) существенно отклонятся
- г) несущественно совпадут
- д) несущественно искажутся

16. Единственная вещь, которая является общей для двух наблюдений, это

- а) их истинные значения
- б) их повторные значения
- в) их первичные значения
- г) их точные значения
- д) их искаженные значения

17. Надежность можно себе представить в качестве пропорции

- а) вероятности в измерении
- б) наглядности в измерении
- в) истины в измерении
- г) метода в измерении
- д) принципа в измерении

18. Мы не можем рассчитать надежность, но мы можем её

- а) угадать
  - б) распознать
  - в) отменить
  - г) оценить
  - д) озвучить
19. В числителе формулы для оценки надежности находится
- а) оценка регрессии истинного значения
  - б) оценка дисперсии истинного значения
  - в) оценка корреляции истинного значения
  - г) оценка суммы истинного значения
  - д) оценка разности истинного значения
20. В знаменателе формулы для оценки надежности находится
- а) регрессия измерения
  - б) дисперсия измерения
  - в) корреляция измерения
  - г) сумма измерения
  - д) разность измерения
21. Если измерение абсолютно надежно, то ошибка измерения
- а) равна 0
  - б) равна 1
  - в) равна 3,14
  - г) равна стандартному отклонению
  - д) равна стандартной ошибке
22. Если измерение абсолютно ненадежно, то ошибка измерения
- а) равна 0
  - б) равна 1
  - в) равна 3,14
  - г) равна стандартному отклонению
  - д) равна стандартной ошибке
23. Надежность всегда находится в интервале
- а) от  $-1$  до  $+1$
  - б) от 0 до  $+1$
  - в) от  $-1$  до 0
  - г) от  $-\infty$  до  $+\infty$
  - д) от  $\max$  до  $\min$
24. Надежность, равная 0,8, означает, что вариабельность характеризуется истинностью
- а) на 20%
  - б) на 40%
  - в) на 80%
  - г) на 100%
  - д) на 200%

25. При оценке степени, в которой согласуются между собой оценки одного и того же явления различными наблюдателями, применяется

- а) надежность внешних наблюдений
- б) тест-ретестовая надежность
- в) надежность параллельных форм
- г) внутренне согласованная надежность
- д) старый проверенный способ

26. При оценке согласованности измерения в различные моменты времени, применяется

- а) надежность внешних наблюдений
- б) тест-ретестовая надежность
- в) надежность параллельных форм
- г) внутренне согласованная надежность
- д) старый проверенный способ

27. При оценке согласованности результатов по пунктам одного и того же тестового задания, применяется

- а) надежность внешних наблюдений
- б) тест-ретестовая надежность
- в) надежность параллельных форм
- г) внутренне согласованная надежность
- д) старый проверенный способ

28. Допустим, что внешние наблюдатели одинаковым образом распределили по категориям 86 наблюдений из 100. В этом случае процент согласия внешних наблюдателей составляет

- а) 0%
- б) 14 %
- в) 86%
- г) 100%
- д) 200%

29. Когда измерение является непрерывным, то для оценок внешних наблюдателей требуется рассчитывать

- а) ошибки
- б) погрешности
- в) отклонения
- г) искажения
- д) корреляции

30. Оценка тест-ретестовой надежности производится тогда, когда одна и та же тестовая методика применяется к одной и той же либо к аналогичной выборке

- а) не менее одного раза
- б) не менее двух раз
- в) не более одного раза
- г) не более двух раз

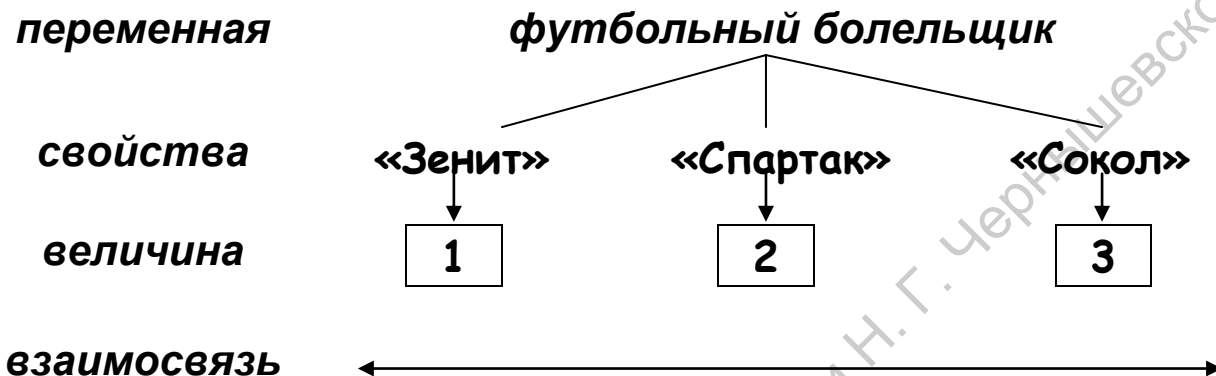
- д) тестовая методика вообще не применяется
31. Оценка тест-ретестовой надежности подразумевает, что в перерыве между двумя измерениями измеряемый конструкт
- а) не претерпевает значительных изменений
  - б) претерпевает незначительные изменения
  - в) претерпевает значительные изменения
  - г) не претерпевает значительных искажений
  - д) претерпевает значительные искажения
32. Чем короче временной интервал между двумя измерениями измеряемого конструкта, тем
- а) меньше корреляция
  - б) больше корреляция
  - в) симметричнее корреляция
  - г) негативнее корреляция
  - д) позитивнее корреляция
33. Проблема в связи с оценкой надежности параллельных форм состоит в необходимости генерирования
- а) большого количества экспериментов
  - б) большого количества испытуемых
  - в) большого количества опросников
  - г) большого количества пунктов опросника
  - д) большого количества оценок испытуемых
34. Для оценки тест-ретестовой надежности следует предложить одному и тому же внешнему наблюдателю выставить кодированные оценки
- а) дважды
  - б) трижды
  - в) сто раз
  - г) ежедневно
  - д) не выставлять кодированные оценки
35. Если тестовых заданий действительно много, то чаще всего для оценки внутренней согласованности используется
- а) тест-ретестовый оценочный показатель
  - б) расколотые корреляции
  - в) средняя корреляция по общему баллу
  - г) средняя корреляция между пунктами методики
  - д) альфа Кронбаха
36. Когда мы последовательно и систематическим образом измеряем неверное значение для всех испытуемых, то такое измерение
- а) валидно и надежно
  - б) валидно, но не надежно
  - в) надежно, но не валидно
  - г) ни надежно, ни валидно
  - д) не репрезентативно



## ТЕМА 4. ШКАЛИРОВАНИЕ

### УРОВНИ ИЗМЕРЕНИЯ

Уровень измерения соотносится с взаимосвязью между теми величинами, которые приписываются свойствам переменной. Что это означает?



Начнем с понятия переменной. В нашем примере «футбольный болельщик». Данная переменная обладает различными свойствами. Предположим, что в конкретном спортивном контексте у неё есть три релевантных свойства: болельщик «Зенита», болельщик «Спартака» и болельщик «Сокола». В целях анализа данных по этой переменной, мы произвольным образом приписали этим трем свойствам значения 1, 2 и 3.

**Уровень измерения** описывает взаимосвязь между этими тремя величинами. В данном случае, номера выступают сокращенными обозначениями того, что в буквенной записи представляется более длинным. Мы не подразумеваем того, что больший номер ассоциирован с чем-то большим, а меньший номер ассоциирован с чем-то меньшим. Мы не подразумеваем того, что болельщик «Спартака» в 2 раза в чем-либо отличается от болельщика «Зенита». Мы не подразумеваем того, что болельщик «Зенита» занимает первое место со своим номером 1 или обладает приоритетом. Мы используем номера только в целях упрощения обозначения и в целях замены номерами словесных наименований. Рассмотренный сейчас уровень измерения называется «номинальным» (уровень наименований).

### Почему уровень измерения важен?

Во-первых, знание уровня измерения помогает принять решение о том, каким образом следует интерпретировать данные по конкретной переменной. Когда известно, что измерение номинальное (аналогично описанному выше), то понятно, что числовые значения – это всего лишь краткие кодовые обозначения более длинных названий.

Во-вторых, знание уровня измерения помогает принять решение о выборе адекватной формы статистического анализа с учетом приписанных количественных значений. Если измерение является номинальным, то в отно-

шении него никогда не вычисляется среднее значение, никогда не рассчитывается t-критерий и т. п.

Существуют четыре уровня измерения, которые называются:

- ❖ Номинальное измерение
- ❖ Порядковое измерение
- ❖ Интервальное измерение
- ❖ Измерение отношений

При **номинальном** измерении, количественное обозначение оказывается всего лишь уникальным «наименованием» свойства. Не предполагается какого-либо упорядочения рассматриваемых случаев. Например, номера на форме футболистов – это пример измерения на номинальном уровне. Футболист с номером 10 (Пеле) ничем не хуже игрока с номером 30, и ни в коем случае не уступает ему в 3 раза.

При **порядковом** измерении, свойства допускают ранжирование. В этом случае дистанция между номерами не имеет никакого значения. Например, при проведении анкетного опроса, можно кодировать уровень образования следующим образом:

- 0 = начальное образование
- 1 = неполное среднее образование
- 2 = полное среднее образование
- 3 = незаконченное высшее образование
- 4 = полное высшее образование
- 5 = ученая степень

При таком измерении, более высокий номер соответствует более высокому уровню образования. Однако, дистанция между 0 и 1 равна ли дистанции между 3 и 4? Конечно, нет. При порядковом измерении интервалы между значениями не играют роли.

При **интервальном** измерении дистанция между измерениями играет роль. Например, когда мы измеряем температуру (по Фаренгейту<sup>6</sup>) то расстояние между 30° и 40° такое же, что и расстояние между 70° и 80°. Интервал между величинами имеет значение. По этой причине, целесообразно рассчитывать среднее значение интервальной переменной, однако, нет смысла поступать таким образом в отношении порядковой переменной. Но обратите внимание на то, что соотношения интервалов измерений не имеют смысла – при 80° не бывает в два раза более жарко, чем при 40° (хотя приписываемое значение оказывается вдвое большим).

При измерении **отношений** всегда существует абсолютный нуль. Это означает, что применительно к переменной отношений можно создать имеющее смысл отношение (или дробь). Переменной отношений является, например, вес. В прикладных психологических исследованиях большинство обсчитываемых переменных – это переменные отношений, например, количество испытуемых. Почему? Потому что может быть нуль испытуемых, по-

---

<sup>6</sup> C° = 5/9 (F° – 32°)

тому что имеет смысл утверждать, что в одной группе испытуемых вдвое больше, чем в другой группе.

	<b>Отно- шений</b>	<b>Абсолютный нуль</b>
	<b>Интер- вальный</b>	<b>Дистанция имеет смысл</b>
<b>Поряд- ковый</b>	<b>Свойства можно упорядочить</b>	
<b>Номи- нальный</b>	<b>Свойства только называются; слабее всех</b>	

Важно понимать, что уровни измерений иерархическим образом соподчиняются. На нижних уровнях измерения, предположения характеризуются меньшей ограничительностью, а анализ данных оказывается менее чувствительным. При подъеме на каждый следующий уровень иерархии, любой данный уровень включает в себя все свойства предшествующего уровня и добавляет к ним новые свойства. Вообще, более высокие уровни измерения (интервальный или уровень отношений) более предпочтительны перед низкими уровнями измерений (номинальным или порядковым).

## ШКАЛИРОВАНИЕ

Шкалированием называется область измерения, связанная с конструированием измерительного инструментария, которая сопрягает качественные конструкты с количественными метрическими единицами. Шкалирование возникло из усилий психологов и педагогов измерить такие «неизмеряемые» конструкты, как самоуважение и авторитарность. Во многих отношениях шкалирование остается одним из наиболее непонятых аспектов измерения в психологических исследованиях. Оно направлено на решение сложнейшей исследовательской задачи – измерение абстрактных понятий.

Большинство людей даже не понимают, что такое шкалирование. Шкалы обычно подразделяются на две широкие категории:

- ❖ одномерные
- ❖ многомерные

Методы одномерного шкалирования были разработаны в первой половине XX века и, как правило, носят имена своих авторов. Мы рассмотрим три типа методов одномерного шкалирования:

- ❖ Шкалирование Тёрстоуна или шкалирование с равными интервалами

- ❖ Шкалирование Лайкерта или суммирующее шкалирование
- ❖ Шкалирование Гуттмана или «кумулятивное» шкалирование

На рубеже 1950-1960х гг. теоретики измерений разработали более передовые техники, позволяющие осуществлять многомерное шкалирование.

## ОБЩИЕ ВОПРОСЫ ШКАЛИРОВАНИЯ

С. Стивенс сформулировал наиболее простое и прямолинейное определение шкалирования:

### «Шкалирование есть приписывание чисел объектам в соответствии с некоторыми правилами»

Что же это значит? В большинстве случаев шкалирования, объектами являются тексты высказываний, особенно те высказываний, в которых выражаются отношения и представления. Рассмотрим следующий пример.

Даны три утверждения, описывающие отношение к иммиграции. Для того, чтобы шкалировать эти утверждения, им необходимо присвоить числа. Как правило, стремятся к тому, чтобы в результате получилась о крайней мере шкала интервалов, что и представлено на схеме. Что означают слова «в соответствии с некоторыми правилами»? Если присмотреться к приведенным утверждениям, то можно заметить, что по мере перехода от верхнего утверждения к нижнему, отношение к мигрантам становится всё более строгим. Если человек соглашается с каким-либо утверждением из приведенного списка, то существует вероятность того, что этот человек согласится с теми утверждениями, которые находятся выше в данном списке. В данном случае правилом является правило кумулятивности.



иммигрантам проживать по соседству с вами?

Вы готовы разрешить своему ребенку вступить в брак с иммигрантом?



...В СООТВЕТСТВИИ С НЕКОТОРЫМИ ПРАВИЛАМИ...

Итак, что такое шкалирование? Это то, каким образом мы получаем те числа, которые могут быть осмысленным образом приписаны объектам, это комплекс процедур. Существуют несколько подходов к шкалированию. Но сначала обратим внимание на то, что люди часто путают понятие шкалы и понятие шкалы ответов. Шкала ответов – это способ собрать ответы испытуемых по данной методике. Шкала ответов может быть дихотомической: «согласен/не согласен»; «истинно/ложно»; «да/нет». Либо шкала ответов может быть интервальной с делениями от 1 до 5 или от 1 до 7. Но если не выполняется ничего кроме прикладывания шкалы ответов к объекту или к утверждению, то это нельзя назвать шкалированием. Шкалирование предполагает процедуры, которые выполняются независимо от испытуемого, т. е. подход к объекту с применением количественного значения. При настоящем шкалированном исследовании, процедура шкалирования применяются для разработки измерительного инструмента (шкалы), а шкала ответов применяется для сбора ответов испытуемых. Но простое присвоение делений шкалы от 1 до 5 пункту теста – это не шкалирование! Различия представлены в следующей таблице:

Шкала	Шкала ответов
получается в результате <b>процесса</b>	применяется для сбора <b>ответов</b> на тестовое задание
каждое деление на шкале является <b>значением шкалы</b>	тестовое задание <b>не связано</b> с значением деления шкалы

соответствует **ком-** применяется в от-  
**плексу тестовых** ношении **единич-**  
**заданий** **ного тестового за-**  
**дания**

### Цели шкалирования

Зачем заниматься шкалированием? Почему бы просто не создать текстовые утверждения ли вопросы и не применить формат ответов для сбора результатов?

- ❖ Во-первых, иногда шкалирование проводят с целью проверки гипотезы. Бывает необходимо узнать, является ли понятие или конструкт одномерным или многомерным.
- ❖ Иногда шкалирование проводят в качестве части исследовательного исследования. Бывает необходимо узнать то, какие измерения лежат в основании комплекса оценок. К примеру, если вы придумываете ряд вопросов, то шкалирование позволяет определить то, насколько вопросы согласуются друг другом, а также то, измеряют ли они один конструкт либо множественные конструкты.
- ❖ Наиболее распространенной причиной шкалирования выступает необходимость балльного выражения ответов испытуемых. Когда испытуемый выполняет ряд тестовых заданий, часто оказывается необходимо присвоить ответам единственное число, отражающее совокупные установки либо представления испытуемого. К примеру, на схеме, которую мы видели выше, желательно иметь возможность присвоить такое единственное число, которое описывало бы общее отношение испытуемого к иммигрантам.

### Метрика

У шкалы может быть любое количество размеров. Большинство разрабатываемых исследователями шкал имеют несколько размеров. Что такое размер? Метрику можно себе представить в виде количества линий. Когда требуется измерить какой-либо конструкт, то необходимо принять решение о том, можно ли измерить конструкт при помощи одной линии, либо потребуется больше линий. Например, рост – это одномерный конструкт. Конструкт роста вполне можно измерить при помощи одной-единственной линии (например, с помощью линейки). Вес тоже одномерный – он измеряется на весах. Жажду тоже можно считать одномерным конструктом – в каждый момент времени вы испытываете жажду в большей или в меньшей степени. Ясно видно, что рост и вес – конструкты одномерные. Но как быть с таким конструктом, как самоуважение? Если вы считаете, что самоуважение можно измерить при помощи линейки от высокого самоуважения к низкому, то, вероятно, для вас самоуважение – это одномерный конструкт.



Каким может быть двумерный конструкт? Во многих моделях интеллекта постулируются два главных измерения – математические способности и вербальные способности. О данном типе двумерной модели можно сказать, что она характеризуется двумя типами умственных способностей. У некоторых людей сильнее развиты вербальные способности, а слабее развиты математические способности. У других людей наоборот. Однако, если конструкт действительно двумерный, то отразить уровень испытуемого по такому конструкту с использованием одной единственной линии невозможно. Иными словами, для того, чтобы охарактеризовать умственные способности человека, потребуется отметить место, которое он занимает в пространстве с двумя измерениями: *x* и *y*.

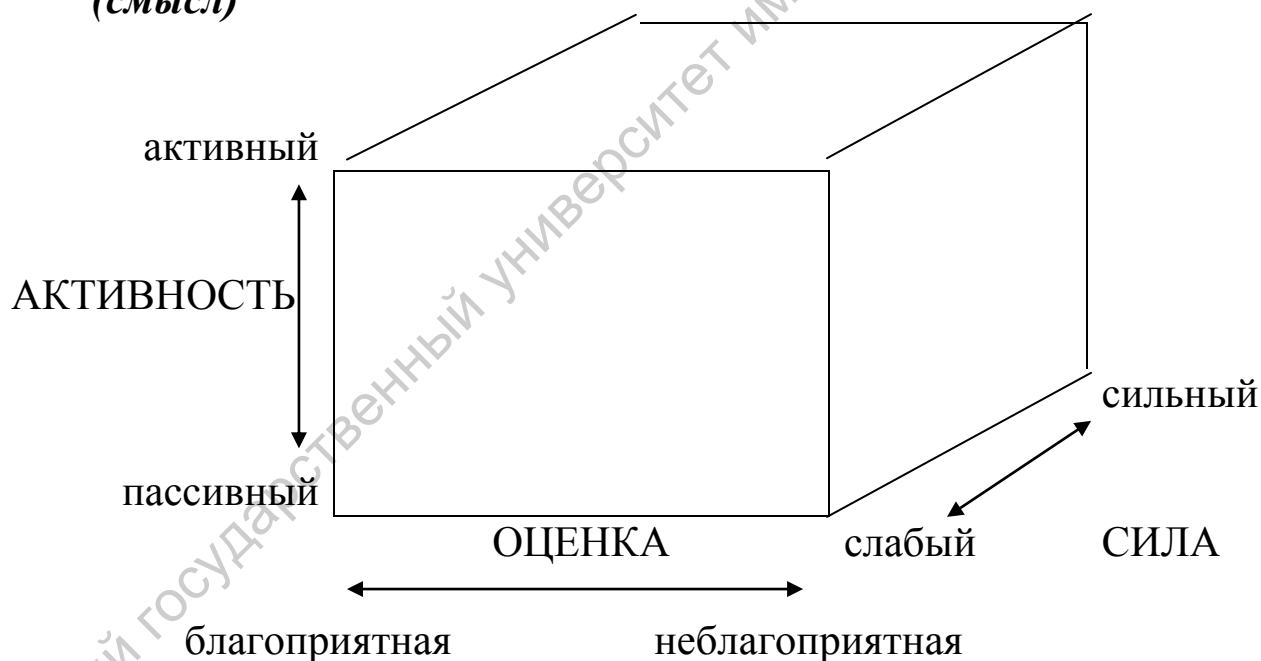


Сделаем следующий шаг: каким может быть трехмерный конструкт? Психологи, изучающие проблемы смысла, теоретически предположили, что смысл можно описать при помощи трех измерений. Иными словами, любые объекты можно отделить от остальных объектов по трем показателям. Эти три измерения были названы: *активность*, *оценка* и *сила*. Эта общая теория смысла получила название **семантического дифференциала**. В данной теории утверждается, что любой объект может быть расположен относительно этих трёх измерений. Например, балет. Если вы любите балет, то, вероятно,

по шкале активности вы отведете ему высокое место, по шкале оценки – благоприятное место, а по шкале силы – сильное место. С другой стороны, такой объект, как книга, допустим, роман. По шкале активности ему может быть отведено низкое место (книга пассивна), по шкале оценки – благоприятное место (если роман вам понравился), а по шкале силы – среднее место. Теперь рассмотрим такой объект, как посещение стоматолога. Большинство людей отводят ему низкое место по шкале активности (это пассивная активность), по шкале оценки – неблагоприятное место, а по шкале силы – слабое (мало что заставляет человека почувствовать себя настолько слабым!). Те исследователи, которые придерживаются теории семантического дифференциала, полагают, что смысл любых конструктов может быть прекрасно описан благодаря трем перечисленным измерениям. Иначе говоря, для того, чтобы описать смысл объекта, нужно точкой отметить его местоположение где-то внутри трехмерного пространства (т. е. внутри куба).

**Семантический  
дифференциал  
(смысл)**

**Трехмерный конструкт**



### **Одномерность или многомерность?**

Каковы преимущества применения многомерной модели? Одномерные конструкты обычно легче поддаются пониманию. Одномерных конструктов просто бывает либо меньше, либо больше, вот и всё. Вы либо выше, либо ниже, либо тяжелее, либо легче. Важно также понять, что одномерная шкала выступает основанием для понимания более сложных многомерных конструктов. Самая главная причина, по которой применяется одномерное шкалирование, заключается в том, что исследователь считает тот конструкт, кото-



рый он измеряет, одномерным в действительности. Как вы видели, многие знакомые конструкторы (рост, вес, температура) действительно оказываются одномерными. Однако, если конструктор, который вы изучаете, в действительности является многомерным, то одномерная шкала либо одна линия не сможет описать его как следует. Если попробовать измерить учебные способности по единому измерению, то получится ряд оценок испытуемых, расположенных от высоких баллов к низким. Но какое место в этом ряду займет тот, у кого высокие оценки по математике, но ужасные по языку, и наоборот? Одномерное шкалирование неприменимо к таким случаям.

### **Основные типы одномерных шкал**

Существуют три главных типа одномерных методов шкалирования. Они аналогичны друг другу в том отношении, что каждый из них измеряет интересующий исследователя конструктор с применением прямой линии. Но эти методы существенно различаются в отношении того, каким образом делениям на шкале присваиваются значения. Эти три метода называются:

- ❖ Шкалирование Тёрстоуна или шкалирование с равными интервалами
- ❖ Шкалирование Лайкерта или суммирующее шкалирование
- ❖ Шкалирование Гуттмана или «кумулятивное» шкалирование

### **ШКАЛИРОВАНИЕ ТЁРСТОУНА**

Тёрстоун был одним из первых и наиболее продуктивных теоретиков шкалирования. Он придумал три различных метода создания одномерной шкалы:

- ❖ метод равно возникающих интервалов;
- ❖ метод последовательных интервалов;
- ❖ метод парных сравнений.

Эти три метода различаются тем, каким образом конструируются шкальные значения по тестовым заданиям. Однако, во всех трех случаях, получающаяся в результате шкала ранжируется испытуемыми одинаковым образом. Для иллюстрации подхода Тёрстоуна, выберем простейший из трех методов – метод равно возникающих интервалов.

### **Метод равно возникающих интервалов**

**Формирование фокуса.** Метод равно возникающих интервалов начинается как почти любой метод шкалирования – с разработки фокуса шкалируемого проекта. Поскольку это метод одномерного шкалирования, допустим, что тот конструктор, который мы собрались шкалировать, имеет веские причины считаться одномерным. Описание этого конструктора должно быть ясным в максимальной степени, настолько, чтобы те, кто будут формировать утверждения, ясно себе представляли, что они в действительности собрались

измерять. Фокус шкалируемого проекта можно себе представить в форме команды – это команда, отдаваемая тем людям, которые будут придумывать утверждения. Например, команда может формулироваться следующим образом:

### **Придумайте утверждения, описывающие отношение людей к больным СПИДОМ.**

Необходимо убедиться в том, что каждый, кто занят формулированием утверждений, ясно себе представляет, что вы хотели поместить в фокус своей команды. Особенно следует проверить понимание технических терминов и аббревиатур (например, что такое СПИД?)

**Формулирование будущих делений на шкале.** Теперь можно приступать к формулированию утверждений. Потребуется большой набор предварительных утверждений (например, 80 – 100), поскольку из этого объема будут отбираться окончательные формулировки для шкалы. Также необходимо проверить, чтобы все утверждения имели аналогичные друг другу формулировки – чтобы они не различались по грамматике и по структуре. К примеру, формулировки должны быть такими, чтобы восприниматься в качестве утверждений, с которыми можно согласиться либо не согласиться. Вам не подойдут такие формулировки, когда одни фразы являются утверждениями, а другие – вопросами.

В нашем примере фокус направлен на разработку шкалы отношений к СПИДУ, а утверждения могут иметь примерно следующий вид:

- люди заболевают СПИДОМ потому что ведут себя аморально
- СПИДОМ можно заразиться в общественно туалете
- СПИД – это гнев Господень
- каждый, у кого СПИД – голубой
- СПИД – это эпидемия, и мы все заболеем
- люди, у которых СПИД, плохие
- люди, у которых СПИД, настоящие люди
- СПИД – это лекарство, а не болезнь
- СПИДОМ можно заразиться при гетеросексуальном сексе
- люди, у которых СПИД, похожи на моих родителей
- СПИДОМ можно заразиться через сиденье в туалете
- у женщин СПИД не бывает
- я ко всем отношусь одинаково – есть у них СПИД или нет
- СПИД дорого обходится обществу
- СПИД – это то, то бывает у других, не у меня
- жить со СПИДОМ невозможно
- дети не могут заразиться СПИДОМ
- СПИД это смертный приговор
- поскольку от СПИДА можно предохраниться, то вместо того, чтобы его лечить, нужно расходовать средства на предохранение
- люди, которые заразились СПИДОМ, заслужили это
- СПИД никому не оказывает предпочтения, любой может заразиться

- СПИД это самое худшее, то может с тобой случиться
- СПИД это хорошо, потому то помогает контролировать население
- если у тебя СПИД, то всё равно можно жить нормальной жизнью
- люди, у которых СПИД, не заслуживают нашей помощи
- к тому времени, когда я заболею СПИДОМ, придумают лекарство
- со мной СПИДА никогда не случится
- СПИДОМ нельзя заразиться при оральном сексе
- СПИД передается также, как простуда
- СПИД не делает различий
- СПИДОМ можно заразиться при поцелуях
- СПИД передается по воздуху
- презервативы всегда предохраняют от передачи СПИДА
- люди, которые заболели СПИДОМ, получили по заслугам
- если у тебя СПИД, то в течение года умрешь
- СПИД бывает у плохих людей, а я хорошая, у меня никогда не будет СПИДА
- меня не волнует, заболею ли я СПИДОМ, потому что ученые скоро найдут лекарство
- СПИД отвлекает от других болезней, которые более заслуживают нашего внимания
- заразить СПИДОМ свою семью – это самое худшее, что можно сделать
- СПИДОМ мало кто болеет, и вряд ли я когда-нибудь встречу больного СПИДОМ
- если мой брат заболеет СПИДОМ, то больше не буду с ним разговаривать
- люди, которые болеют СПИДОМ нуждаются в нашем понимании, но не обязательно в особом лечении
- СПИД вездесущий безжалостный убийца, который подкарауливает в темном переулке, молча ожидая, когда наивная жертва пройдет мимо, чтобы наброситься
- если у меня будут моногамные отношения, то я не заболею СПИДОМ
- государственное переливание крови безопасно
- все должны быть бдительны
- людей, у которых СПИД, следует помещать в карантин, изолируя от остального общества
- я живу не в городе, СПИД мне не грозит
- я знаю достаточно о том, как передается СПИД, поэтому могу оказывать помощь пациентам, больным СПИДОМ
- вирус СПИДА в меня никогда не проникнет
- каждый, то заразился СПИДОМ, получил поделом
- тот, у кого СПИД, похож на меня
- люди, инфицированные СПИДОМ, не предохранялись во время секса
- СПИД всех нас достанет
- к людям, у которых СПИД, следует относиться как ко всем остальным

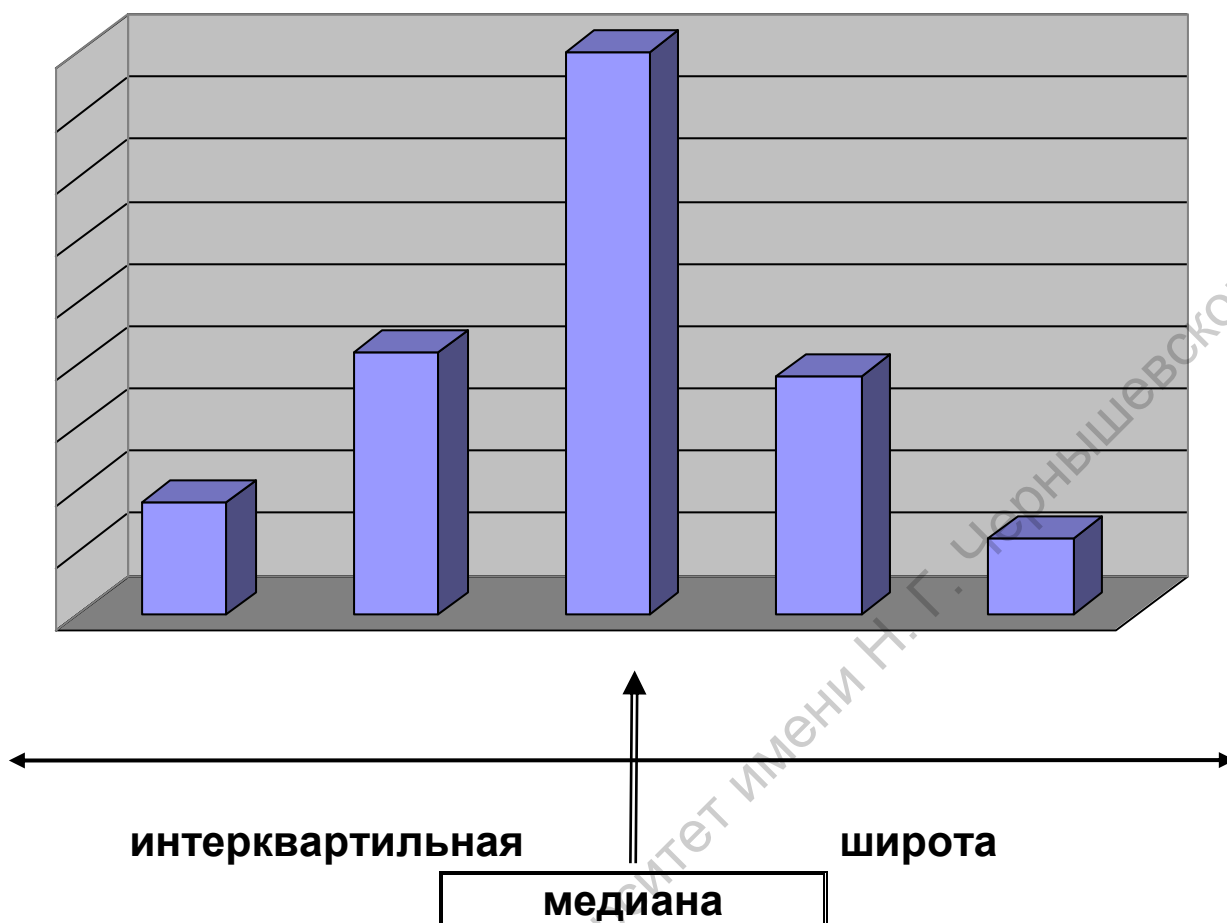
- СПИДОМ может заразиться каждый, если не будет вести себя осторожно
- заразиться СПИДОМ легко
- вероятность заразиться СПИДОМ очень небольшая
- шумиха вокруг СПИДА это эмоциональное напоминание о тех, кто умер незаслуженно или напрасно
- в Москве народу со СПИДОМ больше, чем принято считать
- людей убивает не вирус СПИДА, а осложнения после других болезней, потому что иммунная система перестает действовать
- СПИД всё больше становится проблемой для гетеросексуальных женщин и их детей, чем для наркоманов и гомосексуалистов
- скоро будет лекарство от СПИДА
- всех беременных женщин следует обязать делать тест на ВИЧ

**Ранжирование пунктов шкалы.** Теперь у нас имеется комплекс утверждений. Следующим шагом участники (например, судьи) поведут ранжирование каждого утверждения по шкале от 1 до 11 с точки зрения того, насколько каждое утверждение передает благожелательное отношение к больным СПИДОМ. Обратите на это особо пристальное внимание! Вам НЕ НУЖНО мнение участников о СПИДЕ, или о том, согласны ли они с конкретным утверждением. Вам нужно, чтобы они оценили «благожелательность» каждого утверждения с точки зрения отношения к СПИДУ, где

- ❖ **1** = «крайне неблагоприятное отношение к людям со СПИДОМ»,  
а
- ❖ **11** = «крайне благоприятное отношение к людям со СПИДОМ».

Обратите внимание также на то, что в принципе можно было бы попросить судей оценить то, насколько каждое утверждение выражает негативное отношение к СПИДУ. Составленная подобным образом шкала имела бы более высокие значения применительно к испытуемым с более негативными установками.

**Для каждого утверждения определяется распределение оценок**



## **находятся медиана и интерквартильная широта**

Расчёт балльных значений для каждого деления шкалы. Следующим шагом проводится анализ данных ранжирования. Для каждого утверждения требуется вычислить медиану и интерквартильную широту. Медиана – это значение, выше которого оказываются 50% ранговых оценок, и ниже тоже 50% ранговых оценок.

Итак, первый квартиль (Q1) – это значение, ниже которого оказываются 25% случаев. А выше третьего квартиля оказываются 75% случаев. Иными словами, первый квартиль – это 25й перцентиль. Третий квартиль (Q3) – это 75й перцентиль. Интерквартильная широта – это разность третьего и первого квартилей, либо  $Q1 - Q3$ . На диаграмме вверху представлена гистограмма по одному утверждению, и схема нахождения медианы и интерквартильной широты. Эти значения легко вычисляются любой статистической компьютерной программой. В целях облегчения окончательного отбора тех утверждений, которые войдут в состав шкалы, рекомендуется составить таблицу медиан и интерквартильных широт, располагая значения медианы в возрастающем порядке, а внутри них – расположить в убывающем порядке интерквартильные широты. Получится примерно следующая таблица:

Номер утверждения	Медиана	Q1	Q3	Интерквартильная широта
23	1	1	2,5	1,5
8	1	1	2	1
12	1	1	2	1
34	1	1	2	1
39	1	1	2	1
54	1	1	2	1
56	1	1	2	1
57	1	1	2	1
18	1	1	1	0
25	1	1	1	0
51	1	1	1	0
27	2	1	5	4
45	2	1	4	3
16	2	1	3,5	2,5
42	2	1	3,5	2,5
24	2	1	3	2
44	2	2	4	2
36	2	1	2,5	1,5
43	2	1	2,5	1,5
33	3	1	5	4
48	3	1	5	4
20	3	1,5	5	3,5
28	3	1,5	5	3,5
31	3	1,5	5	3,5
19	3	1	4	3
22	3	1	4	3
37	3	1	4	3
41	3	2	5	3
6	3	1,5	4	2,5
21	3	1,5	4	2,5
32	3	2	4,5	2,5
9	3	2	3,5	1,5
1	4	3	7	4
26	4	1	5	4
47	4	1	5	4
30	4	1,5	5	3,5
13	4	2	5	3
11	4	2	4,5	2,5
15	4	3	5	2
40	5	4,5	8	3,5

2	5	4	6,5	2,5
14	5	4	6	2
17	5,5	4	8	4
49	6	5	9,75	4,75
50	8	5,5	11	5,5
35	8	6,25	10	3,75
29	9	5,5	11	5,5
38	9	5,5	10,5	5
3	9	6	10	4
55	9	7	11	4
10	10	6	10,5	4,5
7	10	7,5	11	3,5
46	10	8	11	3
5	10	8,5	11	2,5
53	11	9,5	11	1,5
4	11	10	11	1

**Отбор окончательных делений шкалы.** Далее следует отобрать окончательные утверждения, которые составят шкалу. Отбирать следует те утверждения, которые характеризуются равными интервалами в диапазоне медианы. В нашем примере отбирать следует по одному утверждению для каждого из 11 значений медианы. По каждому значению медианы следует найти такое утверждение, которое характеризуется наименьшей интерквартильной шириной. По поводу таких учреждений среди судей возникло меньше всего разногласий, следовательно, вариабельность данного утверждения является наименьшей. Не следует в данном случае отводить статистике роль решающего фактора. Следует рассмотреть все утверждения каждого уровня значений медианы и выбрать то, в котором содержится максимум смысла. Если окажется, что наилучшее с точки зрения статистики утверждение имеет неясный смысл. Тогда следует выбрать утверждение, обладающее внятным смыслом, близкое к нему по статистическим показателям.

Произведя отбор утверждений, мы получили следующую шкалу:

- (1) люди, которые заболели СПИДОМ, получили по заслугам
- (2) СПИД это хорошо, потому то помогает контролировать население
- (3) со мной СПИДА никогда не случится
- (4) если у меня будут моногамные отношения, то я не заболею СПИДОМ
- (5) заразиться СПИДОМ легко
- (5) поскольку от СПИДА можно предохраниться, то вместо того, чтобы его лечить, нужно расходовать средства на предохранение
- (6) люди, у которых СПИД, похожи на моих родителей
- (8) если у тебя СПИД, то всё равно можно жить нормальной жизнью
- (9) СПИД никому не оказывает предпочтения, любой может заразиться
- (9) СПИДОМ может заразиться каждый, если не будет вести себя осторожно
- (10) СПИД всех нас достанет

(11) к людям, у которых СПИД, следует относиться как ко всем остальным

Значение в скобках соответствует делению шкалы. Чем выше деление шкалы, тем более благожелательно отношение к людям, заболевшим СПИДОМ. Обратите внимание на то, что у нас есть два утверждения, соответствующие отметке 5 на шкале, два утверждения, соответствующие отметке 9 на шкале, и ни одного утверждения, соответствующего отметке 7 на шкале. Также обратите внимание на то, что утверждения могут располагаться в произвольном порядке, независимо от значений по шкале.

**Как применять составленную шкалу.** Теперь у нас есть шкала для измерения установок в отношении людей, больных СПИДОМ. Её можно предложить испытуемому и попросить выразить своё согласие или несогласие с каждым утверждением. Для определения общего балла испытуемого по шкале, рассчитывают среднее значение по шкале для каждого утверждения, с которым испытуемый выразил своё согласие. Например, допустим, испытуемый ответил следующим образом:

<input checked="" type="checkbox"/>	<input type="checkbox"/>	Люди, у которых СПИД, похожи на моих родителей
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Поскольку от СПИДА можно предохраниться, то вместо того, чтобы его лечить, нужно расходовать средства на предохранение
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Люди, которые заболели СПИДОМ, получили по заслугам
<input checked="" type="checkbox"/>	<input type="checkbox"/>	СПИД всех нас достанет
<input checked="" type="checkbox"/>	<input type="checkbox"/>	К людям, у которых СПИД, следует относиться как ко всем остальным
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Со мной СПИДА никогда не случится
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Заразиться СПИДОМ легко
<input checked="" type="checkbox"/>	<input type="checkbox"/>	СПИД никому не оказывает предпочтения, любой может заразиться
<input checked="" type="checkbox"/>	<input type="checkbox"/>	СПИДОМ может заразиться каждый, если не будет вести себя осторожно
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Если у тебя СПИД, то всё равно можно жить нормальной жизнью
<input type="checkbox"/>	<input checked="" type="checkbox"/>	СПИД это хорошо, потому что помогает контролировать население



■ согласен	□ не согласен	Если у меня будут моногамные отношения, то я не заболею СПИДОМ
---------------	------------------	--

В приведенном примере испытуемый согласился с восемью утверждениями. Если вычислить среднее значение по шкале, т. е. сложить значения тех утверждений, с которыми согласился испытуемый:

$$6 + 5 + 10 + 11 + 9 + 9 + 8 + 4 = 62$$

и разделить их на 8, то получим итоговое значение для испытуемого, равное **7,75**. Оно соответствует месту испытуемого на нашей линейке, измеряющей отношение к людям, которые больны СПИДОМ. Теперь посмотрим пример с другим испытуемым:

□ согласен	■ не согласен	Люди, у которых СПИД, похожи на моих родителей
□ согласен	■ не согласен	Поскольку от СПИДА можно предохраниться, то вместо того, чтобы его лечить, нужно расходовать средства на предохранение
■ согласен	□ не согласен	Люди, которые заболели СПИДОМ, получили по заслугам
□ согласен	■ не согласен	СПИД всех нас достанет
□ согласен	■ не согласен	К людям, у которых СПИД, следует относиться как ко всем остальным
■ согласен	□ не согласен	Со мной СПИДА никогда не случится
□ согласен	■ не согласен	Заразиться СПИДОМ легко
□ согласен	■ не согласен	СПИД никому не оказывает предпочтения, любой может заразиться
□ согласен	■ не согласен	СПИДОМ может заразиться каждый, если не будет вести себя осторожно
□ согласен	■ не согласен	Если у тебя СПИД, то всё равно можно жить нормальной жизнью
■ согласен	□ не согласен	СПИД это хорошо, потому то помогает контролировать население

■ согласен	□ не согласен	Если у меня будут моногамные отношения, то я не заболею СПИДОМ
---------------	------------------	---

В этом примере испытуемый согласился лишь с 4 утверждениями, каждое из которых располагается на отрицательной стороне шкалы. Средний балл этого испытуемого по нашей шкале составляет 2,5 – значительно меньше первого испытуемого, а значит, отношение второго испытуемого к больным СПИДОМ более негативное.

### Другие методы Тёрстоуна

Остальные методы шкалирования по Тёрстоуну аналогичны рассмотренному методу равно возникающих интервалов. Каждый из них начинается с определения фокуса того конструкта, который полагается одномерным, и предполагает придумывание широкого круга потенциальных утверждений шкалы. В результате всех методов Тёрстоуна получается шкала, содержащая сравнительно немного делений, которые испытуемый оценивает с позиции «согласен/не согласен». Главное различие заключается в способе получения данных от судей. Например, метод парных сравнений предусматривает вынесение каждым судьей решения по каждой паре суждений. Поскольку суждений много, эта процедура занимает довольно много времени. Если утверждений, скажем 57, то пар суждений получится 1569, и все их требуется сравнивать!

### ШКАЛИРОВАНИЕ ЛАЙКЕРТА

Аналогично шкалированию Тёрстоуна, шкалирование Лайкерта является одномерным. Рассмотрим основные этапы создания шкалы Лайкерта или «суммирующей» шкалы.

**Определение фокуса.** Подобно любым методам шкалирования, первый шаг заключается в определении того, что вы собираетесь измерять. Поскольку это метод одномерного шкалирования, то предполагается, что конструкт, который вы желаете измерить, является по своей природе одномерным. Рекомендуются операционализировать формулировку конструкта, придав её вид инструкции для тех людей, которые будут создавать первоначальный круг утверждений.

**Разработка утверждений.** Следующим шагом необходимо придумать утверждения, из которых будут отобраны деления шкалы. Это должны быть такие утверждения, которые допускают оценивание по шкале ответов от 1 до 5 либо от 1 до 7 с вариантами ответов от «несогласен» до «согласен». Допускается самостоятельное придумывание утверждений разработчиком шкалы с опорой на собственное глубокое понимание тематики проблемы. Однако, часто бывает полезно привлечь к этому этапу процесса разработки шкалы некоторое количество людей. Допускаются разновидности мозгового штурма. На данном этапе желательно получить как можно больше возможных вариантов, лучше всего, если их окажется от 80 до 100.

**Ранжирование утверждений.** Следующий шаг заключается в оценке сформулированных утверждений группой судей. Обычно ранжирование вариантов проводится с применением шкалы от 1 до 5, где

- ❖ 1 = вариант, решительно благоприятный для измеряемого конструкта
- ❖ 2 = вариант, в некоторой степени неблагоприятный для измеряемого конструкта
- ❖ 3 = нет однозначного мнения
- ❖ 4 = вариант, в некоторой степени благоприятный для измеряемого конструкта
- ❖ 5 = вариант, решительно благоприятный для измеряемого конструкта

Обратите внимание на то, что, аналогично другим методам шкалирования, судьи не высказывают своего мнения – ни оценивают, насколько благоприятно каждое предложенное утверждение в отношении рассматриваемого конструкта.

**Отбор утверждений.** Следующий шаг состоит в вычислении коэффициентов корреляции между всеми парами утверждений, на основе оценок, присвоенных судьями. При принятии решения о том, какие из первоначальных утверждений войдут в окончательную шкалу, применяют следующие разновидности анализа:

- ❖ Отбрасывают все утверждения, у которых оказались малые корреляции с суммой оценок по всем утверждениям. В большинстве пакетов статистических программ данная разновидность статистического анализа проводится сравнительно просто.
  - Сначала создается новая переменная, которая представляет собой сумму рангов всех отдельных утверждений, предложенных каждым участником.
  - Далее, эта новая переменная включается в вычисление корреляционной матрицы (в качестве последней переменной списка).

Насколько малым должен быть коэффициент корреляции, для того, чтобы исключить утверждение? Строгого правила нет. Например, можно вычеркнуть все утверждения, коэффициент корреляции которых меньше 0,6.

- ❖ Для каждого утверждения
  - вычисляется средний ранг, присвоенный верхней четвертью судей и нижней четвертью судей. Затем
  - высчитывается t-критерий с целью определения различий между средним значением в верхней четверти судей и средним значением в нижней четверти судей. Чем больше значение t-критерия, тем больше различие между верхней четвертью судей и нижней четвертью судей. Говоря практически, утверждения, для которых значения t-критерия больше, более спо-

способствуют выявлению различий по шкале. В конце вычислений,

- рекомендуется самостоятельно рассудить, какие утверждения целесообразно оставить.

В окончательном виде шкала должна состоять из небольшого набора утверждений (примерно 10 – 15). Это должны быть утверждения, обладающие

- ✓ большим коэффициентом корреляции с общей суммой оценок и
- ✓ сильнее способствующие выявлению различий по шкале (большими значениями t-критерия).

**Применение шкалы.** Теперь всё готово к тому, что применить шкалу Лайкерта. Испытуемого просят проранжировать каждое утверждение по шкале оценок. Шкала оценок может, например, включать оценки от 1 до 5, где

- 1 = категорически не согласен
- 2 = не согласен
- 3 = нет мнения
- 4 = согласен
- 5 = абсолютно согласен

Шкалы оценок могут иметь другой вид: от 1 до 7, от 0 до 4 и т. п. Во всех вариантах нечетных шкал оценок присутствует среднее значение, которое называется нейтральным, либо «мнение отсутствует». Возможен также вариант четной шкалы оценок с вынужденным выбором, в которой отсутствует срединное значение. Такой тип шкалы оценок принуждает испытуемого принять решение о том, к чему он более склонен – либо более склонен согласиться с утверждением, либо более склонен не согласиться.

Итоговый балл испытуемого по шкале представляет собой сумму оценок по всем пунктам шкалы (поэтому шкалу Лайкерта иногда называют «суммирующей» шкалой). В некоторых разновидностях шкалы Лайкерта, деления на шкале направлены в обратную сторону от общего направления шкалы. Такие разновидности называются **реверсивными шкалами (делениями)**. Оценочные значения по делениям таких шкал следует обращать, прежде чем складывать общую сумму. Это означает, что, если испытуемый выбрал оценку 1, то 1 следует считать 5, если испытуемый выбрал оценку 2, то 2 следует считать 4, 3 = 3, 4 = 2, а 5 = 1.

### **Пример: шкала самоуважения сотрудников**

Предлагается пример шкалы Лайкерта из 10 утверждений, предназначенной для оценки самоуважения работников. Обратите внимание на то, что в данной методике отсутствует центр или нейтральная точка – испытуемый должен определить, согласен ли он с предлагаемым утверждением.

Инструкция: Просим Вас оценить, насколько сильно Вы согласны или не согласны с каждым из следующих утверждений, отметив своё мнение в соответствующей клетке.

	Категори- чески не согласен	В чем-то не согла- сен	В чем-то согласен	Абсолют- но согла- сен
1. У меня хорошее са- мочувствие на работе	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. В целом, у меня хо- рошие отношения с другими сотруднича- ми	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Я горжусь своей способностью справ- ляться с трудностями в работе	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Когда на работе я испытываю диском- форт, то знаю, как его преодолеть	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Мог сказать, что другие сотрудники рады со мной работать	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Я знаю, что смогу справляться с порученной работой столько времени, сколько пожелаю	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Я горжусь тем, как- кие у меня отношения с начальником	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Уверен, что посто- янная помощь в рабо- те мне не требуется	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. У меня есть ощу- щение, что я вношу полезный вклад в ра- боту	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Могу сказать, что коллеги меня уважают	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## ШКАЛИРОВАНИЕ ГУТТМАНА

Шкалирование по Гуттману иногда ещё называют **кумулятивным шкалированием** или **анализом шкалограммы**. Цель шкалирования по Гуттману состоит в разработке одномерного континуума для того конструкта, который желательно измерить. Что это значит? В действительности, нам желательно получить набор таких утверждений, чтобы испытуемый, который согласен с отдельными утверждениями из предложенного списка, согласился бы и со всеми предшествующими утверждениями. Говоря более формально, мы бы хотели иметь возможность прогнозировать ответы испытуемого на предложенные утверждения на основании общего балла для всех испытуемых. Например, представим себе кумулятивную шкалу из 10 утверждений. Если испытуемый получает четыре балла, то это значит, что испытуемый согласился с первыми четырьмя утверждениями. Если испытуемый набрал восемь баллов, значит, испытуемый согласился с первыми восемью утверждениями. Задача состоит в том, чтобы найти такой комплекс утверждений, который бы абсолютно вписывался в данную модель. Итак, анализ шкалограммы применяется для определения того, насколько тесно набор утверждений согласуется с понятием кумулятивности. Шкала Гуттмана составляется следующим образом.

**Определение фокуса.** Подобно любым методам шкалирования, начинаем с определения фокуса нашей шкалы. Представим себе, что мы решили разработать кумулятивную шкалу, измеряющую отношение граждан к приезжим. Необходимо удостовериться в том, что предлагаемая формулировка конкретизирует, какой тип иммиграции имеется в виду: законная или незаконная, а также об иммиграции откуда идет речь: из Африки, из Азии, из ближнего зарубежья и т. п.

**Разработка утверждений.** Далее, подобно любым методам шкалирования, разрабатывается большой набор утверждений, в которых отражается измеряемый конструкт. Это можно сделать самому, либо в помощь можно пригласить к участию группу специалистов. Допустим, получены следующие формулировки утверждений:

- ❖ Я бы согласился на брак моего ребенка с приезжим
- ❖ Я считаю, что в нашу страну следует впускать больше приезжих
- ❖ Я бы чувствовал себя комфортно, если бы на одной лестничной площадке со мной поселился приезжий
- ❖ Я бы чувствовал себя комфортно, если бы в моем микрорайоне поселились приезжие
- ❖ Я бы замечательно относился к тому, чтобы по соседству поселились приезжие
- ❖ Я бы чувствовал себя комфортно, если бы мой ребенок встречался с кем-либо из приезжих

Разумеется, предпочтительно, чтобы первичных утверждений было намного больше (желательно от 80 до 100).

**Оценивание утверждений.** Следующим шагом, утверждения предлагаются для оценки группе судей, которым предстоит решить, насколько бла-

благоприятны предложенные утверждения с точки зрения конструкта иммиграции. Они должны отмечать «да» благоприятные утверждения, а «нет» неблагоприятные утверждения. Обратите внимание на то, что от судей не требуется высказываться, согласны ли они лично с предложенными утверждениями. Вместо этого им предстоит судить о том, каким образом предложенные суждения связаны с интересующим нас конструктом.

**Создание кумулятивной шкалы.** Изюминка шкалирования по Гуттману заключена в анализе. Создают матрицу или таблицу, в которой представлены ответы всех участников по всем утверждениям. Далее эту матрицу сортируют таким образом, чтобы те испытуемые, которые согласились с большим количеством утверждений, оказывались вверху матрицы, а те, среди кого количество согласий с утверждениями меньше, оказывались бы внизу матрицы. Если испытуемые согласились с одинаковым количеством утверждений, то утверждения сортируют слева направо таким образом, чтобы слева оказались те утверждения, с которыми согласились большинство испытуемых, а справа те утверждения, с которыми согласились меньше всего испытуемых. Получается примерно следующая таблица:

**сортировка по строкам и столбцам показывает –  
кумулятивная ли это шкала**

Испытуемый	Утверждение 2	Утверждение 7	Утверждение 5	Утверждение 3	Утверждение 8	Утверждение...
7	Да	Да	Да	Да	Да	Да
15	Да	Да	Да	–	Да	–
3	Да	Да	Да	Да	–	–
29	Да	Да	Да	Да	–	–

19	Да	Да	Да	–	–	–
32	Да	Да	–	Да	–	–
41	Да	Да	–	–	–	–
6	Да	Да	–	–	–	–
14	Да	–	–	Да	–	–
33	–	–	–	–	–	–

Исключения

Обратите внимание на то, что если читать столбцы (утверждения) слева направо, то шкала очень близка к кумулятивной. Речь идет о том, что если испытуемый соглашался с утверждением 7, то всегда соглашался с утверждением 2. А тот испытуемый, который соглашался с утверждением 5, всегда соглашался с утверждениями 7 и 2. Из матрицы видно, однако, что кумулятивность шкалы не абсолютная. Хотя в общем случае испытуемый, который соглашался с утверждением 3, был склонен согласиться с утверждениями 5, 7 и 2, однако, из этого правило оказалось несколько исключений.

Мы рассматриваем матрицу, включающую всего лишь несколько утверждений. Если утверждений множество, то необходимо проводить анализ данных, который называется **анализом шкалограммы**, чтобы определить те подгруппы утверждений из нашего первоначального массива, которые лучше всего удовлетворяют свойству кумулятивности. Далее из этих подгрупп отбираются элементы окончательной шкалы. Существует несколько статистических приемов исследования матрицы в целях поиска кумулятивной шкалы. Поскольку абсолютно кумулятивная шкала возможна в редких случаях, то обычно её тестируют, чтобы выяснить, насколько она хороша. Эти статистические приемы также позволяют оценить шкальное значение каждого утверждения. Эти шкальные значения применяются при итоговом вычислении баллов испытуемого.

**Применение шкалы.** После того, как отобраны утверждения, и шкала приобрела готовый вид, применять её относительно просто. Следует лишь предъявить испытуемому утверждения и попросить испытуемого отметить те из них, с которыми испытуемый согласен. Применительно к нашей гипотетической шкале о приезжих, утверждения могут расположиться в следующем кумулятивном порядке:

- ❖ Я считаю, что в нашу страну следует впускать больше приезжих
- ❖ Я бы чувствовал себя комфортно, если бы в моем микрорайоне поселились приезжие
- ❖ Я бы замечательно относился к тому, чтобы по соседству поселились приезжие
- ❖ Я бы чувствовал себя комфортно, если бы на одной лестничной площадке со мной поселился приезжий



- ❖ Я бы чувствовал себя комфортно, если бы мой ребенок встречался с кем-либо из приезжих
- ❖ Я бы согласился на брак моего ребенка с приезжим

Разумеется, когда утверждения предъявляются респонденту, рекомендуется менять их очередность. В итоге шкала принимает следующий вид:

**Инструкция:** Просим Вас отметить те утверждения, с которыми Вы согласны.

- Я бы согласился на брак моего ребенка с приезжим
- Я считаю, что в нашу страну следует впускать больше приезжих
- Я бы чувствовал себя комфортно, если бы на одной лестничной площадке со мной поселился приезжий
- Я бы чувствовал себя комфортно, если бы в моем микрорайоне поселились приезжие
- Я бы замечательно относился к тому, чтобы по соседству поселились приезжие
- Я бы чувствовал себя комфортно, если бы мой ребенок встречался с кем-либо из приезжих

Каждое деление шкалы (каждое утверждение) имеет собственное значение, вычисленное в результате анализа шкалограммы. Для определения балла испытуемого по данной шкале просто складывают балльные значения, соответствующие каждому утверждению, с которым испытуемый согласился. Суммарный балл по шкале в нашем примере отражает отношение испытуемого к приезжим.

### **Вопросы для самоконтроля по теме «Шкалирование»**

1. С чем соотносится уровень измерения? Что это означает?
2. Почему уровень измерения важен?
3. Сколько уровней измерения существует? Как они называются?
4. Что такое номинальное измерение?
5. Что такое порядковое измерение?
6. Что такое интервальное измерение?
7. Что такое измерение отношений?
8. Каким образом соподчиняются уровни измерений?
9. Какая область измерения называется шкалированием?
10. Как возникло шкалирование?
11. На какие две широкие категории обычно подразделяются шкалы?
12. Какие методы одномерного шкалирования Вы знаете?

13. Сформулируйте определение шкалирования по С. Стивенсу. Что оно означает? Приведите пример.
14. Что такое шкала ответов?
15. Приведите пример дихотомической и интервальной шкалы ответов.
16. В чем отличие шкалы ответов от настоящего шкалирования?
17. Каковы цели шкалирования?
18. Какая размерность может быть у шкалы?
19. Приведите примеры одномерных конструкторов.
20. Каким может быть двумерный конструктор? Приведите пример.
21. Каким может быть трехмерный конструктор? Приведите пример.
22. Каковы преимущества применения многомерной модели?
23. Чем различаются одномерные методы шкалирования?
24. С чего начинается разработка методов одномерного шкалирования?
25. Что такое интерквартильная широта?
26. Какие шкалы называются реверсивными?
27. Как называется шкалирование по Гуттману?
28. В чем заключается цель шкалирования по Гуттману?

#### Тестовые задания по теме 4

1. Взаимосвязь между значениями переменной выражается
  - а) количеством измерений
  - б) надежностью измерений
  - в) валидностью измерений
  - г) дисперсией измерений
  - д) уровнем измерений
2. При каком измерении, количественное обозначение оказывается всего лишь уникальным наименованием свойства
  - а) номинальное измерение
  - б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
3. При каком измерении, свойства допускают ранжирование
  - а) номинальное измерение
  - б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
4. При каком измерении всегда существует абсолютный нуль
  - а) номинальное измерение

- б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
5. При каком измерении играет роль дистанция между измерениями
- а) номинальное измерение
  - б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
6. В прикладных психологических исследованиях большинство обчислимых переменных – это
- а) номинальные переменные
  - б) порядковые переменные
  - в) интервальные переменные
  - г) переменные отношений
  - д) в принципе любые переменные
7. Уровни измерений соподчиняются
- а) параллельно
  - б) последовательно
  - в) синхронно
  - г) иерархически
  - д) взаимно
8. Область измерения, связанная с конструированием измерительного инструментария, которая сопрягает качественные конструкты с количественными метрическими единицами, называется
- а) тестирование
  - б) ранжирование
  - в) шкалирование
  - г) функционирование
  - д) коррелирование
9. Суммирующим шкалированием называется
- а) шкалирование Тёрстоуна
  - б) шкалирование Лайкерта
  - в) шкалирование Гуттмана
  - г) шкалирование с равными интервалами
  - д) кумулятивное шкалирование
10. Согласно определению С. Стивенса, шкалирование есть приписывание чисел объектам в соответствии
- а) со всеми правилами
  - б) с любыми правилами
  - в) с некоторыми правилами
  - г) с одним правилом
  - д) правил шкалирования не существует

11. Наиболее распространенной причиной шкалирования выступает необходимость

- а) понять, что ответил испытуемый
- б) исправить ответ испытуемого
- в) подсказать испытуемому правильный ответ
- г) выразить ответ испытуемого в баллах
- д) узнать, на какой вопрос ответил испытуемый

12. Какое количество размеров может быть у шкалы:

- а) один размер
- б) два размера
- в) много размеров
- г) сколько угодно размеров
- д) столько, сколько испытуемых

13. Метрику шкалы можно себе представить в виде количества

- а) испытуемых
- б) ответов
- в) мнений
- г) баллов
- д) линий

14. Сколько размеров имеет конструкт под названием семантический дифференциал

- а) один размер
- б) два размера
- в) три размера
- г) много размеров
- д) столько, сколько испытуемых

15. Самая главная причина, по которой применяется одномерное шкалирование, заключается в том, что исследователь считает тот конструкт, который он измеряет

- а) надежным
- б) валидным
- в) основным
- г) вспомогательным
- д) одномерным

16. Значение, ниже которого оказываются 25% случаев, называется

- а) первый процентиль
- б) первый квартиль
- в) медиана
- г) мода
- д) стандартное отклонение

17. Какому процентилю соответствует третий квартиль

- а) первому процентилю
  - б) двадцать пятому процентилю
  - в) пятидесятому процентилю
  - г) семьдесят пятому процентилю
  - д) тысячному процентилю
18. Интерквартильная широта – это
- а) разность третьего и первого квартилей
  - б) разность восьмого и четвёртого квартилей
  - в) разность любых квартилей
  - г) разность квартилей и процентилей
  - д) косинус угла альфа
19. Шкалу Лайкерта иногда называют «суммирующей» шкалой, потому что итоговый балл испытуемого по шкале представляет собой
- а) корреляцию оценок по всем пунктам шкалы
  - б) дисперсию оценок по всем пунктам шкалы
  - в) стандартную ошибку оценок по всем пунктам шкалы
  - г) стандартное отклонение оценок по всем пунктам шкалы
  - д) сумму оценок по всем пунктам шкалы
20. Когда деления на шкале направлены в обратную сторону от общего направления шкалы, то такая разновидность называются
- а) кумулятивной шкалой
  - б) реверсивной шкалой
  - в) интервальной шкалой
  - г) порядковой шкалой
  - д) номинальной шкалой
21. Цель шкалирования по Гуттману состоит в разработке
- а) параллельного измерения
  - б) валидного континуума
  - в) одномерного континуума
  - г) надежного континуума
  - д) многомерного конструкта
22. Если по шкале Гуттмана испытуемый набрал восемь баллов, значит, с первыми восемью утверждениями испытуемый
- а) не знаком
  - б) не согласился
  - в) знаком слабо
  - г) согласился
  - д) ни о чём не говорит
23. Когда судьи оценивают утверждения для формируемой шкалы, то согласие судей с оцениваемыми утверждениями
- а) желательно
  - б) обязательно
  - в) единодушное
  - г) ни о чём не говорит
  - д) не важно

## **Семинар 1. Измерения, шкалы и статистика.**

Вопросы для обсуждения:

1. Психологические шкалы и психологические измерения.
2. Типы шкал.
3. Измерительные шкалы.
4. Порядковые и интервальные измерения.
5. Измерения отношений.
6. Переменные и их измерение.

Темы для рефератов:

1. Символы, данные и операции.
2. Этапы многомерного шкалирования.
3. Применение многомерного шкалирования.
4. Многомерное шкалирование и шкалирование Тёрстоуна.
5. Многомерное шкалирование и иерархический кластер-анализ.
6. Многомерное шкалирование и факторный анализ.

Литература к семинару 1:

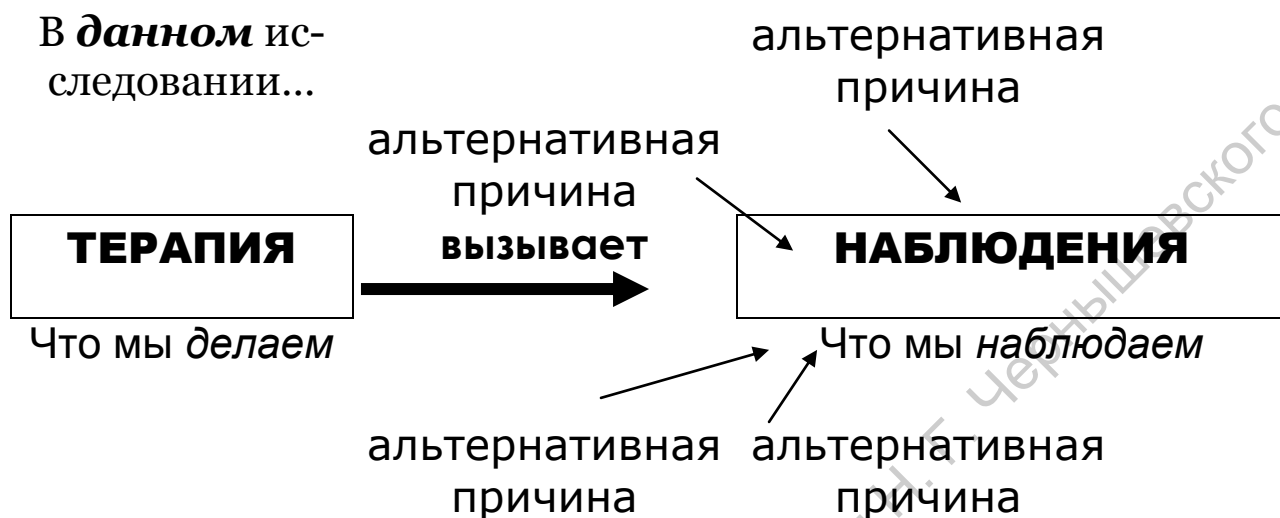
1. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., 1976. С. 12 – 23
2. Дейвисон М. Многомерное шкалирование. Методы наглядного представления данных. М., 1988. С. 13 – 19, 204 – 212
3. Крылов В. Ю. Геометрическое представление данных в психологических исследованиях. М., 1990. С. 46 – 51
4. Психология и математика. М., 1976. С. 15 – 18

## **МОДУЛЬ 2**

### **ТЕМА 5. ВНУТРЕННЯЯ ВАЛИДНОСТЬ**

Внутренней валидностью называется приблизительная истинность в отношении выводов о причинно-следственных либо о причинных взаимосвязях. Таким образом, внутренняя валидность релевантна лишь в тех исследованиях, в которых делаются попытки установить каузальные взаимосвязи. Она не релевантна, например, в большинстве исследований, связанных с наблюдением и с описанием. Однако, применительно к исследованиям, в которых оцениваются эффекты социальных программ либо интервенций, внутренняя валидность играет первостепенную роль. В контексте подобных исследований, необходимо суметь сделать вывод о том, что применяющаяся программа либо терапия создает некоторое различие – либо улучшает баллы по тестам, либо уменьшает симптоматику. Однако, существует множество причин, помимо применяющейся программы экспериментального воздействия либо программы терапии, которые способны обеспечивать увеличение тестовых баллов либо уменьшение симптомов. Основной вопрос применительно к внутренней валидности заключается в том, можно ли объяснить на-

блюдаемые изменения применяющейся программой либо терапией (т. е. причиной), а **не** другими вероятными причинами (которые иногда называют «альтернативными объяснениями» полученных результатов).



Применительно к внутренней валидности труднее всего понять то, что она релевантна лишь в отношении конкретного рассматриваемого исследования. Иными словами, внутренняя валидность характеризуется «нулевой обобщаемостью». Внутренняя валидность подразумевает только лишь подтверждения того, что то, что было сделано в данном исследовании (т. е. программа) вызывает то, что наблюдается (т. е. результат). Внутренняя валидность не указывает на то, хотели ли вы выполнить именно ту программу, которую выполнили, ни на то, хотели ли вы наблюдать именно то, что в итоге наблюдается – эти вопросы относятся к сфере валидности конструкторов. Может быть так, что исследование обладает внутренней валидностью, но не обладает валидностью конструкторов. Например, представьте себе исследование, в котором изучаются эффекты новой обучающей компьютерной программы по математике для студентов второго курса. Предположим, что программа уникальна в том отношении, что она включает в себя сильный игровой компонент, и вы считаете, что именно этот игровой компонент призван обеспечить повышение математических знаний. Наконец, давайте представим себе, что вы оказались неправы (это трудно, не правда ли?), и что уровень математических знаний действительно повысился, и произошло это в результате ваших личных усилий, но не имеет никакого отношения к рассматриваемой компьютерной программе. На самом деле, углубление математических знаний произошло по причине того, что преподаватель уделяет больше внимания ученикам, а компьютерная программа не имеет значения. В данном исследовании присутствует внутренняя валидность, потому то вы сделали нечто, повлиявшее на то, что в итоге наблюдается – вы создали причину, заставившую *нечто* произойти. Однако, в данном исследовании отсутствует валидность конструкторов, а именно, конструктор «компьютерная программа по математике» неточным образом описывает действительную причину, кото-

рую, вероятно, точнее было бы обозначить как «личное внимание со стороны преподавателя».

Поскольку центральным вопросом внутренней валидности оказывается вопрос о причине, то мы начнем с рассмотрения того, какие условия необходимо соблюдать для того, чтобы в исследовательском проекте установить каузальную взаимосвязь. Далее мы рассмотрим различные причины, ставящие под угрозу внутреннюю валидность, иначе говоря, какого рода критику встречают утверждения исследователя о том, что он сумел установить влияние программы на результат. Удобства ради, все причины, угрожающие внутренней валидности, будут разделены на три категории:

- ❖ причины, связанные с единственной группой – критика того, когда изучается экспериментальное воздействие лишь на одну группу
- ❖ причины, связанные с множественностью групп – критика исследований с участием несколько групп (например, экспериментальной группы и контрольной группы).
- ❖ социальные причины, угрожающие внутренней валидности – причины, вызванные тем, что психологические исследования проводятся в реальном человеческом контексте, где люди реагируют не только на то, что на них воздействует, но также на то, что происходит с другими окружающими их людьми.

## УСТАНОВЛЕНИЕ ПРИЧИНЫ И СЛЕДСТВИЯ

Каким образом устанавливается причинно-следственная (каузальная) взаимосвязь? Каким критериям необходимо соответствовать? Вообще, существуют три критерия, которым необходимо соответствовать прежде, чем можно будет говорить о том, что имеются подтверждения каузальной взаимосвязи.

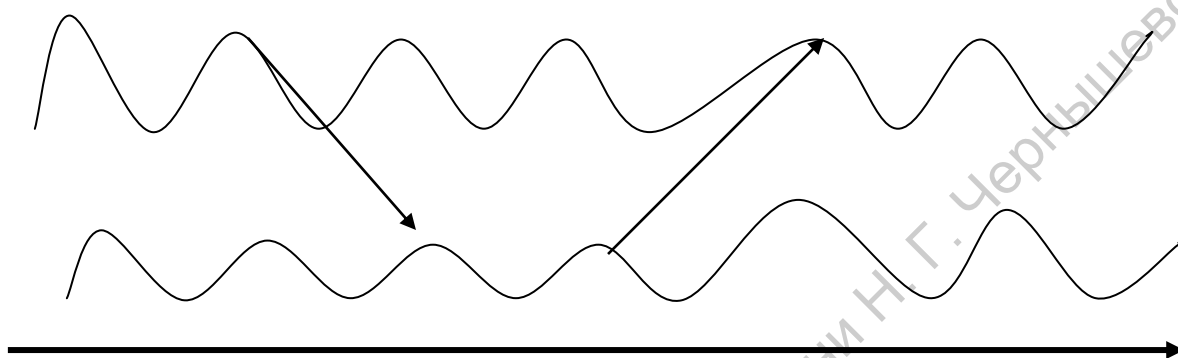
### Предшествование во времени

Прежде всего, необходимо суметь показать, что ваша причина произошла **прежде** вашего эффекта. На первый взгляд, ничего сложного, верно? Разумеется, моя причина произошла раньше эффекта. Вы когда-нибудь слышали, чтобы эффект происходил раньше причины? Пока мы не заблудились в дебрях логических рассуждений, возьмем классический пример из экономики: верно ли, что инфляция вызывает безработицу? На первый взгляд, представляется вполне достоверным, что при увеличении инфляции работодатели, стремясь к уменьшению расходов, увольняют работников. Поэтому инфляция, по крайней мере, отчасти, представляется причиной, приводящей к безработице. Однако, и инфляция, и безработица существуют одновременно. Можем ли мы говорить о том, что флуктуации занятости способны повлиять на инфляцию? Если происходит увеличение рабочей силы (т. е. уменьшение безработицы), то увеличивается потребительский спрос, приводящий к росту цен (т. е. к инфляции), по крайней мере, до тех пор, пока



предложение не удовлетворит увеличение спроса. Так что здесь причина, а что – эффект? В подобных циклических ситуациях оказываются задействованы взаимодействующие друг с другом циклические процессы, каждый из которых может быть и причиной, и следствием для второго процесса. В данной ситуации определение каузальной взаимосвязи оказывается затруднено.

### Циклические флуктуации



### Ковариация причины и следствия

Что это означает? Прежде, чем вы сможете продемонстрировать, что у вас действительно существует каузальная взаимосвязь, вы должны суметь показать, что у вас есть взаимосвязь любого типа. В качестве примера, рассмотрим силлогизм:

**если X, то Y  
если не X, то не Y**

Если наблюдается, что в каждом случае, когда есть X, есть также Y, а когда отсутствует X, то отсутствует также Y, значит, вам удалось продемонстрировать, что между X и Y существует взаимосвязь. Не знаю, как вам, а я иногда прихожу к мысли о том, что рассуждать на языке иксов и игреков не просто. Давайте переформулируем этот силлогизм на языке терапевтического воздействия:

**если терапия, то результат  
если нет терапии, то нет результата**

Либо, на разговорном языке: если вы проводите терапию, то вы наблюдаете уменьшение симптомов, а если не проводите терапию, то не наблюдаете уменьшения симптомов. Это доказывает, что существует взаимосвязь между терапией и уменьшением симптомов. Однако, обратите внимание на то, что этим силлогизмом не доказывается, что уменьшение симптомов происходит благодаря терапии – возможно, что наряду с предпринятой терапией

существует какой-то иной фактор, вызвавший наблюдаемый результат, но не сама терапия. Описанные до сих пор взаимосвязи являются простыми бинарными взаимосвязями. Иногда, желательнее бывает узнать, вызывают ли разные объёмы применения экспериментальной программы различные результаты – это континуальная взаимосвязь:

**если больше терапии, то сильнее уменьшаются симптомы  
если меньше терапии, то слабее уменьшаются симптомы**

### **Отсутствие достоверных альтернативных объяснений**

Только лишь демонстрация существования взаимосвязи сама по себе не означает, что это каузальная взаимосвязь. Возможно, что существует какая-либо другая переменная или фактор, которые вызывают наблюдаемый результат. Его/её иногда называют проблемой «третьей переменной» либо проблемой «пропущенной переменной». *Данная проблема находится в центре проблематики внутренней валидности.* Каковы возможные достоверные альтернативные объяснения? Для ответа на этот вопрос следует обратиться к рассмотрению причин, ставящих под угрозу внутреннюю валидность:

- ❖ причины, связанные с единственной группой
- ❖ причины, связанные с множественностью групп
- ❖ социальные причины

Каждая из указанных причин представляет собой особый тип альтернативного объяснения.

Для того, чтобы иметь основания утверждать, что вами продемонстрировано существование внутренней валидности – то есть, что вами установлено существование каузальной взаимосвязи – необходимо «исключить» альтернативные достоверные объяснения. Как это сделать? Один из главных способов заключается в планировании исследования. Рассмотрим причину, вызванную экспериментированием с единственной группой, причину *историческую*. Предположим, экспериментальная группа была протестирована до начала программы эксперимента, с целью найти линию отсчета. Затем провели экспериментальную программу и снова подвергли группу тестированию по результатам эксперимента. Наблюдается заметное улучшение показателей тестирования, которое исследователю хотелось бы объяснить влиянием экспериментальной программы. Одним из альтернативных достоверных объяснений оказывается историческое – то есть, эффект вызывается не экспериментальной программой, а неким предшествующим конкретным событием. Например, уменьшение количества курильщиков стало не результатом принятой кампании по борьбе с курением, но сократилось в результате предупреждений Минздравсоцразвития, которые испытуемые читали на сигаретах в интервале между началом и завершением эксперимента. Каким образом эту альтернативную причину можно исключить из плана научного исследования? Один из простейших способов заключается в добавлении в эксперимент ещё одной группы – контрольной группы, которую исследователь будет

сравнивать с экспериментальной группой, а единственное различие между экспериментальной группой и контрольной группой заключается в том, что контрольная группа не подвергается воздействию экспериментальной программы. Однако, на сигаретах в контрольной группе тоже написано предупреждение Минздравсоцразвития. Если окажется, что в контрольной группе не наблюдается уменьшение курильщиков, даже несмотря на предупреждения Минздравсоцразвития, значит, предупреждения Минздравсоцразвития действительно исключаются в качестве альтернативного достоверного объяснения того, почему наблюдается уменьшение курильщиков.

В большинстве прикладных психологических исследований, связанных с оцениванием экспериментальных программ, не составляет труда удовлетворить

- *критерию предшествования во времени*, потому что экспериментальная программа проводится прежде, чем измеряется достигнутый эффект. Установить существование
- *ковариации* также достаточно просто, потому что исследователь управляет проводимой экспериментальной программой и может устроить так, чтобы одни люди подвергались экспериментальной программе, а другие бы не подвергались (если есть X и если отсутствует X). Обычно, труднее всего удовлетворить третьему критерию –
- *исключить альтернативные объяснения* наблюдаемого эффекта. По этой причине планирование научного исследования имеет важное значение, и тесно связано с идеей внутренней валидности.

## ПРИЧИНЫ, СВЯЗАННЫЕ С ЕДИНСТВЕННОЙ ГРУППОЙ

### Случай единственной группы

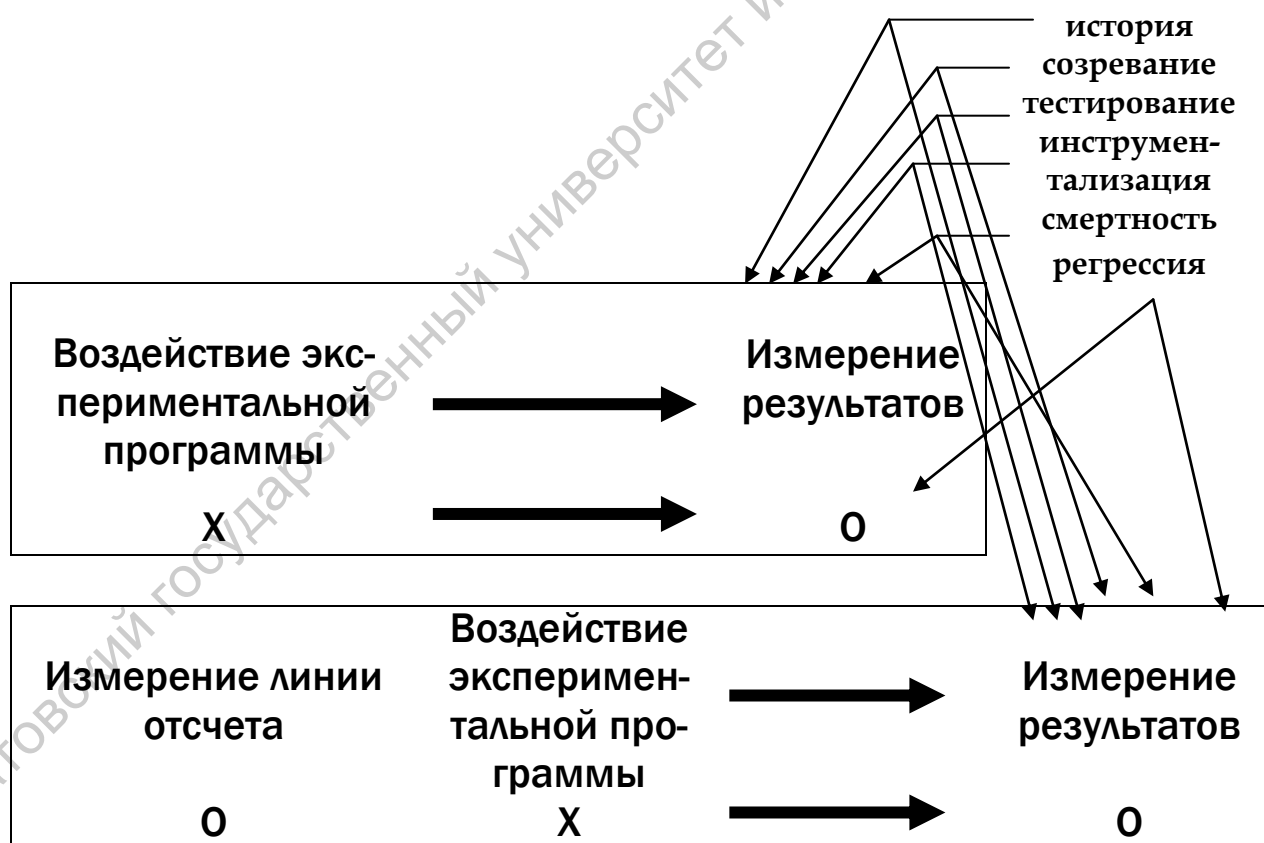
Что понимается под «единственной группой»? Рассмотрим план исследования, состоящий из двух единичных групп, а затем рассмотрим те причины, которые имеют больше всего отношения к внутренней валидности. В верхней части рисунка представлен план «только постэкспериментального» исследования с участием единственной группы. На группу испытуемых оказывается воздействие со стороны экспериментальной программы, а затем проводится постэкспериментальное измерение. В нижней части рисунка представлен план «доэкспериментального/послеэкспериментального» исследования с участием единственной группы. В этом случае проводится измерение испытуемых до эксперимента с целью определить линию отсчета, затем осуществляется экспериментальная программа воздействия на испытуемых, а после этого проводится измерение испытуемых после эксперимента.

Для того, чтобы внести в эту схему больше конкретики, представим себе, что исследуются эффекты программы компенсирующего обучения математике для студентов второго курса с применением таких инструментов из-

мерения математических знаний, как стандартизированный тест математических успехов. При планировании «только постэкспериментального» исследования, к студентам второго курса применяется экспериментальная компенсирующая программа. А после этого проводится постэкспериментальное измерение математических успехов. Если имеются причины полагать, что у испытуемых отсутствуют знания в области математики, то можно не проводить измерения уровня отчета математических знаний. Если имеются основания полагать, что каждый испытуемый получит «ноль» баллов в результате тестирования, то нет смысла проводить тестирование. В случае плана «доэкспериментального/послеэкспериментального» исследования отвергается предположение о том, что знания у испытуемых отсутствуют.

### Два плана с участием единичных групп:

Причины, связанные с единственной группой



Для того, чтобы определить тот уровень, с которого начинаются успехи в изучении математики, необходимо измерить исходный уровень на начало обучения. Гипотетически можно предполагать, что изменения в постэкспериментальном измерении по сравнению с доэкспериментальным измерением, вызываются применяемой нами специальной программой обучения матема-

тике. Это программа компенсирующего обучения, поскольку по ней обучаются только те студенты, у которых с применением некоторых скрининговых механизмов установлены потенциально слабые математические способности.

### **Причины, связанные с единственной группой**

Для любого из названных сценариев необходимо продумать то, что будет происходить в том случае, если наблюдается некоторый уровень математических успехов по окончании воздействия экспериментальной программы, либо если наблюдается изменение между доэкспериментальным уровнем и послеэкспериментальным уровнем математических успехов. Мы стремимся к выводу о том, что результат, который мы наблюдаем, оказался вызван той математической программой, которую мы применили в ходе эксперимента. В чем здесь возможность ошибки с нашей стороны? Рассмотрим некоторые причины, в силу которых внутренняя валидность может быть подвергнута критике – некоторые альтернативные достоверные объяснения наблюдаемого эффекта.

#### **❖ Историческая причина**

Наблюдаемый результат вызван не применением экспериментальной компенсирующей программы математического обучения, но некоторым произошедшим историческим событием. Например, известно, что студенты второго курса смотрели по ТВ программу «Улица Сезам». Известно, что в каждом выпуске этой телепрограммы представлено некоторое элементарное математическое понятие. Возможно, что причина наблюдаемого результата заключается именно в этой телепрограмме, а не в компенсирующей программе математического обучения.

#### **❖ Причина в созревании**

Возможно, что студенты обнаружили бы точно такие же результаты, даже если бы с ними не проводили компенсирующей программы математического обучения. В этом случае оказывается, что исследователь всего лишь измеряет обычное созревание или развитие понимания, которое совершается в качестве составной части общего развития человека, т. е. компенсирующая программа математического обучения не оказывает эффекта. Чем объяснение с точки зрения общего созревания отличается от объяснения с точки зрения исторической причины? Вообще, когда говорят о конкретном событии либо о цепи событий, которые могли вызвать наблюдаемый результат, то это называется исторической причиной. Когда говорят обо всех событиях, которые обычно совершаются в жизни любого человека за определенный период (не уточняя, какие именно из них сыграли причинную роль), то это называется созреванием.

#### **❖ Причина в тестировании**

Эта причина существует только тех исследованиях, которые организованы в соответствии с «доэкспериментальным/послеэкспериментальным» планом. Возможно, что измерение, предшествовавшее экспериментальному воздействию компенсирующей программы математического обучения привело студентов к пониманию того, что прежде вызывало трудности с математикой – эти студенты «дошли до кондиции» во время предварительного тестирования, поэтому, когда началось обучение математике по компенсирующей программе, они оказались к нему готовы, а если бы предварительного тестирования с ними не проводилось, то не было бы и готовности к обучению. Именно это имеется в виду, когда говорят о том, что причина кроется в тестировании – проведение предварительного тестирования до эксперимента (без применения компенсирующей программы математического обучения как таковой) воздействует на то, какие результаты испытуемые обнаружат после применения к ним экспериментальной программы.

### ❖ **Инструментальная причина**

Подобно причине, связанной с тестированием, инструментальная причина также существует лишь в тех исследованиях, которые организованы в соответствии с «доэкспериментальным/послеэкспериментальным» планом. Возможно, различие между доэкспериментальным уровнем и послеэкспериментальным уровнем вызвано не применением компенсирующей программы математического обучения, а различием между тестами, применявшимися до эксперимента, и тестами, применявшимися после эксперимента. Именно это имеется в виду, когда речь идет об инструментальной причине. При проведении повторного тестирования обычно применяют не тот же самый тест (отчасти из-за рассмотренной нами выше причины, содержащейся в тестировании!), но как правило предлагают испытуемым «альтернативную форму» того же теста. Предполагается, что альтернативная форма эквивалентна с точки зрения типов тестовых заданий и с точки зрения уровня трудности, но как быть, если эквивалентность не совершенная? Возможно, различие между доэкспериментальным уровнем и послеэкспериментальным уровнем вызвано изменением измерительного инструментария. Инструментальная проблема наиболее вероятна в тех случаях, когда инструментом выступает человек, проводящий наблюдение. Через какое-то время наблюдатель устает, ему наскучивает наблюдение. И наоборот, чем более человек практикуется, тем лучшим наблюдателем становится. В любом случае, возможно, что наблюдаемый эффект вызывается не экспериментальной программой, а изменением измерительного инструментария.

### ❖ **Причина в смертности**

Смертность не означает, что люди, участвующие в исследовании, умирают (хотя, если такое происходит, то квалифицируется как причина, связанная со смертностью!) В данном случае слово «смертность» употребляется метафорически. Оно означает, что испытуемые «умирают» для исследования.

Как правило, это значит, что они выпадают из исследования. Что в этом плохого? Допустим, что между доэкспериментальным тестированием и послеэкспериментальным тестированием существует некоторый временной интервал. Допустим также, что из исследования «выпадают» те студенты, которые до эксперимента получили низкие баллы по тесту математических способностей. Если посмотреть на среднее различие между тестированием до эксперимента и тестированием после эксперимента по всем балльным показателям для каждого испытуемого, то оказывается, что в итоге тестирования до эксперимента мы включили низкие балльные значения, а в итогах тестирования после эксперимента низкие балльные значения отсутствуют, поскольку испытуемые с низкими баллами покинули наше исследование. Происходит искусственное завышение баллов послеэкспериментального тестирования. Проблему не удастся решить за счет сравнения средних значений до и после эксперимента тех испытуемых, которые всё-таки остались в нашем исследовании. Эта подгруппа испытуемых безусловно окажется неrepresentative о отношению к первоначально составленной выборке испытуемых. более того, известно, что, по причине регрессии, о которой сказано будет далее, результаты оставшихся студентов при тестировании после эксперимента окажутся хуже в силу неслучайного отсева из исследования либо в силу смертности в ходе исследования. Когда причина заключается в смертности, то исследователь нередко способен измерить степень влияния этой причины путем сравнения подгруппы отсева с подгруппой оставшихся испытуемых по показателям измерения до эксперимента. Если между этими подгруппами не выявлено резкого различия, то есть основание полагать, что смертность распределяется по всей выборке испытуемых и не сильно искажает полученные результаты. Однако, если различия по тестированию до эксперимента между этими подгруппами велики, то необходимо решать проблему потенциального искажения результатов исследования по причине смертности.

### ❖ Причина в регрессии

Причина, заключающаяся в регрессии, носит также наименования «регрессионного артефакта» и «регрессии к среднему». Данный статистический феномен возникает каждый раз, когда выборка из генеральной совокупности составлена неслучайным образом, а два измерения не абсолютно коррелируют друг с другом. Попробуем ещё раз с начала. Предположим, что имеются результаты двух измерений – до эксперимента и после эксперимента. Можно ручаться за то, что они коррелируют между собой не абсолютно. Далее, предположим, что показатели измерения до эксперимента в нашей выборке низкие. Регрессионная причина означает, что средний показатель тестирования, выявленный в группе до эксперимента, окажется возросшим (по отношению к генеральной совокупности) даже если с группой ничего не делать, т. е. не оказывать на испытуемых никакого экспериментального воздействия. На первый взгляд, регрессию понять трудно. Попробуем отсюда. Если в экспериментальную программу включить только тех испытуемых, которые по

результатам тестирования до эксперимента составили группу в 10% испытуемых с самыми низкими баллами, то какова вероятность того, что после эксперимента они снова составят группу в 10% испытуемых с самыми низкими баллами? Никакой вероятности нет. У большинства из них после эксперимента окажутся низкие баллы, но мало вероятно, что они снова будут в числе десяти процентов испытуемых с самыми низкими баллами.

Например, допустим, что нескольким испытуемым при тестировании до эксперимента удалось угадать несколько правильных ответов на тестовые задания, и их баллы оказались на уровне одиннадцатого перцентиля. В следующий раз им так сильно не повезет. Но если мы отбираем десять процентов испытуемых с самыми низкими баллами по итогам тестирования до эксперимента, то ниже их баллы уже быть не могут – они могут только повышаться по отношению к той генеральной совокупности, из которой была сформирована выборка. Это явление чисто статистическое, именно это имеют в виду, говоря о регрессии к среднему.

Как быть с перечисленными причинами, ставящими под угрозу внутреннюю валидность? Есть несколько способов, наиболее распространенным из которых оказывается исключение всех этих причин при составлении плана научного исследования. Например, вместо проведения исследования с участием единственной группы, можно включить в исследование также контрольную группу. При таком сценарии получаем две группы: на одну группу оказывается экспериментальное воздействие, а на другую группу экспериментального воздействия не оказывается. В действительности, единственное различие между этими двумя группами должно заключаться в применении программы экспериментального воздействия. Если это требование соблюдается, то контрольная группа испытывает такое же влияние со стороны истории, со стороны созревания, со стороны тестирования, со стороны инструментализации, и обладает такими же показателями смертности и регрессии к среднему. Иными словами, хорошая контрольная группа – это наиболее эффективный способ исключить те причины, которые провоцируют критику внутренней валидности в исследованиях с участием единственной группы. Разумеется, при добавлении контрольной группы план исследования перестает быть ориентированным на единственную группу. Тогда возникают два главных типа проблем, угрожающих внутренней валидности исследования: причины, связанные с множественностью групп и социальные причины.

## **РЕГРЕССИЯ К СРЕДНЕМУ**

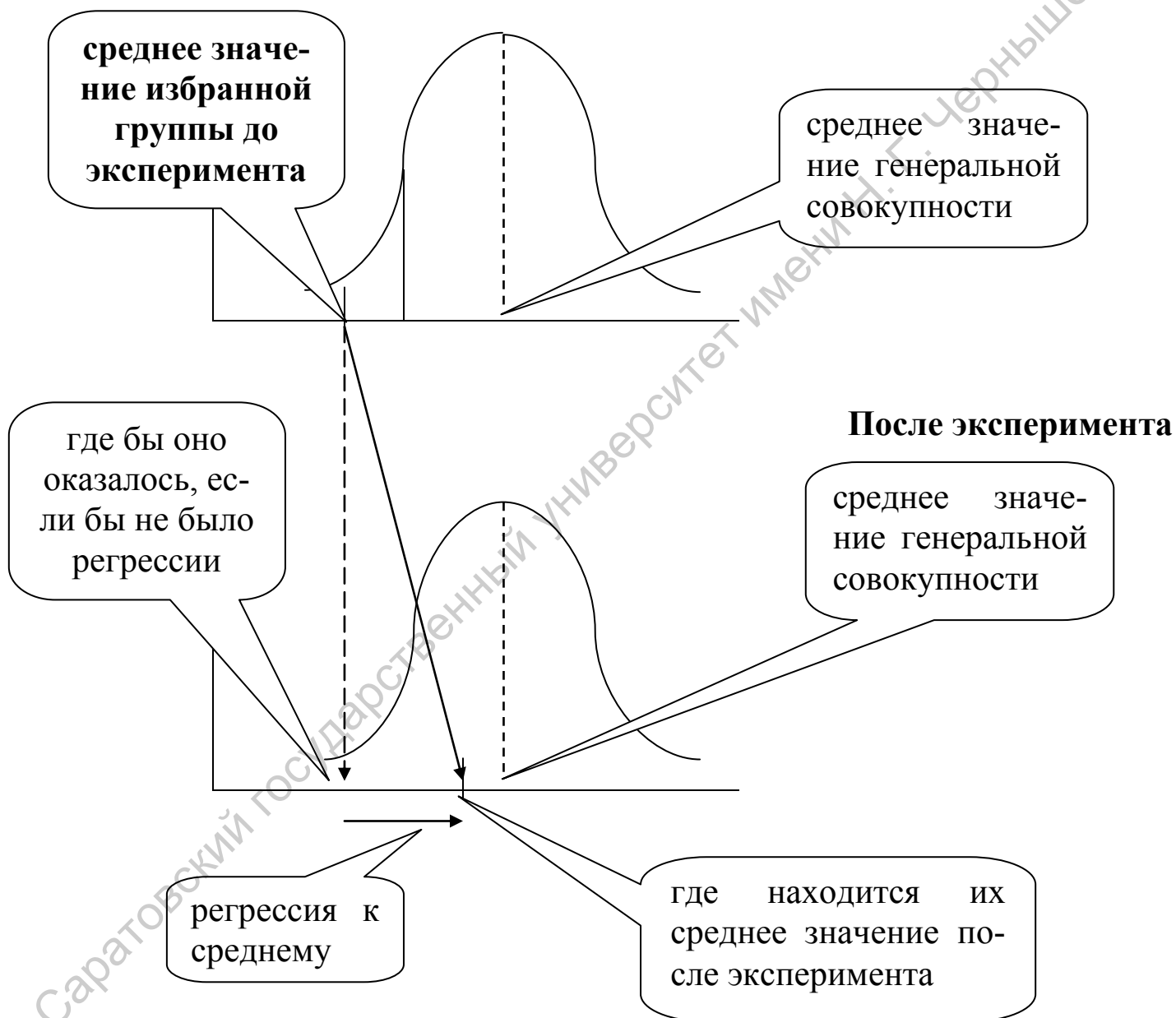
Угроза внутренней валидности со стороны регрессии, носит также наименование «регрессионного артефакта» и «регрессии к среднему». Данный статистический феномен возникает в тех случаях, когда выборка из генеральной совокупности формируется случайным образом, а два измерения неабсолютно коррелируют друг с другом.

На рисунке проиллюстрирован феномен регрессии к среднему. В верхней части рисунка представлено распределение генеральной совокупности до



эксперимента. Баллы по тестированию до эксперимента распределены «нормальным» образом, частотное распределение внешне напоминает «колоколообразную» кривую. Допустим, что выборка в нашем исследовании образована исключительно из тех испытуемых, которые получили низкие баллы при тестировании до эксперимента. В верхней части рисунка показано, где проходит среднее значение до эксперимента – ясно, что среднее значение нашей выборки значительно ниже среднего значения в генеральной совокупности. Какими следует ожидать значения после эксперимента?

### До эксперимента



Прежде всего, предположим, что экспериментальная программа либо терапия, которую применяли к испытуемым, вообще не работает («нулевой» случай). Наше наивное предположение заключается в том, что баллы нашей выборки испытуемых после эксперимента окажутся толь же малыми, как и до эксперимента. Но этого не происходит! В нижней части рисунка показано,

где было бы среднее значение выборки, если бы регрессии не возникало, и где действительно оказывается среднее значение после эксперимента. На самом деле, после эксперимента среднее значение выборки смещается ближе к среднему постэкспериментальному значению для генеральной совокупности, по сравнению с тем, где находилось среднее значение нашей выборки до эксперимента по сравнению с преэкспериментальным средним значением для генеральной совокупности. Иными словами, оказывается, что, при сравнении баллов до и после эксперимента, среднее значение выборки *регрессирует в направлении среднего значения* генеральной совокупности.

### **Почему это происходит?**

Начнем с простого объяснения и от него пойдем дальше. Для того, чтобы понять, почему происходит регрессия к среднему, рассмотрим конкретный случай. Допустим, мы отобрали 10% из генеральной совокупности, исходя из тех баллов по тестированию, которые были получены до эксперимента. Какова вероятность того, что при тестировании после эксперимента нижние десять процентов снова окажутся состоящими из той же самой группы? Никакой вероятности нет. Вероятно, большинство из этой группы и после эксперимента получают низкие баллы, но даже если несколько человек из группы их не получают, то среднее значение в группе окажется ближе к среднему значению генеральной совокупности после эксперимента, чем оно было до эксперимента. То же самое верно и в обратном направлении. Если составить выборку из 10% тех испытуемых, которые получили до эксперимента самые высокие баллы, то мало вероятно, что после эксперимента они снова окажутся в составе 10% испытуемых, получивших наивысшие баллы (даже если большинство из этих испытуемых действительно получают наивысшие баллы). Даже если только несколько человек из этих 10% получают после эксперимента баллы ниже, чем до тестирования, то среднее значение выборки после эксперимента окажется ближе к среднему значению популяции после эксперимента, по сравнению со средним значением до эксперимента.

Вот некоторые вещи, которые необходимо знать о явлении регрессии к среднему:

#### **❖ Это статистическое явление**

Регрессия к среднему происходит по двум причинам. Во-первых, она получается по причине того, что выборка из генеральной совокупности составлена асимметрично. Если составлять выборку случайным образом, то можно будет наблюдать (с учетом случайной ошибки), что среднее значение до эксперимента одно и то же и в генеральной совокупности, и в вашей выборке. Поскольку выборка уже находится на отметке среднего значения генеральной совокупности до эксперимента, то среднее значение самой выборки уже не может далее регрессировать в направлении среднего значения генеральной совокупности!

### ❖ **Это групповое явление**

На основании явления регрессии к среднему невозможно предвидеть того, каким образом сместятся баллы индивидуального испытуемого. Даже с учетом того, что среднее значение группы испытуемых сместится в направлении среднего значения популяции, баллы индивидуальных испытуемых могут сместиться в другом направлении.

### ❖ **Это происходит с любыми двумя переменными**

Вот в чем заключается распространенная исследовательская ошибка. Вы проводите экспериментальную программу, а затем не обнаруживаете совокупного группового эффекта. Поэтому вы принимаете решение рассмотреть группу из тех испытуемых, которые при тестировании после эксперимента показали наилучшие результаты, чтобы выяснить – насколько улучшились их показатели по сравнению с тестированием до эксперимента. Вы отбираете тех испытуемых, у которых баллы самые высокие. Некоторые из них до эксперимента получили не такие высокие баллы (хотя у большинства были высокие баллы и до эксперимента). Поэтому, среднее значение до эксперимента в такой группе должно быть ближе к среднему значению в генеральной совокупности, по сравнению с результатами тестирования после эксперимента. Вы удовлетворены своими успехами, но в этот момент вам предлагают обратить внимание на «неудачников» – на тех испытуемых, которые после тестирования получили самые низкие баллы. Когда вы находите их баллы до эксперимента, то оказывается, что перед экспериментом они не были среди худших. Если они оказываются среди худших до эксперимента и после эксперимента, то следует, с вашей точки зрения, утверждать, что эффект вашей экспериментальной программы отсутствует. Но вы видите, что все намного хуже – похоже на то, что в результате оказанного вами экспериментального воздействия показатели этих испытуемых ухудшились по отношению к генеральной совокупности! Что вы станете делать в этом случае?

В этом случае требуется понять, что описанный только что паттерн результатов происходит всегда, когда измеряются любые две переменные! Он происходит и перспективно (т. е. при переходе от преэкспериментальных данных к постэкспериментальным данным), и ретроактивно (т. е. при переходе от постэкспериментальных данных к преэкспериментальным данным)! Данный паттерн отмечается при переходе от одних данных к другим, которые были получены одновременно (например, рост и вес)! Он отмечается даже если никакого экспериментального воздействия или терапии вообще не оказывается.

### ❖ **Это явление относительное**

Это явление не имеет никакого отношения к общей тенденции к созреванию. Обратите внимание на то, что на вышеприведенном рисунке не выполнена разметка оси X ни до эксперимента, ни после эксперимента. Допустим, разница между баллами до эксперимента и баллами после эксперимента

для любого человека в генеральной совокупности составляет 20 баллов (в среднем). Но регрессия к среднему всё равно будет воздействовать, даже в этом случае. То есть, испытуемые с низкими баллами в среднем улучшат свои баллы более чем на 20 баллов, и тогда их среднее значение окажется ближе к среднему значению генеральной совокупности.

### ❖ Регрессия может быть направлена вверх или вниз

Если ваша выборка состоит из тех, чьи баллы ниже среднего балла в генеральной совокупности, то регрессия к среднему заставит их **подняться** над другими балами. Но если выборка состоит из испытуемых с высокими баллами, то их среднее сместится **вниз** по отношению к генеральной совокупности. (Обратите внимание на то, что, даже если среднее значение повышается, то всё равно существуют отличия от генеральной совокупности. Допустим, в выборке из испытуемых с высокими баллами показатели после эксперимента повысились на 5 баллов, а во всей выборке в целом они повысились на 15 баллов – это дает основание подозревать регрессию к среднему в качестве альтернативного объяснения такого сравнительно малого улучшения).

### ❖ Чем экстремальнее выборочная группа, тем больше регрессия к среднему

Если по первому измерению выборка не сильно отличается от генеральной совокупности, то регрессия к среднему будет невелика, потому что мало места для регрессии – значения в выборке и так близки среднему значению по популяции. Поэтому, когда имеется выборка, пусть и неслучайная, но представляющая собой хорошую подвыборку из генеральной совокупности, то регрессия к среднему не окажет сильного влияния, хотя и будет присутствовать. Но если выборка оказывается экстремальной по отношению к генеральной совокупности (например  $x\%$  самых больших или самых малых значений), то среднее значение выборки оказывается удалено от среднего значения генеральной совокупности, тем самым, возникает больше места для регрессии.

### ❖ Чем меньше корреляция двух переменных, тем сильнее регрессия к среднему

Другой важный фактор, который влияет на объем регрессии к среднему – это корреляция между переменными. Если две переменные коррелируют *абсолютно* – самый высокий балл по одной переменной оказывается самым высоким баллом по другой переменной, следующий за ним балл по одной переменной оказывается следующим за ним баллом по другой переменной и т. д. – то регрессии к среднему не возникнет. Но на практике такое встречается редко. Из теории измерения известно, что не существует такой вещи как «совершенное» измерение – предположительно, все измерения (согласно теории истинного значения) обладают некоторой случайной ошибкой изме-

рения. Случайной ошибки не возникает лишь при полной надежности – в этом случае можно рассчитывать на абсолютную корреляцию. Поскольку в реальности такого не случается, то приходится допускать, что в измерениях присутствует некоторая степень ненадежности, и что взаимосвязи между измерениями будут неабсолютными, значит, между этими двумя измерениями будет возникать регрессия к среднему, с учетом того, что выборочные подгруппы асимметричны.

### **Формула вычисления процента регрессии к среднему**

В любой ситуации можно точно вычислить процент регрессии к среднему. Формула имеет следующий вид:

$$P_{\text{rm}} = 100 (1 - r)$$

где

$P_{\text{rm}}$  = процент регрессии к среднему

$r$  = корреляция между двумя измерениями

Рассмотрим следующие четыре случая:

**если  $r = 1$ , то регрессии к среднему нет (т. е. 0%)**

**если  $r = 0,5$ , то регрессия к среднему составляет 50%**

**если  $r = 0,2$ , то регрессия к среднему составляет 80%**

**если  $r = 0$ , то регрессия к среднему составляет 100%**

В первом случае, две переменные абсолютно коррелируют, а регрессия к среднему отсутствует. При корреляции, равной 0,5, выборочная группа смещается на **пятьдесят процентов** расстояния от той точки, где не существует регрессии к среднему значению генеральной совокупности. Если корреляция мала и равна 0,2, то выборка регрессирует на 80% такого расстояния. А если корреляция между измерениями отсутствует, то выборка в процессе регрессии проделывает весь путь обратно к среднему значению генеральной совокупности! Стоит задуматься над последним случаем. При нулевой корреляции, знание баллов по одному измерению абсолютно не содержит в себе информации о том, какие баллы получил испытуемый по другому измерению. В таком случае единственное разумное предположение о баллах испытуемого по второму измерению состоит в вычислении среднего значения по второму измерению.

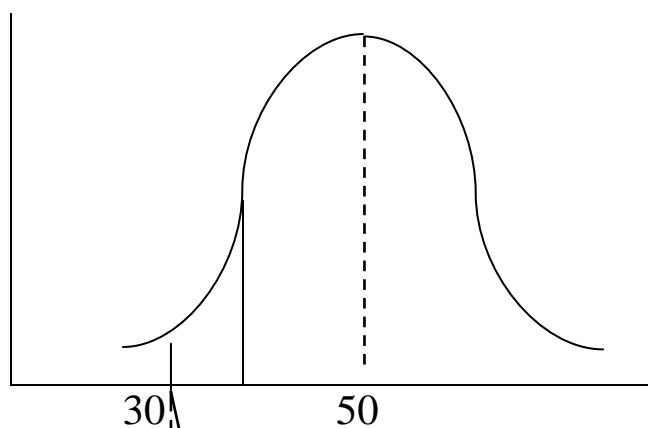
## Вычисление и коррекция регрессии к среднему

Зная формулу вычисления процента регрессии, можно вычислить регрессию к среднему. Все, что нам потребуется знать – это

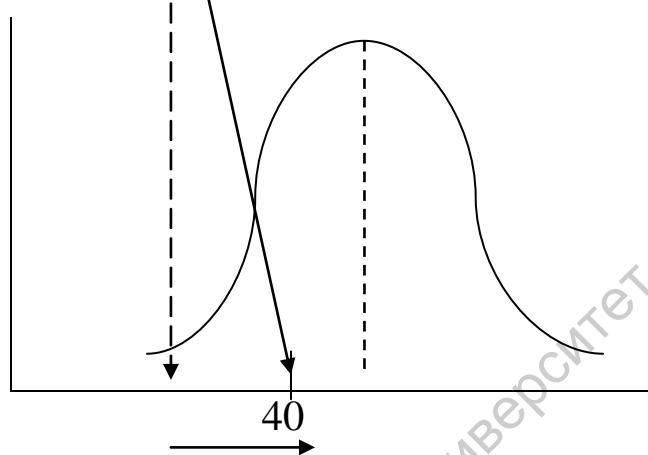
- среднее значение выборки по первому измерению,
- средние значения генеральной совокупности по обоим измерениям, и
- корреляцию между измерениями.

Рассмотрим простой пример. Предположим, что значение для генеральной совокупности до эксперимента равно 50, и что мы включили в выборку испытуемых с низкими баллами по тестированию до эксперимента, среднее значение для выборки при этом составило 30. Для начала, предположим, что мы не оказываем никакого экспериментального воздействия (т. е. «нулевой» случай), и что измеренные характеристики генеральной совокупности со временем не изменяются (т. е. устойчивое состояние). Зная все это, можно ожидать, что среднее значение в генеральной совокупности после эксперимента будет равно 50, а среднее значение выборки после эксперимента будет равно 30, *если не будет регрессии к среднему*.

**До эксперимента**



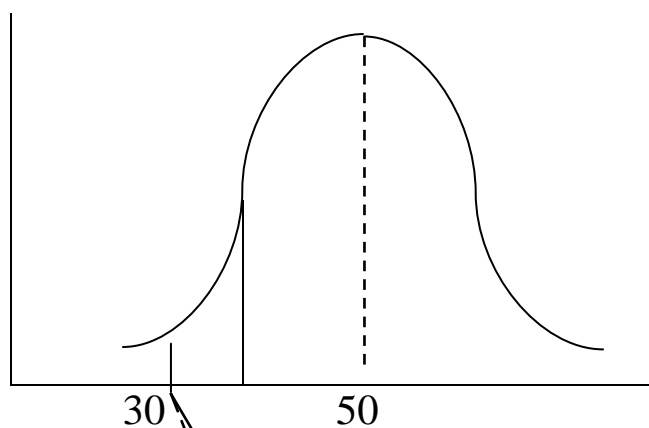
**После эксперимента**



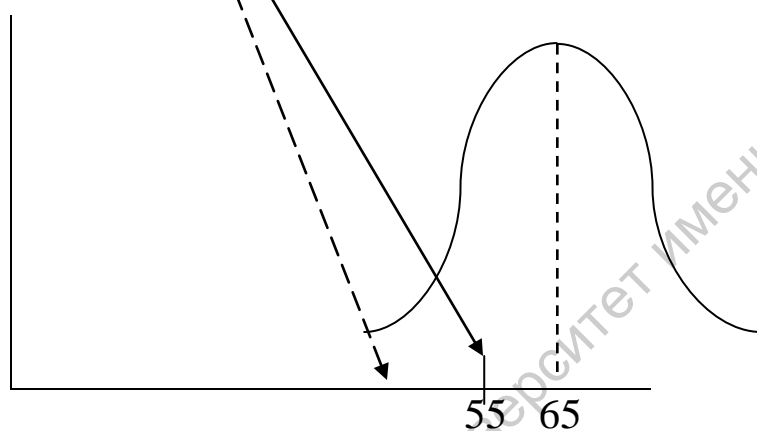
псевдоэффект регрессии к среднему

Далее, допустим, что корреляция между показателями генеральной совокупности до эксперимента и после эксперимента составляет 0,5. Применяя формулу, находим, что регрессия в выборочной группе составляет 50% расстояния от той точки, в которой не существует регрессии к среднему значению генеральной совокупности, или 50% расстояния от 30 баллов к 50 баллам. В таком случае, мы будем наблюдать в выборочной группе среднее значение, равное 40, что составляет 10-балльный псевдоэффект или артефакт регрессии.

**До эксперимента**



**После эксперимента**



псевдоэффект регрессии к среднему

Теперь ослабим первоначальные допущения. Например, допустим, что различие между тестированием до эксперимента и тестированием после эксперимента составило в генеральной совокупности в среднем 15 баллов, и эта разница единообразно распределена во всей генеральной совокупности, т. е. дисперсия генеральной совокупности остается неизменной в обоих случаях тестирования. В данном случае можно ожидать, что выборка, которая по результатам тестирования до эксперимента в среднем получает 30 баллов, после тестирования получит в среднем 45 баллов (т. е.  $30 + 15$ ), если не возникнет регрессии к среднему (т. е. если коэффициент корреляции  $r = 1$ ). Однако, в нашем примере коэффициент корреляции  $r = 0,5$ , поэтому мы рассчитываем на то, что регрессия к среднему покрывает 50% расстояние от среднего значения, равного 45 баллам, до среднего значения в генеральной совокупности после тестирования, которое равно 65. То есть, вопреки псевдоэффекту, равняющемуся 10 баллам, мы наблюдаем среднее значение в нашей выборке после эксперимента, равное 55 баллам.



Регрессия к среднему – это наиболее хитроумная угроза валидности. Она действует исподволь, и даже первоклассному исследователю иногда не удается обнаружить потенциальный артефакт регрессии.

## ПРИЧИНЫ, СВЯЗАННЫЕ С МНОЖЕСТВЕННОСТЬЮ ГРУПП

### Главная проблема

План исследования с привлечением нескольких групп обычно предполагает, что групп не менее двух, и что измерения проводятся до и после эксперимента. Чаще всего, на одну группу оказывается экспериментальное воздействие, а на другую группу экспериментального воздействия не оказывается – это так называемая «контрольная» группа. Но иногда, экспериментальная группа подвергается воздействию экспериментальной программы, а контрольная группа подвергается воздействию какой-либо иной, например, стандартной программы. В этом случае проводится сравнение относительно достигнутых результатов. В типичном случае, план исследования с участием нескольких групп формируется таким образом, чтобы группы можно было бы непосредственно сравнивать одну с другой. В таких планах основная проблема, относящаяся к внутренней валидности, определяется тем, в какой степени группы допускают сравнение. Если они допускают сравнение, то единственное различие между ними заключается в том, какую программу применяют к каждой из них, а различия, наблюдаемые после эксперимента объясняются этой программой. Однако, здесь есть одно большое *если*. Начнем с того, что если группы не допускают сравнения между собой, то нельзя узнать, в какой мере результат вызван применявшейся программой, а в какой мере – исходными различиями между группами.

Когда план исследования включает в себя несколько групп, то существует одна угроза внутренней валидности: группы перед началом эксперимента невозможно сравнивать. Эта причина называется **искаженным отбором**. Искраженный отбор – это любой фактор, кроме фактора экспериментальной программы, приводящий к межгрупповым различиям после эксперимента. Когда мы подозреваем, что межгрупповые результаты различны не о причине экспериментальной программы, но по причине априорных межгрупповых различий, тогда есть основание говорить об искаженном выборе. Несмотря на то, что термин «искаженный отбор» употребляется в качестве категориального обозначения всех априорных различий, тем не менее, когда мы конкретно знаем о том, в чем заключаются различия между группами, то название конкретной причины сопрягается со словом «отбор». Проблемы, связанные с участием в исследовании нескольких групп испытуемых, непосредственно параллельны проблемам, связанным с проведением исследования единичной группы. Например, то, что применительно к единичной группе называлось словом «история», применительно к нескольким группам имеет аналог «отбор-история».

Подобно проблемам, угрожающим внутренней валидности в связи с единичной группой в исследовании, мы рассмотрим простой пример, в котором второкурсников обучают по новой компенсирующей математической программе. План исследования предполагает проведение измерения до эксперимента и после эксперимента, а студентов поделим на две группы – первая обучается по новой программе, а вторая – по прежней программе. В нашем случае возникают следующие проблемы, угрожающие внутренней валидности:

#### ❖ Причина «отбор-история»

Отбором-историей называется любое иное событие в интервале между тестированием до эксперимента и тестированием после эксперимента, которое неодинаково переживается группами. Поскольку эта причина связана с отбором, это означает, что группы некоторым образом отличаются друг от друга. Поскольку эта причина историческая, то это означает, что группы различаются по тому, каким образом они реагируют на историческое событие. Например, дети в двух группах могут различаться своим отношением к телевидению. Например, в экспериментальной группе дети чаще смотрят программу «Улица Сезам», чем дети в контрольной группе. Поскольку в детском шоу «Улица Сезам» детей в интересной форме знакомят с элементарными математическими понятиями, то вполне возможно, что более высокие баллы в экспериментальной группе при тестировании после эксперимента вовсе не говорят об эффекте, оказанном компенсирующей математической программой – они говорят о том, что две группы различно относятся к некоторому релевантному событию, в нашем случае к передаче «Улица Сезам», в интервале между тестированием до эксперимента и тестированием после эксперимента.

#### ❖ Причина «отбор-созревание»

Причина «отбор-созревание» вызывается различным темпом нормального развития групп в интервале между тестированием до эксперимента и тестированием после эксперимента. В этом случае, две группы различаются темпом формирования математических понятий. Важно понимать различие между исторической причиной и причиной, вызванной созреванием. Вообще, история предполагает некоторое отдельно взятое событие или последовательность событий, а созревание предполагает нормально совершающийся процесс развития. В каждом случае, если группы созревают с различным темпом, то нельзя предполагать, что различия, обнаруженные после исследования, вызваны воздействием экспериментальной программы. Эти различия могут быть следствием «отбора-созревания».

#### ❖ Причина «отбор-тестирование»

Причина «отбор-тестирование» возникает тогда, когда возникает *неодинаковый* межгрупповой эффект при тестировании после эксперимента, по

сравнению с данными тестирования до эксперимента. Вероятно, в каждой группе испытуемые по-разному «доходят до кондиции» при тестировании, либо, при тестировании до эксперимента они научились не одному и тому же. В этих случаях, различия, наблюдаемые при тестировании после эксперимента, нельзя объяснить воздействием экспериментальной программы. Эти различия могут явиться следствием причины «отбор-тестирование».

#### ❖ Причина «отбор-инструментализация»

Причина «отбор-инструментализация» обозначает любые неодинаковые изменения в тестировании, которому каждая группа подвергается до и после эксперимента. Иными словами, тестовые задания изменяют неодинаковым образом для двух групп. Допустим, что тестирование заключается в том, что наблюдатели определяют рейтинги успеваемости учеников в классе. Возможно, что в одной группе наблюдатели работают лучше, а в другой группе наблюдатели устали и им надоело. Различия в показателях тестирования после эксперимента легко могут быть вызваны подобной неодинаковой инструментализацией, которая называется «отбор-инструментализация», а не экспериментальной программой.

#### ❖ Причина «отбор-смертность»

Причина «отбор-смертность» возникает там, где имеется неодинаковый неслучайный отсев между началом и окончанием эксперимента. В нашем примере, в каждой группе учащихся мог произойти отсев в силу разных обстоятельств. Различия в результатах тестирования после эксперимента в таком случае объясняются именно отсевом или «отбором-смертностью», а не экспериментальной программой.

#### ❖ Причина «отбор-регрессия»

Наконец, причина «отбор-регрессия» возникает тогда, когда в двух группах существуют неравные степени регрессии к среднему. Это может происходить, если одна группа занимает более экстремальное положение с точки зрения показателей тестирования до эксперимента, чем другая группа. В контексте нашего примера, допустим, что в экспериментальной группе отмечается непропорциональное количество учащихся со слабыми математическими способностями, потому что, по мнению преподавателей, этим детям в большей степени требуется наверстывать математику (а учителя не понимают важности «контрольной» программы и контрольной группы!) Поскольку в экспериментальной группе низкие балльные показатели встречаются крайне часто, то среднее значение экспериментальной группы сильнее отдалено от среднего значения генеральной совокупности, поэтому регрессия к среднему в экспериментальной группе проявится сильнее, по сравнению с контрольной группой. Соответственно, различие между тестированием до эксперимента и тестированием после эксперимента окажется больше в экспериментальной группе, но не благодаря воздействию экспериментальной про-

граммы компенсирующего обучения математике, а благодаря подобному артефакту «отбора-регрессии».

При переходе от плана исследования одной группы к плану исследования нескольких групп, что мы получаем в обмен на увеличение затрат на вторую группу? Если вторая группа – контрольная, и она допускает сравнение с экспериментальной группой, то можно исключить причины, угрожающие внутренней валидности в случае исследования единственной группы, потому что все эти причины отразятся в контрольной группе и уже не смогут служить объяснением того, почему возникают межгрупповые различия после эксперимента. Однако, важно, чтобы эти группы допускали сравнение между собой. Каким образом создать две группы, которые были бы по-настоящему сопоставимы? Единственный известный способ заключается в проведении рандомизированного или «истинного» эксперимента. Однако, в контексте прикладного исследования рандомизация не всегда возможна в силу факторов логистики, этики и иных. Тогда две выборки остаются неслучайным образом, но по возможности максимально эквивалентными. Например, в одном классе проводится экспериментальная компенсирующая программа, а в другом классе, который будет контрольной группой, проводится обычная стандартная программа. В этом случае остается надеяться на то, что обе группы эквивалентны, и можно даже найти причины верить в это. Однако, поскольку неэквивалентность групп тем не менее возможна, и поскольку их вероятностная эквивалентность не проверена надлежащими процедурами, то подобные планы исследований называются квазиэкспериментальными планами. Тестирование подобных групп до эксперимента позволяет определить, подобны ли эти группы друг другу по сопоставимым характеристикам, чтобы далее вынести суждение о том, насколько достоверно существование искаженного отбора.

Но, даже перейдя от плана исследования единичной группы к плану исследования нескольких групп, мы не приобретаем уверенности в том, что располагаем крепкой внутренней валидностью. Существуют также социальные причины, порождаемые тем, что в прикладном психологическом исследовании присутствует человеческое взаимодействие.

## **СОЦИАЛЬНЫЕ ПРИЧИНЫ**

### **Что такое «социальные» причины?**

Прикладные психологические исследования являются одной из сфер человеческой деятельности. На результаты этих исследований воздействуют социальные факторы, связанные с человеческими взаимоотношениями. Социальные причины, угрожающие внутренней валидности, относятся к тем социальным факторам в исследовательском контексте, которые способны вызывать постэкспериментальные различия, непосредственно не связанные с самим экспериментальным воздействием. Большинство этих причин возникают потому, что различные группы (например, экспериментальная группа и

контрольная группа) либо ключевые фигуры, участвующие в исследовании (например, менеджеры, администраторы, преподаватели и т. п.), знают о существовании друг друга и о той роли, которую каждый из них играет в исследовательском проекте, либо находятся в контакте друг с другом. Большинство подобных причин могут быть сведены к минимуму благодаря изолированию двух групп друг от друга, однако, при этом возникают другие проблемы (например, сложно составить случайные выборки, а затем их изолировать; это напоминает редуцирование обобщаемости или внешней валидности). Основные причины, связанные с социальным взаимодействием, следующие:

#### ❖ **Диффузия или имитация экспериментального воздействия**

Эта причина возникает тогда, когда контрольная группа узнает о программе экспериментального воздействия либо непосредственно, либо через вторые руки от участников экспериментальной группы. В школьном контексте, ученики в разных группах в одной и той же школе могут делиться опытом во время обеденного перерыва. Либо ученики из контрольной группы, видя, что делает контрольная группа, могут попробовать подражать контрольной группе. В любом случае, если имитационная диффузия оказывает своё влияние на постэкспериментальные показатели контрольной группы, то под грозой оказывается способность экспериментатора определить, влияет ли программа экспериментального воздействия на результат эксперимента. Обратите внимание на то, что рассматриваемая угроза валидности действует в направлении выравнивания результатов в обеих группах, сводя к минимуму шансы на то, чтобы увидеть эффект экспериментального воздействия, даже если такой эффект действительно существует.

#### ❖ **Компенсирующее соперничество**

В этом случае, контрольная группа знает о том, что делает экспериментальная группа, и проникается по отношению к ней установкой на соперничество. Учащиеся из контрольной группы могут лицедреть специальную программу по математике, которую проходят в экспериментальной группе, и начать завидовать. В результате в контрольной группе возникает стремление превзойти экспериментальную группу: «Мы им покажем, на что мы способны». Иногда, в подобных контекстах, учителя из бюлагих побуждений даже поощряют испытуемых к соперничеству друг с другом. Педагогический смысл подобных побуждений состоит в мотивировании учеников в обеих группах к тому, чтобы лучше учиться, но возможность увидеть эффект экспериментального воздействия при этом снижается. Если соперничество между контрольной и экспериментальной группами влияет на постэкспериментальные показатели, то становится сложнее установить эффект воздействия экспериментальной программы. Подобно диффузии и имитации, данная проблема обычно действует в направлении выравнивания постэкспериментальных показателей в обеих группах, увеличивая вероятность того, что исследова-

тель не увидит эффекта от воздействия экспериментальной программы, даже если экспериментальная программа в действительности обладает эффектом.

#### ❖ Деморализация от обиды

Это почти полная противоположность компенсирующему соперничеству. В этом случае, учащиеся из контрольной группы осведомлены о том, чем занимаются в экспериментальной группе. Однако, вместо стремления к соперничеству, в контрольной группе возникают негодования и злость, и контрольная группа перестает стараться. В отличие от двух названных выше проблем, данная проблема усиливает постэкспериментальные различия между группами, за счет чего программа экспериментального воздействия начинает выглядеть более эффективной, чем она эффективна на самом деле.

#### ❖ Компенсирующее уравнивание экспериментального воздействия

Это единственная из четырех проблем, связанная скорее не с самими испытуемыми, а с теми людьми, которые призваны обеспечивать создание экспериментального контекста. Когда испытуемые из экспериментальной и контрольной групп осведомлены тех условиях, в которых находится каждая группа, то они могут захотеть оказаться в противоположной группе, в зависимости от того, насколько желательной представляется испытуемым программа экспериментального воздействия. Зачастую, либо сами испытуемые, либо их родители или учителя пытаются повлиять на исследователей, чтобы перевести ребенка в другую группу. Экспериментаторы начинают думать, что учащиеся распределены по группам несправедливым образом, и пытаются компенсировать преимущества в каждой группе за счет перевода учеников из группы в группу. Если программ компенсирующего обучения математике преподаются на компьютерах, то можно быть уверенными в том, что родители потребуют «компенсации» для некомпьютеризированной группы учащихся. Такая компенсация может быть предоставлена в виде компьютеризированного преподавания других учебных дисциплин. Если в результате подобных компенсаций постэкспериментальные показатели испытуемых в обеих группах уравниваются, тогда становится затруднительно определить эффективность программы экспериментального воздействия, даже если эффективность действительно существует. Например, в результате предпринятой компенсации самоуважение испытуемых в контрольной группе может возрасти, лишая исследователя возможности определить: вызывает ли специальная программа обучения математике изменения в самоуважении, при сравнении с традиционной программой.

Каждый раз, проводя прикладное психологическое исследование, приходится иметь дело с реальностью человеческих взаимоотношений и их влиянием на процесс исследования. Описанные выше проблемы зачастую удается свести к минимуму за счет формирования нескольких групп, каждая из которых ничего не знает об остальных группах (например, эксперимен-

тальная группа из одной школы, а контрольная группа из другой школы), либо следует инструктировать помощников исследователя о недопустимости перевода испытуемых из одной группы в другую и не разрешать уравнивать программы экспериментального воздействия. Но нам никогда не удастся совершенно устранить влияние человеческих взаимоотношений, затрудняющее оценку причинно-следственных взаимосвязей.

### **Вопросы для самоконтроля по теме «Внутренняя валидность»**

1. Что такое внутренняя валидность?
2. Каким исследованиям релевантна внутренняя валидность?
3. В чем заключается основной вопрос применительно к внутренней валидности?
4. Почему внутренняя валидность характеризуется «нулевой обобщаемостью»? Что это значит?
5. Может ли быть так, что исследование обладает внутренней валидностью, но не обладает валидностью конструкторов? Приведите пример.
6. Какие причины угрожают внутренней валидности? Перечислите и охарактеризуйте.
7. Что такое ковариация причины и следствия?
8. Означает ли демонстрация существования взаимосвязи то, что это каузальная взаимосвязь? Почему?
9. Что такое проблема «третьей переменной»?
10. Для чего необходимо «исключить» альтернативные достоверные объяснения? Как это сделать?
11. Что понимается под «единственной группой»? Приведите пример.
12. В каких случаях возникает регрессия к среднему? Что это такое?
13. Возможно ли на основании явления регрессии к среднему предвидеть то, каким образом сместятся баллы индивидуального испытуемого?
14. Регрессия к среднему происходит перспективно или ретроактивно?
15. Регрессия направлена вверх или вниз?
16. Чем меньше корреляция двух переменных, тем сильнее регрессия к среднему? Почему?
17. По какой формуле вычисляется процент регрессии к среднему?
18. Каков процент регрессии, если корреляция равна 0,5?
19. Что необходимо знать для того, чтобы вычислить регрессию к среднему? Приведите пример.
20. Какая угроза внутренней валидности существует, когда план исследования включает в себя несколько групп?
21. Что такое искаженный отбор?
22. Что такое «социальные» угрозы внутренней валидности? Перечислите их и охарактеризуйте.

## Тестовые задания к теме 5

1. Приблизительная истинность в отношении выводов о причинно-следственных либо о причинных взаимосвязях называется
  - а) валидность конструкторов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность
2. В большинстве исследований, связанных с наблюдением и с описанием, внутренняя валидность
  - а) возможна
  - б) реальна
  - в) допускается
  - г) релевантна
  - д) не релевантна
3. В тех исследованиях, в которых делаются попытки установить каузальные взаимосвязи, внутренняя валидность
  - а) возможна
  - б) реальна
  - в) допускается
  - г) релевантна
  - д) не релевантна
4. Чем характеризуется внутренняя валидность? характеризуется «нулевой обобщаемостью»
  - а) 65% доверительным интервалом
  - б) 95% доверительным интервалом
  - в) 99% доверительным интервалом
  - г) стопроцентной обобщаемостью
  - д) нулевой обобщаемостью
5. Какая валидность подразумевает подтверждения того, что то, что было сделано в данном исследовании (т. е. программа) вызывает то, что наблюдается (т. е. результат)?
  - а) валидность конструкторов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность
6. Какая валидность указывает на то, хотели ли вы выполнить именно ту программу, которую выполнили, либо на то, хотели ли вы наблюдать именно то, что в итоге наблюдается?
  - а) валидность конструкторов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность



- д) конвергентная валидность
7. Какая валидность занимается решением проблемы «третьей переменной»?
- а) валидность конструкторов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность
8. Для того, чтобы иметь основания утверждать, что вами продемонстрировано существование внутренней валидности, необходимо
- а) подобрать альтернативные достоверные объяснения
  - б) разобрать альтернативные достоверные объяснения
  - в) включить альтернативные достоверные объяснения
  - г) исключить альтернативные достоверные объяснения
  - д) обосновать альтернативные достоверные объяснения
9. Допустим, мы отобрали 10% из генеральной совокупности, исходя из тех баллов по тестированию, которые были получены до эксперимента. Какова вероятность того, что при тестировании после эксперимента нижние десять процентов снова окажутся состоящими из той же самой группы?
- а) 100%
  - б) 99%
  - в) 95%
  - г) 65%
  - д) 0%
10. Возможно ли на основании явления регрессии к среднему предвидеть то, каким образом сместятся баллы индивидуального испытуемого?
- а) всегда возможно
  - б) в ряде случаев возможно
  - в) возможно, если приобрести необходимый навык
  - г) возможно в редких случаях
  - д) никогда не возможно
11. Феномен регрессии к среднему возникает
- а) только с психологическими переменными
  - б) только с математическими переменными
  - в) только с независимыми переменными
  - г) только с зависимыми переменными
  - д) с любыми переменными
12. Какое отношения имеет регрессия к среднему к общей тенденции к созреванию?
- а) никакого отношения
  - б) косвенное отношение
  - в) непосредственное отношение
  - г) случайное отношение
  - д) чисто деловое отношение
13. Если ваша выборка состоит из тех, чьи баллы ниже среднего балла в генеральной совокупности, то регрессия к среднему заставит их

- а) опуститься
- б) подняться
- в) замереть
- г) отжаться
- д) исчезнуть

14. Если выборка состоит из испытуемых с высокими баллами, то регрессия к среднему заставит их

- а) опуститься
- б) подняться
- в) замереть
- г) отжаться
- д) исчезнуть

15. Если по первому измерению выборка не сильно отличается от генеральной совокупности, то регрессия к среднему будет

- а) большой
- б) небольшой
- в) средней
- г) достаточной
- д) не возникнет

16. Когда имеется выборка, пусть и неслучайная, но представляющая собой хорошую подвыборку из генеральной совокупности, то регрессия к среднему

- а) не возникнет
- б) не окажет сильного влияния
- в) не окажет никакого влияния
- г) окажет сильное влияние
- д) окажет такое влияние, какое нам нужно

17. Если выборка оказывается экстремальной по отношению к генеральной совокупности, то регрессия к среднему

- а) не возникнет
- б) не окажет сильного влияния
- в) не окажет никакого влияния
- г) окажет сильное влияние
- д) окажет такое влияние, какое нам нужно

18. Если две переменные коррелируют абсолютно, то регрессия к среднему

- а) не возникнет
- б) не окажет сильного влияния
- в) не окажет никакого влияния
- г) окажет сильное влияние
- д) окажет такое влияние, какое нам нужно

19. Насколько смещается выборочная группа, если корреляция  $r = 0,5$ ?

- а) на 100%
- б) на 80%
- в) на 50%
- г) на 20%
- д) на 0%

20. Насколько смещается выборочная группа, если корреляция  $r = 0,2$ ?
- а) на 100%
  - б) на 80%
  - в) на 50%
  - г) на 20%
  - д) на 0%
21. Зная формулу вычисления процента регрессии, можно вычислить
- а) линейную корреляцию
  - б) ранговую корреляцию
  - в) регрессию к среднему
  - г) альфу Кронбаха
  - д) расколотые корреляции

## ТЕМА 6. ВАЛИДНОСТЬ ВЫВОДОВ

В большинстве психологических исследований анализ данных предполагает три главных шага, выполняемых грубо в таком порядке:

- ❖ Организация данных для анализа (подготовка данных)
- ❖ Описание данных (дескриптивная статистика)
- ❖ Проверка гипотез и моделей (статистические выводы)

Подготовка данных предполагает

- ✓ проверку или отслеживание данных;
- ✓ проверку данных на точность;
- ✓ ввод данных в компьютер;
- ✓ преобразование данных;
- ✓ разработку и документирование структуры базы данных, интегрирующую различные измерения.

Дескриптивная статистика применяется для описания базовых свойств данных в исследовании:

- проводится простое обобщение выборки и измеренных показателей;
- проводится простой графический анализ;
- закладывается основа для количественного анализа полученных данных;
- описывается то, что показывают данные.

Статистические выводы предполагают изучение вопросов, моделей и гипотез. Во многих случаях, выводы на основе статистических расчетов намного шире, нежели непосредственно представленные данные. Например, при помощи статистических выводов можно пытаться на основании данных по выборке определить, что думают люди. Либо, при помощи статистических выводов можно высказывать суждения относительно вероятности того, что наблюдаемое различие между группами является от чего-то зависимым, либо, наоборот, случайным. Тем самым, при помощи статистических выводов, можно делать общие выводы относительно данных. В этом состоит отличие

статистических выводов от дескриптивной статистики, которая просто описывает то, на что указывают собранные данные.

В большинстве научных исследований, аналитический раздел содержит эти три фазы анализа:

*Описание того, каким образом были подготовлены к анализу собранные данные*, обычно бывает кратким и сосредоточено на уникальных аспектах проведенного исследования, если такие имеются, например, произведенные преобразования конкретных данных.

*Дескриптивная статистика* обычно располагает к многословию. В большинстве письменных отчетов о проведенных исследованиях, авторы тщательно отбирают дескриптивную статистическую информацию и организуют её в виде обобщающих таблиц и графиков, демонстрирующих то, что наиболее важно в проведенном исследовании.

*Статистические выводы* обычно связываются исследователем с конкретными исследуемыми вопросами либо гипотезами исследования, либо с тестируемыми моделями, возникающими в качестве составной части принятого анализа. В большинстве аналитических письменных работ, важно не допустить той ошибки, когда «за деревьями не видно леса». Это означает, что, когда подробностей слишком много, то читатель может оказаться не в состоянии уследить за центральной идеей представленных результатов. Нередко, анализ, перенасыщенный деталями, оказывается более уместно вынести в приложение, оставив в основном корпусе текста лишь наиболее важные обобщающие аналитические данные.

## ВАЛИДНОСТЬ ВЫВОДОВ

Из четырех типов валидности (внутренняя валидность, валидность конструкторов и внешняя валидность), валидность выводов, несомненно, рассматривается реже всех и вызывает больше всего непонимания. Вероятно, это вызвано тем фактом, что в первоначальном варианте она называется статистической валидностью выводов, а одно лишь упоминание слова «статистический» способно отпугнуть кого угодно.

В большинстве случаев, валидность выводов оказывается самым важным из четырех видов валидности, потому что она релевантна при решении вопроса о том, существует ли взаимосвязь в наших наблюдениях (а это самый базовый аспект любого анализа).

Вероятно, следует начать с попытки дать определение:

**Валидность выводов – это та степень, в которой выводы, к которым мы пришли относительно взаимосвязи в наших данных, обоснованы.**

Например, если мы проводим исследование, в котором рассматриваются взаимоотношения между социально-экономическим статусом и отношением к смертной казни, то мы стремимся прийти к каким-либо выводам. На основе наших данных, мы можем заключить, что существует позитивная взаимосвязь: люди с более высоким социально-экономическим статусом, склонны более положительно относиться к смертной казни, тогда как люди с более

низким социально-экономическим статусом склонны сильнее возражать против смертной казни. Валидность выводов – это та степень, в которой выводы, к которым мы пришли, достоверны либо вероятны.

Несмотря на то, что ранее валидность выводов считалась частью проблемы статистических выводов, теперь очевидно, что она также релевантна качественному исследованию. К примеру, при полевом исследовании подростков-бродяг методом наблюдения, исследователь может, на основании полевых данных, отметить такую закономерность, что подростки, употребляющие наркотики, более вероятно оказываются вовлеченными в сложные социальные сети и взаимодействуют с более разнообразными группами людей. Поскольку такой вывод основан всецело на личных впечатлениях исследователя, можно задать ему вопрос – присутствует ли здесь валидность вывода, т. е. обоснован ли вывод о взаимосвязи в проведенном наблюдении.

При любом исследовании взаимоотношений, существуют два вероятных вывода – либо в полученных данных имеется взаимосвязь, либо она отсутствует. В любом случае, существует возможность сделать неверный вывод. Можно сделать вывод о том, что взаимосвязь существует, при том, что взаимосвязь фактически отсутствует, либо можно сделать вывод о том, что взаимосвязь отсутствует, при том, что фактически взаимосвязь существует, но не была установлена исследователем. Таким образом, когда мы говорим о валидности выводов, мы должны учитывать обе эти вероятности.

Важно осознать, что валидность выводов – это проблема, с которой приходится иметь дело в любом случае, когда следует сделать вывод о существовании взаимосвязи, пусть даже это – взаимосвязь между некоторой программой (например, обучающей), либо системой психотерапии, и каким-то результатом. Иными словами, валидность выводов имеет отношение также к каузальным взаимоотношениям.

В чём отличие *валидности выводов* от *внутренней валидности*, которая также имеет отношение к каузальным отношениям?

Валидность выводов ориентирована только на то, существует ли взаимосвязь или взаимосвязи не существует. К примеру, при оценивании учебной программы, мы можем сделать вывод о том, что существует позитивная взаимосвязь между данной учебной программой и баллами при тестировании по результатам обучения – те, кто обучались по данной программе, получили при тестировании более высокие баллы, а те, кто обучались по другой учебной программе, получили при тестировании баллы ниже. Валидность выводов в данном случае заключается в том, обоснована или не обоснована взаимосвязь на основании имеющихся данных.

Однако, мы можем сделать вывод о том, что, несмотря на наличие взаимосвязи между программой и результатами тестирования, данная программа не является причиной результатов тестирования. Вероятно, что какой-то другой фактор, но не наша обучающая программа, повлиял на результаты тестирования. К примеру, начнем с того, что наблюдаемые различия в результатах тестирования были вызваны тем фактом, что в группе, занимавшейся по нашей обучающей программе, сами по себе студенты более смыш-

ленные, нежели студенты в сравниваемой группе. Наблюдаемые нами посттестовые различия между этими группами могли быть вызваны именно таким исходным различием, а не оказались результатом воздействия обучающей программы.

Данный вопрос – вероятность того, что некоторый другой фактор повлиял на результат эксперимента – это то, с чем имеет дело внутренняя валидность. Таким образом, вполне возможно, что в проводимом нами исследовании учебная программа и результат тестирования действительно взаимосвязаны (валидность выводов), но при этом на результат влияет иной фактор, а не наша учебная программа (т. е. отсутствует внутренняя валидность).

Угрозами валидности выводов называются различные причины, по которым можно сделать ошибочные выводы о том, что взаимосвязь отсутствует, либо о том, что взаимосвязь существует. Существует несколько таких главных причин, по которым бывает затруднительно сделать вывод о взаимосвязи.

- ❖ Одна такая главная проблема состоит в том, что зачастую взаимосвязь бывает трудно увидеть потому, что наши методики и наблюдения обладают низкой надежностью – они слишком слабы, чтобы противостоять средовому «шуму».
- ❖ Другая причина заключается в том, что взаимосвязь, которую мы разыскиваем, может быть слишком слабой, подобно поиску иголки в стогу сена.
- ❖ Иногда проблема заключается в том, что мы не собрали достаточно информации для того, чтобы заметить взаимосвязь, даже если она существует.

Все эти проблемы соотносятся с идеей *статистической мощности*.

## УГРОЗЫ ВАЛИДНОСТИ ВЫВОДОВ

Угрозой валидности выводов называется фактор, способный привести нас к неверному выводу в отношении взаимосвязи полученных данных. Существуют два возможных неверных вывода:

1. вывод о том, что взаимосвязь отсутствует, хотя в действительности она существует (значит, мы не сумели её увидеть);
2. вывод о том, что взаимосвязь существует, хотя в действительности она отсутствует (значит, мы увидели то, чего не было)

Большинство угроз валидности выводов связаны именно с первой проблемой. Почему? Может быть, в большинстве исследований оказывается чрезвычайно трудно установить взаимосвязь, поскольку исследуемая проблема либо не важная, либо редкая. Когда ищешь иголку в стоге сена, внимание оказывается занято настолько, что не замечаешь почти ничего! Итак, давайте разделим угрозы валидности выводов в соответствии с типами ошибок, которые к ним приводят.

## **Взаимосвязь не найдена, несмотря на то, что она существует (или «иголка в стоге сена»)**

Когда вы ищете иголку в стоге сена, вы вынуждены решать две основные проблемы: слишком маленькая иголка и слишком много сена. Можно рассматривать это в качестве проблемы соотношения сигнала-к-шуму. «Сигнал» – это иголка, это та взаимосвязь, которую вы пытаетесь увидеть. «Шум» состоит из всех факторов, которые затрудняют поиск такой взаимосвязи.

Существуют несколько важных источников шума, каждый из которых создает угрозу валидности выводов.

Одной серьезной угрозой оказывается **низкая надежность методик**. Она может быть вызвана многими факторами, включая

- плохую вербальную формулировку вопросов,
- плохое планирование или компоновка инструментария,
- неразборчивость полевых записей, и т. п.

Если в исследовании проводится оценивание программы, то шум может быть привнесен за счет **низкой надежности реализации** такой программы. Если реализация программы проводится не в соответствии с предписанными процедурами, либо осуществляется непоследовательно, это затрудняет нахождение взаимосвязи между программой и другими факторами, такими как результативность программы.

Шум может вызываться **непродуманной несообразностью окружения**, затрудняющей нашу способность увидеть взаимосвязь. В контексте учебной аудитории, уличное движение за окном, помехи в коридоре и другие бесчисленные несообразные события могут отвлекать исследователя или участников.

Типические особенности людей, привлекаемых для проведения исследования, также затрудняют нахождение взаимосвязи. Такая угроза валидности выводов вызывается **непродуманной гетерогенностью респондентов**. Если группа испытуемых слишком разнообразная, то данные измерений будут широко варьировать. По большей части подобное варьирование может соотноситься с тем явлением, которое вы исследуете, но некоторая часть варьирования вызывается индивидуальными различиями, которые не имеют отношения к искомой взаимосвязи.

Все перечисленные угрозы увеличивают вариабельность контекста исследования и увеличивают долю шума по сравнению с долей сигнала, т. е. той взаимосвязи, которую исследователь пытается обнаружить. Однако, шум – это только часть проблемы. Нужно также рассмотреть другую часть – ту, которая связана с сигналом, или насколько прочна взаимосвязь. Существует одна широкая угроза валидности выводов, которая способна поглотить все факторы, производящие шум, и которая учитывает

- ❖ силу сигнала,
- ❖ объем собранной информации,
- ❖ степень риска, на который готов пойти исследователь при принятии решения о наличии или об отсутствии взаимосвязи.

Эта угроза называется **низкой статистической мощностью**. Поскольку эта угроза слишком важна для понимания того, каким образом мы принимаем решения относительно взаимосвязей, мы рассмотрим её отдельно, когда пойдёт речь о статистической мощности.

### **Взаимосвязь найдена, несмотря на то, что её не существует (или «увидели то, чего не было»)**

Почти во всех исследованиях, за исключением наиболее тривиальных, исследователь проводит значительную часть своего времени за анализом данных относительно взаимосвязей. Разумеется, проведение скрупулезного анализа крайне важно, но большинство людей осознают тот факт, что если играть с данными достаточно долго, то может получиться так, что результаты выскочат сами по себе, и эти результаты будут в пользу проверяемой гипотезы. Говоря общепонятным языком, если данные анализировать снова и снова, слегка меняя условия, тогда удастся выловить нужный результат.

В статистическом анализе, мы пытаемся определить вероятность того, что то, что нам удалось обнаружить, реально существует или, наоборот, оказалось случайностью. От степени вероятности будет зависеть то, примем ли мы полученный статистический результат в качестве подтверждения существования взаимосвязи.

В психологии исследователями зачастую применяется совершенно произвольное значение степени вероятности, известное как *0,05 уровень значимости*, на основании которого решается, можно ли считать полученный результат достоверным. В действительности, значение 0,05 означает, что полученный результат имеет вероятность происходить случайным образом по крайней мере 5 раз из каждых 100 раз, подвергаемых статистическому анализу. В основе почти любого статистического анализа лежит допущение о том, что каждый анализируемый случай независим от других. Но это допущение может оказаться неверным, если проводить неоднократный анализ одних и тех же данных.

К примеру, допустим, что мы провели 20 статистических испытаний и в каждом испытании мы применяли 0,05-уровневый критерий определения взаимосвязи. В каждом испытании существует вероятность, равная 5 из 100, что вы увидите взаимосвязь даже если никакой взаимосвязи нет. Именно это имеют в виду, когда говорят о том, что результат оказался случайным.

Вероятность, равная 5 из 100, равна дроби  $5/100$ , что в свою очередь равно  $1/20$ . В нашем примере, вы проводите 20 отдельных анализов. Допустим, оказалось так, что из двадцати результатов, лишь один статистически значим на уровне 0,05. Означает ли это, что мы установили статистически значимую взаимосвязь? Если бы вы проводили только один анализ, тогда можно было бы заключить, что в результате была найдена взаимосвязь. Но если вы провели 20 анализов, тогда один статистически значимый результат из этих 20 можно считать случайным. Такая угроза валидности выводов называется **проблемой степени ошибки**. Эта проблема состоит в том, что вы



пытались «поймать рыбку», неоднократно подвергая данные анализу и рассматривая каждый анализ как независимый от других анализов. Напротив, когда данные анализируют много раз, следует оценивать степень вероятности ошибки (т. е. уровень значимости) с учетом количества проведенных анализов. Основная идея здесь заключается в том, что при повторном анализе мы склонны увидеть взаимосвязь там, где она отсутствует.

### **Проблемы, приводящие к обоим ошибкам в выводах**

Любой анализ основан на множестве допущений относительно (1) природы данных, (2) процедур, применяемых при проведении анализа, и (3) соответствия между данными и процедурами. Если не принимать в расчет предположения, на которых основывается статистический анализ, тогда вполне вероятны ошибочные выводы относительно взаимосвязей. В количественных исследованиях подобная угроза называется **нарушение предположений о статистических тестах**. К примеру, в статистическом анализе часто допускается, что данные имеют нормальное распределение – что популяция, из которой извлечена выборка, распределяется в соответствии с кривой нормального распределения или колоколообразной кривой. Если в отношении анализируемых данных это предположение оказывается неверным, то есть вероятность некорректной оценки истинной взаимосвязи. Не всегда возможно определить то, какая именно ошибка допущена – то ли найдена взаимосвязь, которой не существует, то ли не найдена взаимосвязь, которая существует.

Та же самая проблема существует и в качественных исследованиях. В качественных методах есть свои допущения. Например, в ситуации интервью можно допустить, что респондент свободен говорить что пожелает. Если это неверно – если на респондента оказывается давление тем или иным образом – тогда можно ошибочно увидеть такие взаимосвязи в ответах, которых не существует, и/или пропустить те взаимосвязи, которые существуют.

Вышеперечисленные угрозы иллюстрируют некоторые основные трудности и ловушки, связанные с одной из самых основополагающих исследовательских задач – принятии решения о взаимосвязи в данных или в результатах наблюдения. Как справляться с этими трудностями? Существуют *стратегии улучшения валидности выводов*, позволяющие свести к минимуму или вообще устранить описанные выше трудности.

### **УЛУЧШЕНИЕ ВАЛИДНОСТИ ВЫВОДОВ**

Итак, вы столкнулись с проблемой подтверждения достоверности сделанных вами выводов. Что теперь делать? Вот несколько общих указаний, которым можно последовать при планировании исследования, помогающих улучшить валидность выводов.

## Указания по улучшению валидности выводов

**Хорошая статистическая мощность.** В психологических исследованиях эмпирически установлено, что значение статистической мощности должно быть не менее 0,8. Это означает, что у вас должно быть не менее 80 шансов из 100 на обнаружение взаимосвязи. Существуют несколько факторов, влияющих на мощность.

- 1) **Собрать больше информации** – применить выборку большого объема. При этом необходимо учесть затраты во времени и расходы на формирование большей выборки – выгодно ли это исследователю.
- 2) **Увеличить риск сведения к минимуму ошибки первого типа.** Ошибка первого типа (или ошибка типа «пропуск цели») заключается в том, что мы отвергаем нулевую гипотезу, поскольку статистика принимает значение, принадлежащее критической области, в то время, как нулевая гипотеза верна. На практике это означает повысить альфа-уровень. Например, вместо 0,05 уровня значимости, можно использовать 0,10 уровень значимости в качестве точки отсекаания.
- 3) Можно **увеличить размер эффекта.** Размер эффекта – это отношение сигнала о взаимосвязи к контекстуальному шуму. В этой связи существуют две общих стратегии.
  - а) Чтобы **усилить сигнал**, нужно сделать взаимосвязь более заметной. Эта стратегия применяется при исследовании экспериментального контекста, в котором изучается эффект программы (образовательной или терапевтической). Если увеличить дозировку программы (например, увеличить количество часов, проводимых за обучением, или увеличить количество терапевтических сеансов), тогда эффект окажется заметнее.
  - б) Другая стратегия состоит в том, чтобы **уменьшить шум**, или, иными словами, увеличить надежность.

**Хорошая надежность.** Надежность связана с идеей о том, что шум создает «ошибку», затрудняющую возможность увидеть взаимосвязь. Надежность можно улучшить,

- 1) усовершенствовав измерительную методику
- 2) увеличив количество вопросов по шкале
- 3) уменьшив ситуативные факторы, отвлекающие испытуемых в контексте измерения.

**Хорошая реализация.** Когда вы исследуете эффект от интервенции, от терапии или от программы, вы можете улучшить валидность выводов за счет хорошей реализации исследования. Для этого проводится обучение операторов учебных программ и стандартизация протоколов проведения программ.

## СТАТИСТИЧЕСКАЯ МОЩНОСТЬ

Существуют четыре взаимосвязанных компонента, влияющих на то, какие выводы делаются на основе статистических тестов в исследовательском проекте:

- 1) **размер выборки**, или количество единиц (например, количество человек) в исследовании;
- 2) **размер эффекта**, или то, насколько сигнал заметен на фоне шума;
- 3) **альфа-уровень** ( $\alpha$  или уровень значимости), или вероятность того, что наблюдаемый результат оказался случайным;
- 4) **мощность**, или вероятность того, что вы наблюдаете эффект тогда, когда он происходит.

Если известны значения для трех из перечисленных компонентов, то возможно рассчитать значение четвертого компонента. К примеру, необходимо определить разумный объем выборки для исследования. Если есть возможность определить размер эффекта, альфа-уровень и мощность, тогда размер выборки можно легко рассчитать, или, что еще легче, найти по таблице.

Одними компонентами легче манипулировать, по сравнению с другими, в зависимости от обстоятельств исследовательского проекта. Например, если проект предусматривает оценивание образовательной программы или психолого-консультативной программы для конкретного количества потребителей программы, значит, размер выборки уже предопределен.

Либо, допустим, что дозировка лекарства, назначаемого в рамках программы терапии, должна быть малой по причине потенциальных негативных побочных эффектов этого лекарства, тогда размер эффекта будет соответственно мал.

Цель заключается в достижении такого баланса между этими четырьмя компонентами, который обеспечил бы максимальный уровень мощности для определения эффекта, если эффект существует, с учетом программных, логистических и финансовых ограничений в отношении остальных компонентов.

Все статистические выводы связаны с конструированием двух взаимно исключающих гипотез, которые называются:

- а) нулевая гипотеза (обозначается  $H_0$ )
- б) альтернативная гипотеза (обозначается  $H_1$ )

Взятые вместе, эти две гипотезы описывают все возможные выводы о результатах исследования. Основное решение, которое должен принять исследователь, состоит в том, какую из двух гипотез принять, а какую – отвергнуть. Например, в типичном случае нулевая гипотеза может быть такой:

$$H_0 : \text{Эффект от программы} = 0$$

при такой альтернативной гипотезе:

$$H_1 : \text{Эффект от программы} < > 0$$

Нулевая гипотеза названа так потому, что она обычно соотносится со случаем «нет различий» или «нет эффекта». Обычно, в психологических исследованиях ожидается, что исследуемая программа или терапия создает различия. Поэтому в типичном случае теории описываются с использованием альтернативных гипотез.

Следует потратить некоторое время на изучение нижеследующей таблицы.

Во-первых, посмотрите на заголовки её столбцов. В них отражается реальность – существует ли реальный эффект, реальное различие, реальная польза. Разумеется, мы не можем уверенно знать того, что происходит в реальности. Однако, поскольку нами сформулированы взаимно исключающие гипотезы, раз одна из гипотез верна, значит другая гипотеза ошибочна.

В первом столбце таблицы 2 x 2 показан тот случай, когда исследуемая программа не обладает эффектом.

Во втором столбце представлен тот случай, когда программа обладает эффектом или создает различие.

В заголовке левого столбца описан мир, в котором мы реально живем. Не зная заранее, что истинно, а что ложно, мы вынуждены принимать решения о том, какую гипотезу считать правильной. В заголовке левого столбца указаны два решения, которые мы можем принять – либо от нашей программы нет никакого эффекта (первая строка таблицы 2 x 2), либо от нашей программы есть эффект (вторая строка).

Теперь рассмотрим ячейки таблицы 2 x 2. Каждая ячейка обозначена греческим символом. Обратите внимание на то, что сумма по столбцам равняется 1 :

$$\alpha + (1 - \alpha) = 1 \quad \text{и} \quad \beta + (1 - \beta) = 1$$

Почему столбцы можно суммировать, а строки суммировать нельзя? Потому что, если один столбец истинный, то другой – наоборот – если от программы есть реальный эффект (правый столбец), то одновременно не бывает так, чтобы этот эффект отсутствовал. Поэтому, вероятность для каждого столбца составляет 1, т. к. две строки в каждом столбце описывают единственно возможное решение (принять или отвергнуть нулевую или альтернативную гипотезу) в каждой возможной реальности.

Ниже греческих символов приведены типичные значения для каждой ячейки. Особое внимание следует обратить на значения в двух нижних ячейках. В психологических исследованиях значение  $\alpha$  как правило выбирается равным 0,05. Сейчас устанавливается новая традиция, допускающая статистическую мощность не менее 0,80. Ниже типичных значений приводятся типичные названия для каждой ячейки (заглавными буквами). Вы может заметить, что две ячейки описывают ошибки – вы сделали ложный вывод, а в двух других ячейках – вы сделали верный вывод.

## МАТРИЦА ПРИНЯТИЯ СТАТИСТИЧЕСКОГО РЕШЕНИЯ

<p>В реальности</p> <p>Какие выводы мы сделали</p>	<p><math>H_0</math> (нулевая гипотеза) верна</p> <p><math>H_1</math> (альтернативная гипотеза) ложна</p>	<p><math>H_0</math> (нулевая гипотеза) ложна</p> <p><math>H_1</math> (альтернативная гипотеза) верна</p>
<p>Мы принимаем нулевую гипотезу (<math>H_0</math>)</p> <p>Мы отвергаем альтернативную гипотезу (<math>H_1</math>)</p> <p>Мы говорим... «Взаимосвязь отсутствует» «Нет различий, нет пользы» «Наша теория неверна»</p>	<p style="text-align: center;"><math>1 - \alpha</math> (например, 0,95)</p> <p style="text-align: center;"><b>УРОВЕНЬ ДОВЕРИТЕЛЬНОСТИ</b></p> <p>Существует вероятность утверждать отсутствие взаимосвязи, различий, пользы, которые фактически отсутствуют</p> <p>Вероятность того, что теория не подтверждается корректным образом</p> <p><i>95 раз из 100, когда эффект отсутствует, мы скажем, что эффекта нет</i></p>	<p style="text-align: center;"><math>\beta</math> (например, 0,20)</p> <p style="text-align: center;"><b>ОШИБКА ВТОРОГО ТИПА</b></p> <p>Существует вероятность утверждать отсутствие взаимосвязи, различий, пользы, которые фактически присутствуют</p> <p>Вероятность того, что теория подтверждается, если она истинна</p> <p><i>20 раз из 100, когда существует эффект, мы утверждаем, что эффекта нет</i></p>

Мы отвергаем нулевую гипотезу ( $H_0$ )	$\alpha$ (например, 0,05)	$1 - \beta$ (например, 0,90)
Мы принимаем альтернативную гипотезу ( $H_1$ )	ОШИБКА ПЕРВОГО ТИПА (УРОВЕНЬ ЗНАЧИМОСТИ)	МОЩНОСТЬ
Мы говорим... «Взаимосвязь существует»	Существует вероятность утверждать наличие взаимосвязи, различий, пользы, которые фактически отсутствуют	Существует вероятность утверждать наличие взаимосвязи, различий, пользы, которые фактически присутствуют
«Есть различия, есть польза»	Вероятность того, что теория подтверждается некорректным образом	Вероятность того, что теория подтверждается корректным образом
«Наша теория верна»	5 раз из 100, когда эффект отсутствует, мы утверждаем, что эффект существует	80 раз из 100, когда эффект присутствует, мы говорим, что эффект существует
	Мы должны стремиться к тому, чтобы ошибка первого типа была минимальной, если мы хотим избежать риска сделать ложное заключение о том, что наша программа эффективно работает	Обычно мы стремимся к тому, чтобы мощность была как можно большей.

Иногда бывает трудно запомнить, какая ошибка первого типа, а какая ошибка – второго типа. Если вы запомните, что ошибка первого типа – это то же самое, что  $\alpha$  или уровень значимости, тогда вы сможете запомнить, что это – вероятность того, что эффект или различие оказались случайными. Люди чаще всего подвергаются ошибке первого типа, потому что они почти все-

гда хотят получить подтверждение того, что разработанная ими программа эффективна. Если они найдут статистическое подтверждение, тогда они станут громко рекламировать свою программу.

С другой стороны, тщательная проверка вероятности ошибки второго типа объясняется тем, что, когда обнаруживается, что программа не обладает презентабельной эффективностью, сразу начинают искать – почему? В этом случае пытаются показать, что мощность слабая, а  $\beta$  большая, т. е. возникает вероятность отрицания эффекта, даже если он всё-таки есть.

После слов, написанных заглавными буквами, приводятся различные варианты описания значения в каждой клетке: один вариант – с точки зрения результатов, а другой вариант – с точки зрения подтверждения теории. Курсивом набран пример того, каким образом количественное значение вербализуется в словах.

Для того, чтобы лучше понять странную взаимосвязь между столбцами, задумайтесь о том, что происходит, если вы пытаетесь увеличить мощность в исследовании. При увеличении мощности, возрастает вероятность обнаружения эффекта, если есть эффект (нижняя строка). Однако, при этом возрастает вероятность допустить ошибку первого типа!

Несмотря на то, что сумма по строкам не дает 1, очевидно, что взаимосвязь между строками все-таки существует. Поскольку мы хотим, чтобы мощность была высокой, а ошибка первого типа – маловероятной, нужно иметь в виду эту взаимосвязь.

Мы часто говорим об « $\alpha$ » и « $\beta$ », употребляя слова «более высокий» и «менее высокий». Например, мы можем говорить о преимуществах более высокого или более низкого  $\alpha$ -уровня в исследовании. Следует быть осторожными с интерпретацией смысла таких высказываний. Когда речь идет о *более высоких*  $\alpha$ -уровнях, это означает, что мы *увеличиваем* вероятность ошибки первого типа. Поэтому, более *низкий*  $\alpha$ -уровень означает, что тестирование проводится *более строгим* образом. Имея в виду все вышесказанное, давайте рассмотрим некоторые общие случаи, очевидные из приведенной таблицы.

- ❖ Чем ниже  $\alpha$ , тем ниже мощность; чем выше  $\alpha$ , тем больше мощность
- ❖ Чем ниже  $\alpha$ , тем меньше вероятность допустить ошибку первого типа (т. е. отвергнуть нулевую гипотезу, когда она справедлива)
- ❖ Чем ниже  $\alpha$ , тем более строгим является тестирование
- ❖  $\alpha$  равная 0,01 (при сравнении с 0,05 и 0,1) означает, что исследователь работал тщательно, он допускает возможность риска ошибиться, отвергая нулевую гипотезу, когда она справедлива, лишь в 1 случае из 100 (т. е. говорит, что эффект есть, когда на самом деле эффекта нет)
- ❖  $\alpha$  равная 0,01 (при сравнении с 0,05 и 0,1) ограничивает вероятность попадания в нижнюю строку таблицы при формулировании вывода о том, что программа обладает эффектом. Это означает, что одновременно у вас и низкая статистическая мощность, и низкая вероятность совершить ошибку первого типа

- ❖  $\alpha$  равная 0,01 означает, что существует 99%-ная вероятность утверждать, что различия отсутствуют, когда различия действительно реально отсутствуют (см. верхнюю левую ячейку)
- ❖ увеличение  $\alpha$  (например, с 0,01 до 0,05 или до 0,1) увеличивает вероятность ошибки первого типа (т. е. утверждения, что различия существуют, когда на самом деле различий не существует) и уменьшает вероятность допустить ошибку второго типа (т. е. утверждения, что различия отсутствуют, когда в действительности различия существуют) и уменьшает строгость тестирования
- ❖ увеличение  $\alpha$  (например, с 0,01 до 0,05 или до 0,1) увеличивает мощность, потому что нулевая гипотеза начинает отвергаться чаще (т. е. принимается альтернативная гипотеза) и, соответственно, если альтернативная гипотеза истинна, то увеличивается вероятность её принятия (т. е. мощность)

### **Вопросы для самоконтроля по теме «Валидность выводов»**

1. Что называется валидностью выводов?
2. В чём отличие валидности выводов от внутренней валидности?
3. Какие причины затрудняют валидность выводов? Перечислите их и охарактеризуйте. С чем соотносятся эти причины?
4. Что называется угрозой валидности выводов? Какие угрозы валидности выводов Вы знаете?
5. Что означает 0,05 уровень значимости? Приведите пример.
6. В чем заключается проблема степени ошибки?
7. На каких допущениях основан научный анализ?
8. Что такое нарушение предположений о статистических тестах? Приведите пример.
9. Меньше какого значения не может быть статистическая мощность? Что это означает?
10. Какие факторы влияют на статистическую мощность?
11. В чем состоит ошибка типа «пропуск цели»?
12. Что такое размер эффекта?
13. Какие стратегии увеличения размера эффекта Вы знаете?
14. Какими способами можно улучшить надежность?
15. Что такое альфа-уровень?
16. С конструированием каких двух взаимно исключающих гипотез связаны все статистические выводы? Как называются эти гипотезы?
17. Почему нулевая гипотеза так называется?
18. В чем отличие ошибки первого типа от ошибки второго типа?
19. При увеличении мощности, возрастает вероятность обнаружения эффекта?
20. Что означает более низкий альфа-уровень?
21. Как связаны альфа-уровень и мощность?



22. Что означает  $\alpha$  равная 0,01?

23. Вероятность ошибки какого типа увеличивается при увеличении  $\alpha$ ?

### Тестовые задания к теме 6

- По сравнению с непосредственно представленными данными, выводы на основе статистических расчетов во многих случаях
  - намного точнее
  - намного приблизительнее
  - намного шире
  - намного уже
  - одинаково
- Какой характер имеют статистические выводы относительно данных?
  - конкретный
  - общий
  - валидный
  - надежный
  - предварительный
- Что позволяет делать с собранными данными дескриптивная статистика, в отличие от статистических выводов?
  - рассчитывать
  - прогнозировать
  - анализировать
  - описывать
  - наблюдать
- Степень обоснованности, достоверности либо вероятности тех выводов, к которым мы пришли относительно взаимосвязи в наших данных, называется
  - валидность конструкторов
  - валидность выводов
  - внешняя валидность
  - внутренняя валидность
  - конвергентная валидность
- Все факторы, которые затрудняют поиск взаимосвязи, которую вы пытаетесь увидеть, называются
  - сигналом
  - кошмаром
  - шумом
  - гамом
  - валидностью
- Значение степени вероятности, на основании которого решается, можно ли считать полученный результат достоверным, равно
  - 3,14
  - 0,1
  - 0,01

г) 0,05

д) 1

7. Значение степени вероятности 0,05 означает, что из каждых 100 раз, подвергаемых статистическому анализу, полученный результат имеет вероятность происходить случайным образом, по крайней мере

а) 100 раз

б) 95 раз

в) 50 раз

г) 5 раз

д) 0 раз

8. В психологических исследованиях эмпирически установлено, что значение статистической мощности должно быть

а) не менее 1

б) не менее 0,8

в) не менее 0,2

г) не более 1

д) равно 0

9. Ошибка, которую мы допускаем, когда отвергаем нулевую гипотезу, поскольку статистика принимает значение, принадлежащее критической области, в то время, как нулевая гипотеза верна, называется

а) роковая ошибка

б) простительная ошибка

в) ошибка первого типа

г) ошибка второго типа

д) медвежья услуга

10. Отношение сигнала о взаимосвязи к контекстуальному шуму называется

а) размер сигнала

б) размер шума

в) размер взаимосвязи

г) размер эффекта

д) размер отношения

11. Что произойдет с сигналом, если сделать взаимосвязь более заметной?

а) сигнал усилится

б) сигнал ослабеет

в) шум усилится

г) шум станет невыносимым

д) ничего не изменится

12. Что произойдет с шумом, если увеличить надежность?

а) сигнал усилится

б) сигнал ослабеет

в) шум усилится

г) шум станет невыносимым

д) ничего не изменится

13. С идеей о том, что шум создает «ошибку», затрудняющую возможность увидеть взаимосвязь, связана

- а) валидность
- б) надежность
- в) погрешность
- г) вероятность
- д) мощность

14. Хорошая реализация исследования позволяет улучшить

- а) валидность
- б) надежность
- в) погрешность
- г) вероятность
- д) мощность

15. Вероятность того, что вы наблюдаете эффект тогда, когда он происходит, называется

- а) валидность
- б) надежность
- в) погрешность
- г) вероятность
- д) мощность

16. Нулевая гипотеза названа так потому, что она обычно соотносится со случаем

- а) нет погрешности
- б) нет искажения
- в) нет надежности
- г) нет эффекта
- д) нет правил

17. В психологических исследованиях значение  $\alpha$  как правило выбирается равным

- а) 3,14
- б) 0,1
- в) 0,01
- г) 0,05
- д) 1

18. Вероятность того, что эффект или различие оказались случайными, называется

- а) роковая ошибка
- б) простительная ошибка
- в) ошибка первого типа
- г) ошибка второго типа
- д) медвежья услуга

19. Вероятность отрицания эффекта, даже если он всё-таки есть, называется

- а) роковая ошибка
- б) простительная ошибка
- в) ошибка первого типа
- г) ошибка второго типа
- д) медвежья услуга

20. Более низкий  $\alpha$ -уровень означает, что тестирование проводится
- а) более строгим образом
  - б) менее строгим образом
  - в) более рациональным
  - г) менее рациональным
  - д) ничего не означает
21. Чем ниже  $\alpha$ , тем ниже
- а) валидность
  - б) надежность
  - в) погрешность
  - г) вероятность
  - д) мощность
22. Чем ниже  $\alpha$ , тем меньше
- а) валидность конструкта
  - б) надежность измерения
  - в) погрешность измерения
  - г) вероятность ошибки
  - д) мощность критерия
23.  $\alpha$  равная 0,01 означает, что вероятность утверждать, что различия отсутствуют, когда различия действительно реально отсутствуют, равна
- а) 0%
  - б) 1%
  - в) 50%
  - г) 99%
  - д) 100%
24. Как влияет увеличение  $\alpha$  на вероятность ошибки первого типа (т. е. утверждения, что различия существуют, когда на самом деле различий не существует)?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет
25. Как влияет увеличение  $\alpha$  на вероятность ошибки второго типа (т. е. утверждения, что различия отсутствуют, когда в действительности различия существуют)?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет
26. Как влияет увеличение  $\alpha$  на строгость тестирования?
- а) нормализует
  - б) увеличивает
  - в) уменьшает

- г) усредняет
  - д) никак не влияет
27. Как влияет увеличение  $\alpha$  на мощность?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет

## ТЕМА 7. ДЕСКРИПТИВНЫЕ СТАТИСТИКИ

### ПОДГОТОВКА ДАННЫХ

Подготовка данных предполагает

- 1) их проверку и отслеживание
- 2) проверку данных на точность
- 3) ввод данных в компьютер
- 4) преобразование данных
- 5) разработка и документирование структуры базы данных, в которой интегрированы различные измерения

### Проверка и отслеживание данных

В любом исследовательском проекте данные могут поступать в разное время из многих различных источников:

- ❖ почтовые опросы
- ❖ закодированные данные интервью
- ❖ претестовые или посттестовые данные
- ❖ данные наблюдений

В любых исследованиях, за исключением простейших, необходимо создать процедуру отслеживания информации и её учета, которая будет выполняться до тех пор, пока вы не будете готовы приступить к масштабному анализу данных. Исследователи различаются по тому, каким образом они предпочитают регистрировать поступающие данные. В большинстве случаев, вам потребуется такая база данных, доступ к которой вы можете получать когда пожелаете проверить, какой объем данных уже накоплен и чего ещё не хватает. Это можно сделать при помощи стандартных программ компьютеризированных баз данных (например, Microsoft Access, Claris Filemaker) или стандартных статистических программ (например, SPSS, SAS, Minitab, Data-desk), которые выдают информацию о статусе данных в виде простого дескриптивного анализа. Оригинальные данные (например, протоколы тестирования) необходимо сохранять на протяжении определенного времени. Большинство профессиональных исследователей хранят такие данные 5 – 7 лет. Если исследование важное или дорогостоящее, тогда данные кладут в архив. Важно иметь возможность проследить результат на основании тех оригинальных бланков, в которые были собраны данные.

## Проверка данных на точность

Когда данные собраны, их следует исследовать на точность. Незамедлительное выполнение проверки данных на точность позволяет определить проблемы и ошибки при составлении выборки. В качестве первичной проверки данных, следует задаться следующими вопросами:

- ❖ Читабельны ли бланки ответов?
- ❖ Даны ли ответы на все важные вопросы?
- ❖ Даны ли полные ответы?
- ❖ Включена ли вся релевантная контекстуальная информация (например, дата, время, место, кто проводил обследование)?

В большинстве психологических исследований важнейшей проблемой становится качество измерения. Качество всего последующего анализа будет зависеть от проверки того, не способствовал ли процесс сбора данных появлению неточностей.

## Формирование структуры базы данных

Структурой базы данных называется то, каким образом вы собираетесь хранить данные для их последующего анализа. Вы можете использовать ту же самую структуру, при помощи которой вы отслеживали собранные данные, либо, при проведении комплексных исследований, может быть сформирована другая структура. Как было отмечено выше, существуют два варианта хранения данных в компьютере – программы баз данных и статистические программы. Программы баз данных обычно труднее освоить, однако, благодаря программам баз данных манипулирование данными становится более гибким.

В любом исследовательском проекте вам потребуется распечатанная **книга кодов**, в которой описываются данные и указывается путь доступа к данным. Как минимум, книга кодов содержит следующие параметры каждой переменной:

- ✓ название переменной
- ✓ описание переменной
- ✓ формат переменной (номер, дата, текст)
- ✓ инструментарий или метод сбора данных
- ✓ собранные данные
- ✓ респондент или группа
- ✓ местонахождение переменной (в базе данных)
- ✓ примечания

Такая книга кодов незаменима для исследовательской группы. Вместе с базой данных, она образует документацию, при помощи которой последующие исследователи, работающие с той же самой базой данных, смогут извлечь из базы данных дополнительную информацию.

## Ввод данных в компьютер

Существует множество способов ввода данных для компьютерного анализа. Древнейший способ состоит во вводе данных с клавиатуры. Для того, чтобы обеспечить точность вводимых данных, применяется процедура под названием **двойного ввода**. При этой процедуре данные вводятся впервые. Затем, применяется специальная программа, позволяющая ввести данные вторично, при этом второй ввод сопоставляется с первым вводом. Если выявляется противоречие, тогда программа уведомляет пользователя и дает возможность ввести правильные данные. Подобная процедура двойного ввода в значительной степени уменьшает ошибки при вводе данных. Однако, подобные программы двойного ввода не всем доступны и требуют специальной подготовки для работы с ними.

Альтернативой оказывается первичный ввод данных, а затем запуск процедуры проверки данных на точность. Например, можно произвольно просматривать введенные данные в случайном порядке. После того, как данные введены, далее используются различные программы обобщения данных, позволяющие проверить, находятся ли введенные данные в пределах допустимых границ. К примеру, такие обобщающие программы позволяют определить людей, возраст которых 601 год, или когда при интервале значений от 1 до 5 оказалось введено значение 7.

## Преобразования данных

После того, как данные введены, почти всегда приходится преобразовывать сырые данные в пригодные для анализа переменные. Существует много преобразований, которые могут быть выполнены. К наиболее распространенным преобразованиям относятся следующие:

**Пропущенные значения.** Многие аналитические компьютерные программы автоматически считают отсутствующие значения пропущенными. В других программах необходимо указывать конкретное значение, которое будет считаться пропущенным. Например, таким значением можно считать  $-99$ . Для выбора того, как обращаться с пропущенными значениями, необходимо учитывать особенности компьютерной программы.

**Обращение пункта.** При работе со шкалами, иногда применяется обращение отдельных пунктов шкалы, помогающее уменьшить вероятную заданность ответов. В процессе анализа данных, мы стремимся к тому, чтобы все отметки на шкале располагались в одном направлении, т. е. чтобы высокие значения означали одно, а низкие значения чтобы означали другое. В таких случаях, по отдельным шкалам необходимо совершать обращение значений.

Например, допустим, что у нас имеется пятибалльная шкала измерения самоуважения, на которой 1 означает «совершенно не согласен», а 5 означает «абсолютно согласен». Один из пунктов формулируется таким образом: «Обычно я хорошо отношусь к самому себе». Если респондент абсолютно

согласен, тогда он выбирает отметку 5 баллов, и выбранное значение указывает нам на высокое самоуважение.

Напротив, возьмем другой пункт опросника: «Иногда я чувствую, что как личность мало что из себя представляю». Теперь, если респондент абсолютно согласен и выбирает 5 баллов, тогда это говорит о низком самоуважении. Для того, чтобы два пункта опросника можно было сравнивать друг с другом, нужно совершить обращение баллов по одному из них (скорее всего, мы совершим обращение по второму пункту опросника, потому что высокое значение всегда говорит о высоком самоуважении). Потребуется совершить такую трансформацию, при которой первоначальный 1 балл превратится в 5 баллов, 2 поменяется на 4, 3 останется без изменения, 4 поменяется на 2, а 5 станет 1. Большинство программ допускают совершение подобных преобразований по отдельно взятым субшкальным показателям, тем не менее, существует полезная формула:

$$\text{Новое значение} = (\text{Высокое значение} + 1) - \text{Исходное значение}$$

В нашем примере, высоким значением по шкале оказывается 5, поэтому, чтобы получить новое (преобразованное) значение шкалы, необходимо просто вычесть исходное значение из 6 (т. е.,  $5 + 1$ )

**Суммирование шкал.** После того, как выполнены преобразования отдельных пунктов шкалы, рекомендуется сложить либо посчитать среднее значение по отдельным пунктам, чтобы получить общее значение по шкале.

**Категории.** Рекомендуется объединить отдельные переменные в категории. Например, значения доходов (в рублях), объединенные в категорию, укажут диапазон или размах доходов.

## ДЕСКРИПТИВНЫЕ СТАТИСТИКИ

Дескриптивные статистики применяются для описания основных характеристик исследуемых данных. Они позволяют провести простое обобщение данных по выборке и по методикам измерения. В совокупности с простым графическим анализом, дескриптивные статистики образуют основу любого количественного анализа данных.

Дескриптивные статистики принято отличать от статистических выводов. Дескриптивные статистики просто описывают то, что есть и то, что показывают данные. Статистические выводы выходят за пределы непосредственных данных. К примеру, при помощи статистических выводов мы пытаемся установить то, что считают люди. Либо, статистические выводы позволяют нам судить о том, закономерны или случайны наблюдаемые различия между группами. Итак, статистические выводы относятся к общим условиям, а дескриптивные статистики просто описывают то, что есть в наших данных.

Дескриптивные статистики применяются для представления количественных описаний в удобной форме. В научном исследовании проводится



множество измерений, измерению подвергаются большое количество людей. Deskриптивные статистики помогают нам разумным образом упорядочить большие объемы собранных данных. Каждая deskриптивная статистика сводит множество данных к простому общему виду.

Каждый раз, когда мы пытаемся выразить большое количество отдельных наблюдений при помощи единственного показателя, мы рискуем исказить первичные данные либо пропустить важную деталь. Средний балл, полученный студентом по успеваемости, не говорит нам ничего о том, труден или легок был учебный курс, относится ли учебный курс к специализации или к общеобразовательным дисциплинам. Но, даже с учетом этих ограничений, deskриптивная статистика является мощным инструментом обобщения, благодаря которому возможны сравнения между людьми или иными измеряемыми категориями.

### Одномерный анализ

Одномерный анализ связан с одновременным рассмотрением значений переменной. У отдельно взятой переменной имеются три основные характеристики:

- ❖ распределение
- ❖ центральная тенденция
- ❖ дисперсия

В большинстве исследований, описывают все три эти характеристики для каждой переменной.

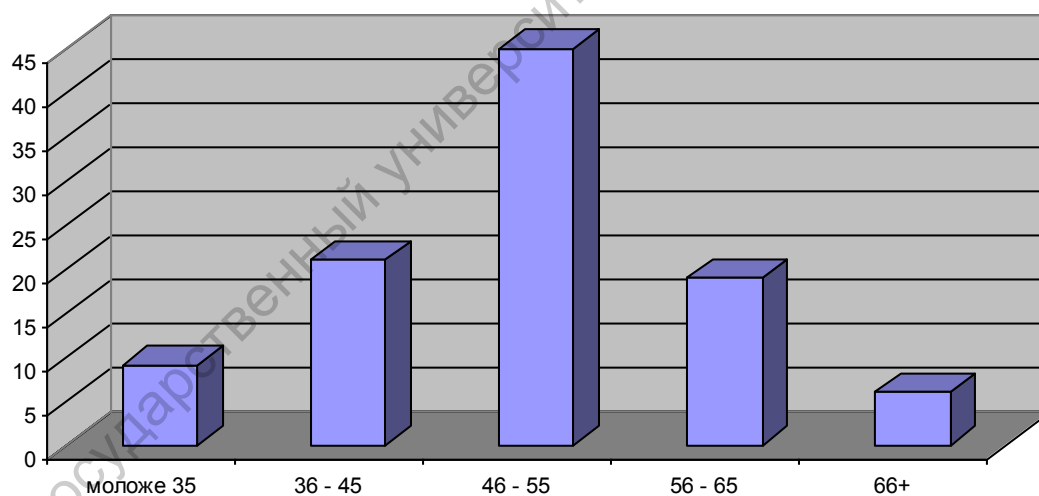
**Распределение.** Распределением называется обобщение частоты отдельных значений или рангов значений переменной. При простейшем распределении перечисляются каждое отдельное значение переменной и количество людей с данным значением. К примеру, типичный образ описания распределения студентов – это год обучения (курс), с указанием количества или процента студентов каждого года обучения (курса). Либо пол описывается путем указания количества или процента мужчин и женщин. В этих случаях у переменной мало значений, поэтому мы можем просто перечислить все значения и подсчитать количество отдельных случаев для каждого значения. Но как быть с такими переменными, как доход или средний балл успеваемости (ЕГЭ)? У подобных переменных большое количество возможных отдельных значений, и каждому значению соответствует небольшое количество людей. В таких случаях, сырые баллы группируются в категории в соответствии с рангами значений. Например, категории ЕГЭ соответствуют каждому отдельному балльному значению. Либо, можно сгруппировать доход в соответствии с четырьмя – пятью рангами значений дохода.

### Таблица частоты распределения

Категория	Процент
Моложе 35	9
36 – 45	21
46 – 55	45
56 – 65	19
66+	6

Одним из наиболее распространенных способов описания единичной переменной является **частота распределения**. В зависимости от конкретного значения, все значения могут быть сгруппированы в категории (например, для переменных «возраст», «цена» или «температура» было бы бессмысленно определять частоту каждого отдельно взятого значения. Лучше поступить наоборот и сгруппировать значения в соответствии с рангом и частотой). Распределение частоты может быть представлено либо в виде таблицы, либо в виде графика.

### Диаграмма частоты распределения



Графики такого типа называют **гистограммами**.

Распределение также можно представить в процентах. К примеру, проценты применяются при описании:

- ❖ процентов людей с различным уровнем дохода
- ❖ процентов людей в различных возрастных диапазонах
- ❖ процентов людей в различных диапазонах баллов по стандартизированным тестам

**Центральная тенденция.** Центральная тенденция распределения – это оценка «центра» распределения значений. Существуют три основных типа оценки центральной тенденции:

- ❖ среднее

- ❖ медиана
- ❖ мода

**Среднее** – наиболее распространенный метод описания центральной тенденции. Чтобы рассчитать среднее, необходимо всего лишь суммировать все значения и поделить сумму на количество значений. К примеру, складываем оценки всех студентов, сдававших экзамен, и делим сумму на количество студентов. Рассмотрим ряд чисел:

**15, 20, 21, 20, 36, 15, 25, 15**

Сумма этих восьми значений равна 167, поэтому среднее  $167:8 = 20,875$ .

**Медиана** – это балл, который находится точно посередине ряда чисел. Один способ расчета медианы заключается в перечислении всех балльных значений по порядку, а затем в нахождении того балла, который окажется в середине ряда. К примеру, если в ряду 500 значений, то значение номер 250 будет медианой. Если мы расположим наши восемь значений, то получим:

**15, 15, 15, 20, 20, 21, 25, 36**

Середина ряда представлена значениями номер 4 и номер 5. Поскольку оба этих значения равны 20, то и медиана равна 20. Если два средних значения неодинаковы, тогда для нахождения медианы придется прибегнуть к интерполированию.

**Модой** называется наиболее распространенное значение в ряду чисел. Для того, чтобы найти моду, значения располагают в ряд как показано выше, а затем считают, сколько раз встретилось каждое значение. Значение, которое встречается чаще всего, и есть мода. В некоторых распределениях бывает более одного модального значения. Например, в бимодальном распределении имеются два значения, которые встречаются чаще других значений.

Итак, для одного ряда из восьми чисел, мы получили три значения:

Среднее = 20,875

Медиана = 20

Мода = 15

Если распределение является нормальным (т. е. колоколообразным), то среднее, медиана и мода равны друг другу.

**Дисперсия.** Дисперсией называют распределение значений вокруг центральной тенденции. Существуют два общих показателя дисперсии: размах (диапазон) и стандартное отклонение.

**Размах** – это максимальное значение минус минимальное значение. В примере с нашим распределением, максимальное значение 36, а минимальное значение 15, поэтому размах

$$36 - 15 = 21.$$

**Стандартное (среднеквадратичное) отклонение** – это более точный показатель дисперсии, потому что одно случайное значение, например 36, способно существенно увеличить размах. Стандартное отклонение показывает отношение ряда чисел к среднему по выборке. Снова возьмем ряд чисел:

**15, 20, 21, 20, 36, 15, 25, 15**

Для того, чтобы рассчитать стандартное отклонение, сначала необходимо найти расстояние между каждым значением и средним. Мы уже знаем, что среднее равно 20,875. Поэтому, вычитая среднее, мы получаем:

$15 - 20,875 = -5,875$	$36 - 20,875 = 15,125$
$20 - 20,875 = -0,875$	$15 - 20,875 = -5,875$
$21 - 20,875 = +0,125$	$25 - 20,875 = +4,125$
$20 - 20,875 = -0,875$	$15 - 20,875 = -5,875$

Обратите внимание на то, что когда числовое значение меньше среднего, тогда разность получается отрицательной, а когда числовое значение больше среднего, тогда разность получается положительной. Далее, все разности возводим в квадрат:

$(-5,875)^2 = 34,515625$	$(15,125)^2 = 228,765625$
$(-0,875)^2 = 0,765625$	$(-5,875)^2 = 34,515625$
$(+0,125)^2 = 0,015625$	$(+4,125)^2 = 17,015625$
$(-0,875)^2 = 0,765625$	$(-5,875)^2 = 34,515625$

Теперь, нам нужно взять эти «квадраты» и сложить их, чтобы получить сумму квадратов. Сумма квадратов равна 350,875. Разделим эту сумму на количество значений минус 1. Результат составляет  $350,875 : 7 = 50,125$ . Это значение называется **дисперсия**. Для того, чтобы получить стандартное отклонение, необходимо извлечь из дисперсии квадратный корень (поскольку мы все отклонения возводили в квадрат):

$$\sqrt{50,125} = 7,079901129253$$

И хотя эти расчеты могут показаться сложными, в действительности они весьма просты. Это видно из формулы стандартного отклонения:

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{(n - 1)}}$$

где:

- $x_i$  = любое значение
- $\bar{x}$  = среднее
- $n$  = количество значений
- $\Sigma$  = знак суммы

В верхней части дроби, в числителе, мы видим, что из каждого частного значения вычитается среднее значение, разности возводятся в квадрат, а квадраты суммируются. В знаменателе из количества значений вычитается 1. Дробь представляет собой дисперсию, а квадратный корень из дисперсии – это стандартное отклонение. По-русски, стандартное отклонение можно записать так:

**корень квадратный из суммы квадратов отклонений от среднего, поделенный на количество значений минус единица**

Эту статистику можно рассчитать при помощи ручки и листа бумаги, однако, когда значений довольно много, расчеты становятся утомительными. С ними легко справляется любая статистическая программа. Например, вводим восемь чисел в SPSS и получаем результат в виде следующей таблицы:

количество	8
среднее	20,8750
медиана	20,0000
мода	15,00
стандартное отклонение	7,0799
дисперсия	50,1250
размах	21,00

что соответствует расчетам, выполненным вручную.

Стандартное отклонение позволяет нам сделать некоторые выводы относительно конкретных значений в нашем распределении. Предположив, что распределение баллов является нормальным или колоколообразным (либо близким к нему!), можно сделать следующие выводы:

- ❖ приблизительно 69% значений в выборке находятся в пределах одного стандартного отклонения от среднего
- ❖ приблизительно 95% значений в выборке находятся в пределах двух стандартных отклонений от среднего

- ❖ приблизительно 99% значений в выборке находятся в пределах трёх стандартных отклонений от среднего

К примеру, поскольку среднее в нашем примере равно 20,8750, а стандартное отклонение составляет 7,0799, то на основании вышеприведенных утверждений, мы можем рассчитать, что примерно 95% значений окажутся в пределах от  $20,8750 - (2 * 7,0799)$  до  $20,8750 + (2 * 7,0799)$ , т. е. между 6,7152 и 35,0348. Такая информация имеет решающее значение, поскольку позволяет нам сравнивать показатели испытуемых по данной переменной с показателями этих испытуемых по другим переменным, даже если эти переменные измеряются по совершенно различным шкалам.

### **Вопросы для самоконтроля по теме «Дескриптивные статистики»**

1. Из чего складывается подготовка данных для анализа?
2. Назовите и охарактеризуйте наиболее распространенные преобразования данных.
3. Что такое обращение баллов? По какой формуле оно рассчитывается?
4. Для чего применяются дескриптивные статистики?
5. В чем отличие дескриптивных статистик от статистических выводов?
6. Назовите три основные характеристики каждой отдельно взятой переменной.
7. Что называется распределением? Приведите пример.
8. Что такое частота распределения?
9. В каком виде можно представить распределение частоты?
10. Что такое центральная тенденция распределения?
11. Что такое медиана?
12. Что называется в статистике модой?
13. Что такое дисперсия? Сколько показателей может быть у дисперсии? Какой из них является более точным?
14. Как рассчитать дисперсию?
15. Какие выводы можно сделать из предположения о нормальном распределении баллов?

### **Тестовые задания к теме 7**

1. Способ хранения данных в базе для их последующего анализа образует её
  - а) систему
  - б) структуру
  - в) функцию
  - г) разновидность
  - д) комплекс
2. Описание данных и путь доступа к данным указывается
  - а) в книге чисел
  - б) в книге перемен

- в) в книге кодов
  - г) в книге жалоб
  - д) в красной книге
3. Основу любого количественного анализа данных образуют
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
4. Судить о том, закономерны или случайны наблюдаемые различия между группами, нам позволяют
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
5. Описывают то, что есть и то, что показывают данные
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
6. Для представления количественных описаний в удобной форме, применяются
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
7. Помогают нам разумным образом упорядочить большие объемы собранных данных
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
8. Обобщение частоты отдельных значений или рангов значений переменной, называется
- а) размышление
  - б) наблюдение
  - в) распределение
  - г) валидность
  - д) надежность
9. Оценка «центра» распределения значений, называется
- а) выраженная тенденция

- б) сглаженная тенденция
  - в) оптимальная тенденция
  - г) центральная тенденция
  - д) современная тенденция
10. Балл, который находится точно посередине ряда чисел, называется
- а) среднее
  - б) мода
  - в) медиана
  - г) дисперсия
  - д) ранг
11. Если в ряду 500 значений, то медианой будет
- а) значение номер 0
  - б) значение номер 1
  - в) значение номер 250
  - г) значение номер 251
  - д) значение номер 500
12. Наиболее распространенное значение в ряду чисел, называется
- а) среднее
  - б) мода
  - в) медиана
  - г) дисперсия
  - д) ранг
13. Если распределение является нормальным (т. е. колоколообразным), то среднее, медиана и мода
- а) больше нуля
  - б) в сумме равны нулю
  - в) возводятся в квадрат
  - г) равны друг другу
  - д) не существуют
14. Распределение значений вокруг центральной тенденции называется
- а) дислексия
  - б) депрессия
  - в) дисперсия
  - г) регрессия
  - д) прогрессия
15. Максимальное значение минус минимальное значение называется
- а) промах
  - б) размах
  - в) замах
  - г) max
  - д) min
16. Отношение ряда чисел к среднему по выборке называется
- а) стандартная ошибка
  - б) стандартное отклонение
  - в) стандартное значение



- г) стандартный размер
- д) стандартная оценка

17. Для того, чтобы рассчитать стандартное отклонение, сначала необходимо найти расстояние между каждым значением и

- а) модой
- б) медианой
- в) средним
- г) дисперсией
- д) отклонением

18. Если среднее равно 20,8750, а стандартное отклонение составляет 7,0799, то в пределах между 6,7152 и 35,0348 окажутся примерно

- а) 100% значений
- б) 99% значений
- в) 95% значений
- г) 50% значений
- д) 0% значений

## ТЕМА 8. КОРРЕЛЯЦИЯ

Корреляция – это одна из наиболее распространенных и самых полезных статистик. Корреляция представляет собой одно число, которым описывается степень взаимосвязи между переменными. Давайте проделаем пример, который продемонстрирует то, как рассчитывается эта статистика.

### Пример корреляции

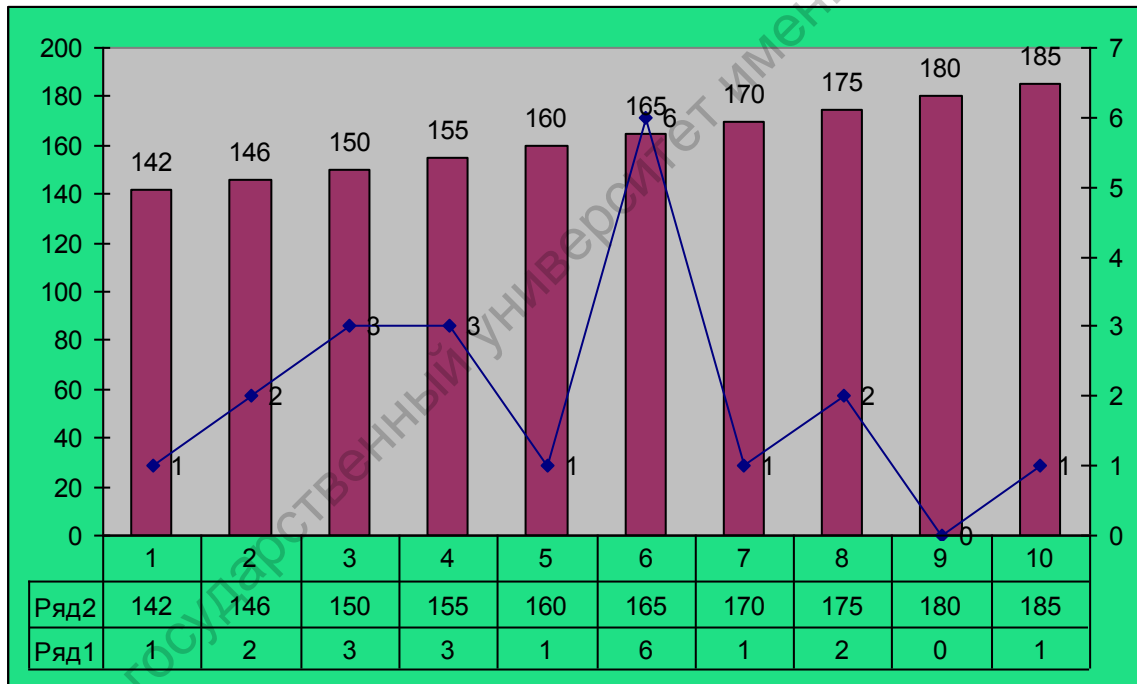
Предположим, что нам необходимо рассмотреть взаимоотношение между двумя переменными: ростом и самоуважением. Допустим, что у нас есть гипотеза, в соответствии с которой рост влияет на самоуважение (направление каузальности здесь самоочевидно – самоуважение не может определять собой рост!). Допустим, мы провели обследование 20 человек, все они мужского пола, потому что известно, что средний рост мужчин и женщин неодинаков, поэтому для простоты исследования мы обследовали только мужчин. Самоуважение было измерено по десяти пунктам опросника, ответы на которые ранжировались по пятибалльной шкале (чем выше балл – тем выше самоуважение). Данные по 20 испытуемым представлены в таблице:

Испытуемый	Рост	Самоуважение
1	166,6	4,1
2	174	4,6
3	152	3,8
4	183,7	4,4
5	142	3,2
6	147	3,1
7	164	3,8
8	166,6	4,1

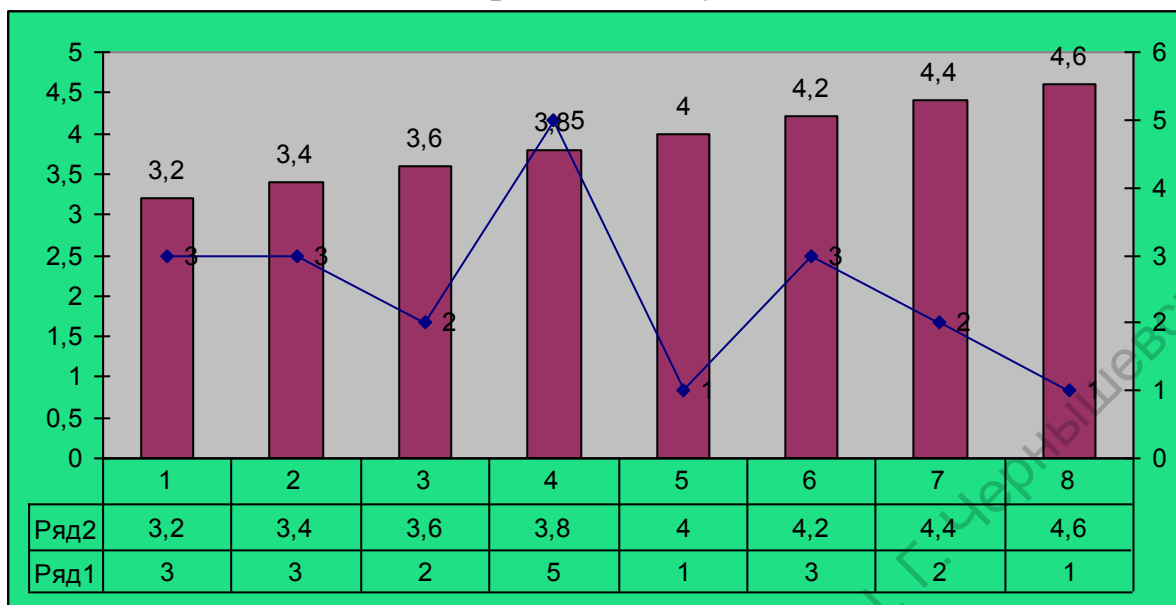
9	174	4,3
10	169	3,7
11	166,6	3,5
12	164	3,2
13	154,4	3,7
14	152	3,3
15	147	3,4
16	154,4	4,0
17	159,3	4,1
18	164	3,8
19	147	3,4
20	149,5	3,6

Давайте рассмотрим гистограмму для каждой переменной.

Гистограмма роста



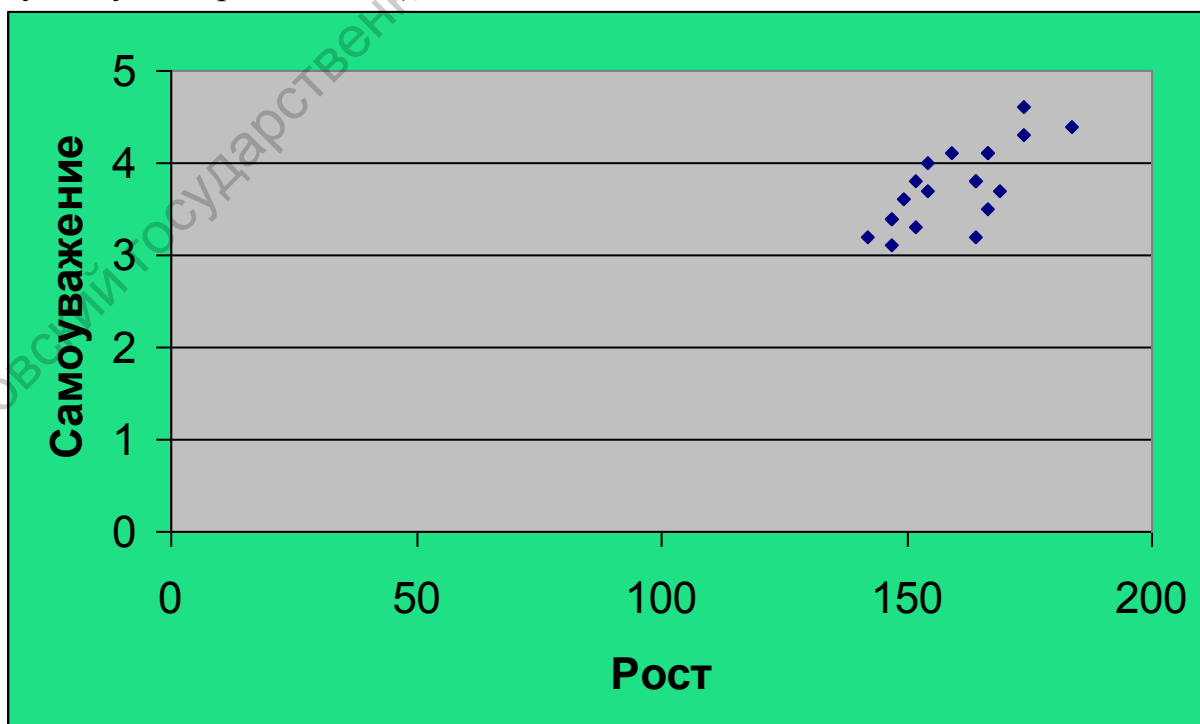
## Гистограмма самоуважения



Данные дескриптивной статистики:

Название переменной	Среднее	Стандартное отклонение	Дисперсия	Сумма	Минимум	Максимум	Размах
Рост	65,4	4,40574	19,4105	1308	58	75	17
Самоуважение	3,755	0,426090	0,181553	75,1	3,1	4,6	1,5

Наконец, посмотрим на простую двумерную диаграмму (т. е. диаграмму с двумя переменными):



Из двумерной диаграммы очевидно, что между переменными существует положительная взаимосвязь, потому что, если через точки провести прямую линию, то она пройдет слева направо, т. е. её наклон будет положительным. Поскольку корреляция есть ни что иное, как количественное выражение взаимосвязи, то можно предполагать, что корреляция окажется положительной.

Что в данном контексте означает выражение «позитивная взаимосвязь»? Оно означает, что более высоким значениям по одной переменной соответствуют более высокие значения по второй переменной, а низким значениям по одной переменной соответствуют низкие значения по второй переменной. Если посмотреть на представленный выше график, то в этом можно убедиться наглядно.

### Расчет корреляции

Теперь мы готовы рассчитать значение корреляции. Формула для расчета корреляции:

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

где:

$N$  = количество пар значений

$\sum xy$  = сумма произведений пар значений

$\sum x$  = сумма значений переменной  $x$

$\sum y$  = сумма значений переменной  $y$

$\sum x^2$  = сумма квадратов значений переменной  $x$

$\sum y^2$  = сумма квадратов значений переменной  $y$

Корреляция обозначается символом  $r$ . По волшебству математики,  $r$  всегда находится в пределах от  $-1,0$  до  $+1,0$ . Если корреляция отрицательная, значит взаимосвязь негативная; если она положительная, значит взаимосвязь позитивная. Расчет корреляции требует знания следующих величин:

Испытуемый	Рост (x)	Самоуважение (y)	x * y	x <sup>2</sup>	y <sup>2</sup>
1	166,6	4,1	683,06	27755,56	16,81
2	174	4,6	800,4	30276	21,16
3	152	3,8	577,6	23104	14,44
4	183,7	4,4	808,28	33745,69	19,36
5	142	3,2	454,4	20164	10,24
6	147	3,1	455,7	21609	9,61
7	164	3,8	623,2	26896	14,44
8	166,6	4,1	683,06	27755,56	16,81
9	174	4,3	748,2	30276	18,49
10	169	3,7	625,3	28561	13,69
11	166,6	3,5	583,1	27755,56	12,25
12	164	3,2	524,8	26896	10,24
13	154,4	3,7	571,28	23839,36	13,69
14	152	3,3	501,6	23104	10,89
15	147	3,4	499,8	21609	11,56
16	154,4	4,0	617,6	23839,36	16
17	159,3	4,1	653,13	25376,49	16,81
18	164	3,8	623,2	26896	14,44
19	147	3,4	499,8	21609	11,56
20	149,5	3,6	538,3	22350,25	12,96
<b>Сумма =</b>	<b>3197,1</b>	<b>75,1</b>	<b>12071,81</b>	<b>513417,83</b>	<b>285,45</b>

В трёх первых столбцах повторена таблица, приведенная выше. В трёх следующих столбцах представлены простые расчеты, выполненные на основании данных о росте и о самоуважении. Это вся информация, которая необходима для того, чтобы рассчитать корреляцию. В нижней строке таблицы имеются следующие значения:

$$N = 20$$

$$\sum xy = 12071,81$$

$$\sum x = 3197,1$$

$$\sum y = 75,1$$

$$\sum x^2 = 513417,83$$

$$\sum y^2 = 285,45$$

Теперь подставляем эти значения в формулу для расчета корреляции, которая приведена выше, и получаем следующее (последовательно, шаг за шагом):

$$r = \frac{20(12071,81) - (3197,1)(75,1)}{\sqrt{[20(513417,83) - (3197,1)^2][20(285,45) - (75,1)^2]}}$$

$$r = \frac{241436,2 - 240102,21}{\sqrt{[10268356,6 - 10221448,41][5709 - 5640,01]}}$$

$$r = \frac{1333,99}{\sqrt{[46908,19][68,99]}}$$

$$r = \frac{1333,99}{\sqrt{3236196,0281}}$$

$$r = \frac{1333,99}{1798,943030810036518872929744221}$$

$$r = \frac{1333,99}{1798,943030810036518872929744221}$$

$$r = 0,74$$

Итак, всего для 20 испытуемых, корреляция оказалась +0,74 – это достаточно сильная позитивная корреляция. Можно полагать, что существует взаимосвязь между ростом и самоуважением, по крайней мере, в этих смоделированных данных!

### Проверка значимости корреляции

Рассчитав корреляцию, следует рассмотреть возможность того, что наблюдаемая корреляция может оказаться случайной. Нас интересуют реальные корреляции, а не случайные. В данном случае, проверке подвергается взаимно исключающие гипотезы:

Нулевая гипотеза:  $r = 0$

Альтернативная гипотеза:  $r < > 0$

Самый простой способ проверки гипотезы – взять статистический справочник, в котором имеется таблица критических значений коэффициентов корреляции  $r$ . Так же, как и при проверке любой гипотезы, сначала следует определить **уровень значимости**. Примем распространенный уровень

значимости  $\alpha = 0,05$ . Это означает, что мы проводим проверку того, что вероятность случайности полученной корреляции не превышает 5 из 100. Прежде, чем начать поиск критического значения по таблице, следует рассчитать количество степеней свободы или  $df$ . Количество степеней свободы равно  $N - 2$ , в нашем примере  $20 - 2 = 18$ . Наконец, следует выбрать, какая проверка будет производиться – односторонняя или двусторонняя. В нашем примере, поскольку у нас не было прочной априорной теории относительно того, какова взаимосвязь роста и самоуважения – позитивная либо негативная, выберем двустороннюю проверку. Собрав воедино все три компонента информации:

- ❖ уровень значимости ( $\alpha = 0,05$ );
- ❖ количество степеней свободы ( $df = 18$ );
- ❖ тип проверки (двусторонняя)

теперь можно приступать к проверке значимости найденной корреляции. По таблице мы находим, что критическое значение

$$r = 0,4438.$$

Это означает, что, если наша корреляция больше  $+0,4438$  либо меньше  $-0,4438$  (напомним, что проверка – двусторонняя), то можно сделать вывод о том, что вероятность того, что корреляция оказалась случайной, равна 5 из 100. Поскольку наша корреляция  $+0,74$  намного превышает критическое значение, взятое из таблицы, мы делаем вывод о том, что наша корреляция не случайна, что корреляция «статистически значима» (с поправкой на параметры проверки). Мы можем отвергнуть нулевую гипотезу и принять альтернативную гипотезу.

### Корреляционная матрица

Итак, мы увидели то, как рассчитывается корреляция между двумя переменными. В большинстве исследований бывает более двух переменных. Допустим, что у нас 10 переменных на уровне интервалов, и мы хотим рассчитать взаимосвязи для всех десяти (т. е. между всеми возможными парами переменных). В нашем примере придется вычислять 45 корреляций (чуть позже о том, как это определить!) Мы можем 45 раз выполнять расчеты, либо, при помощи статистической программы, мы можем получить 45 корреляций одним кликом мышки. Допустим, то мы пошли по второму пути. Вот результат:

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	1,000									
C2	0,274	1,000								
C3	-0,134	-0,269	1,000							
C4	0,201	-0,153	0,075	1,000						
C5	-0,129	-0,166	0,278	-0,011	1,000					
C6	-0,095	0,280	-0,348	-0,378	-0,009	1,000				
C7	0,171	-0,122	0,288	0,086	0,193	0,002	1,000			
C8	0,219	0,242	-0,380	-0,227	-0,551	0,324	-0,082	1,000		
C9	0,518	0,238	0,002	0,082	-0,015	0,304	0,347	-0,013	1,000	
C10	0,299	0,568	0,165	-0,122	-0,106	-0,169	0,243	0,014	0,352	1,000

Таблица такого типа называется **корреляционной матрицей**. Наименования переменных (С1 – С10) приводятся в левом столбце и в верхней строке. Диагональ корреляционной матрицы (т. е. числа, исходящие из верхнего левого угла к нижнему правому углу) всегда состоит из единиц. Это так потому, что данные корреляции – между каждой переменной и ею самой (а переменная всегда превосходно коррелирует сама с собой). Статистические программы воспроизводят только нижний треугольник корреляционной матрицы. В любой корреляционной матрице существуют два треугольника – значения выше и правее диагонали (верхний треугольник) и ниже и левее диагонали (нижний треугольник). Совершенно ни к чему воспроизводить оба треугольника, потому что они зеркально отражают друг друга (корреляция переменной  $x$  с переменной  $y$  – это всегда корреляция переменной  $y$  с переменной  $x$ ). Матрица, которая обладает таким свойством зеркального отображения, называется **симметричной матрицей**.

Для того, чтобы определить корреляцию для любой пары переменных, найдите значение в таблице на пересечении строки и столбца. Например, чтобы найти корреляцию между переменными С5 и С2, нужно найти строку С2 и столбец С5 (в нашем случае – пустое место, поскольку это верхний треугольник), либо строку С5 и столбец С2, и увидеть, что корреляция равна – 0,166.

Итак, как мы узнали, что для 10 переменных нам потребуется 45 корреляций? Существует весьма простая формула, позволяющая определить количество пар (т. е. корреляций) для любого количества переменных:

$$\frac{N * (N - 1)}{2}$$

где  $N$  – это количество переменных, В нашем примере, у нас есть 10 переменных, поэтому

$$(10 * 9)/2 = 90/2 = 45 \text{ пар.}$$

### **Другие корреляции**

Рассмотренные корреляции называются корреляциями Пирсона. Они применяются тогда, когда обе переменных измерены на уровне интервалов. Однако, существует множество других типов корреляций. Например, если переменные порядковые, тогда применяются либо ранговые корреляции Спирмена ( $\rho$  «ро»), либо ранговые корреляции Кендалла ( $\tau$  «тау»). Если одна переменная измерена на непрерывном интервальном уровне, а другая – дихотомическая (т. е. двойная категория), тогда может применяться точечно-бисериальная корреляция.



## Вопросы для самоконтроля по теме «Корреляция»

1. Что такое корреляция?
2. По какой формуле рассчитывается корреляция?
3. Как проверить корреляцию?
4. Как найти степень свободы корреляции?
5. Какие компоненты информации необходимы для проверки корреляции?
6. Что означает критическое значение корреляции?
7. Что такое корреляционная матрица?
8. Какая матрица называется симметричной?
9. Как с помощью матрицы определить корреляцию для любой пары переменных?

## Тестовые задания к теме 8

1. Число, которым описывается степень взаимосвязи между переменными, называется
  - а) размах
  - б) дисперсия
  - в) медиана
  - г) мода
  - д) корреляция
2. Выражение «позитивная взаимосвязь» означает, что более высоким значениям по одной переменной соответствуют по второй переменной
  - а) более низкие значения
  - б) более высокие значения
  - в) равные значения
  - г) нулевые значения
  - д) число  $\pi$
3. Выражение «позитивная взаимосвязь» означает, что более низким значениям по одной переменной соответствуют по второй переменной
  - а) более низкие значения
  - б) более высокие значения
  - в) равные значения
  - г) нулевые значения
  - д) число  $\pi$
4. Рассчитав корреляцию, следует рассмотреть возможность того, что наблюдаемая корреляция может оказаться
  - а) положительной
  - б) отрицательной
  - в) нулевой
  - г) достоверной
  - д) случайной

5. Уровень значимости  $\alpha = 0,05$  означает, что мы проводим проверку того, что вероятность случайности полученной корреляции не превышает 5 из 100
- а) 99 из 100
  - б) 95 из 100
  - в) 50 из 100
  - г) 5 из 100
  - д) 1 из 100
6. Диагональ корреляционной матрицы (т. е. числа, исходящие из верхнего левого угла к нижнему правому углу) всегда состоит из
- а) крестиков
  - б) ноликов
  - в) единичек
  - г) корреляций
  - д) процентов
7. В любой корреляционной матрице существуют
- а) два крестика
  - б) два нолика
  - в) две единички
  - г) два треугольника
  - д) два параллелепипеда

## ТЕМА 9. t-КРИТЕРИЙ

### СТАТИСТИЧЕСКИЕ ВЫВОДЫ

Когда мы пытаемся сформулировать статистические выводы, мы выходим за пределы только непосредственных данных. К примеру, мы пытаемся на основании выборочных данных сделать статистический вывод о том, что считают люди. Либо, мы пытаемся сделать статистический вывод относительно вероятности того, закономерно или случайно наблюдаемое различие между группами. Тем самым, мы прибегаем к статистическим выводам тогда, когда на основании имеющихся данных необходимо сделать обобщенные заключения. В этом заключается отличие статистических выводов от дескриптивной статистики, которая просто описывает то, что содержится в собранных данных.

Рассмотрим те статистические выводы, которые полезны при экспериментальном и при квазиэкспериментальном планах исследования, либо при оценивании результатов осуществления какой-либо программы. Один из простейших критериев, относящихся к формулированию статистических выводов, используется при сравнении средней результативности двух групп по единой методике с целью обнаружить различия. Допустим, мы сравниваем баллы по математическому тесту среди учеников восьмого класса – мальчиков и девочек, либо выясняем, отличается ли успеваемость в группе, которая обучалась по новой программе, с успеваемостью в контрольной группе уча-

щихся. Каждый раз, при сравнении средней результативности в двух группах, полезно применять t-критерий межгрупповых различий.

Формулирование статистических выводов предполагает обращение к семье статистических моделей, известной под именем **общей линейной модели**. В неё входят

- ❖ t-критерий
- ❖ дисперсионный анализ (ANOVA)
- ❖ ковариационный анализ (ANCOVA)
- ❖ регрессионный анализ
- ❖ различные многомерные методы:
  - факторный анализ
  - многомерное шкалирование
  - кластерный анализ
  - дискриминантный функциональный анализ и т. п.

Поскольку общая линейная модель играет большую роль в статистическом анализе, каждому серьезному исследователю необходимо с ней познакомиться. Для начала обратимся к простейшей прямолинейной модели. Это позволит создать у вас представление о линейной модели и подготовит вас к более сложным разновидностям анализа, изложенным ниже.

Основой для понимания того, каким образом производится сравнение групп, выступает индикаторная (фиктивная) переменная. Слово «фиктивная» в названии этой переменной может ввести в заблуждение и заставить подумать, что мы анализируем то, чего не существует. Лучше было бы назвать её приблизительной переменной. На самом деле, индикаторной (фиктивной) переменной называется такая переменная, которая применяет дискретные значения, обычно 0 и 1, обозначающие различные исследуемые группы. Индикаторные (фиктивные) переменные позволяют легко понять сложные вещи. Например, если включить индикаторную (фиктивную) переменную в какую-либо модель, то можно смоделировать два отдельных выражения (по одному выражению для каждой группы) и эти выражения уравнять.

Одна из важнейших разновидностей анализа, применяемая при оценке результативности программ, предполагает сравнение экспериментальной группы и контрольной группы по одной или нескольким переменным, отражающим эффективность сформированных знаний. То, каким образом такой анализ будет производиться, зависит от выбранного нами плана исследования. Планы исследований подразделяются на два основных типа:

- ❖ экспериментальный план
- ❖ квазиэкспериментальный план

Каждый план предполагает особый анализ.

**Экспериментальный анализ.** Простой *посттестовый рандомизированный эксперимент для двух групп* обычно анализируют с применением t-критерия либо дисперсионного анализа (ANOVA). *Факторный экспериментальный план* предполагает дисперсионный анализ (ANOVA) с одним вхо-

дом. *Экспериментальный план ковариационного анализа* предполагает – ничего удивительного – статистическую модель ковариационного анализа.

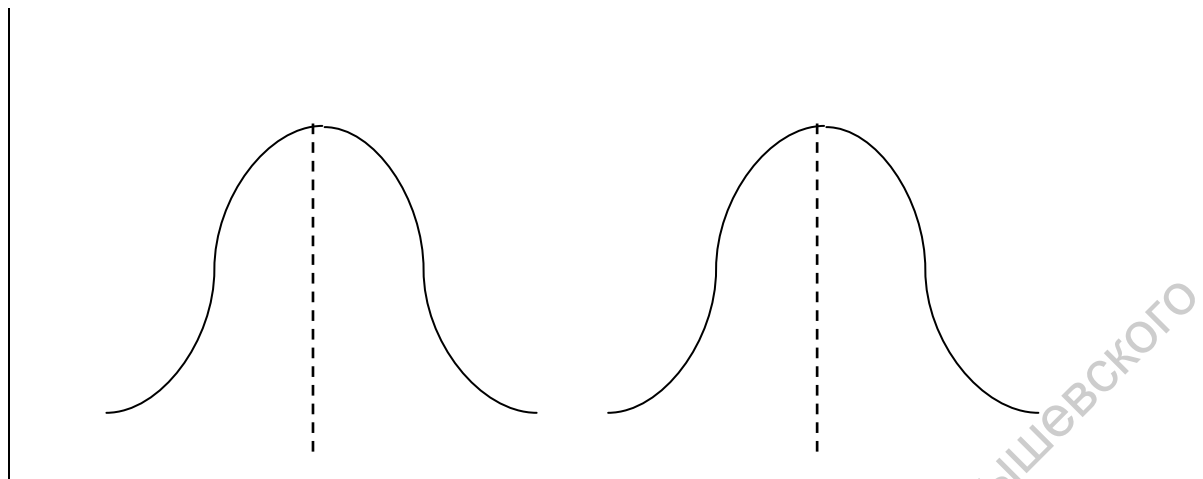
**Квазиэкспериментальный анализ.** Квазиэкспериментальные планы отличаются от экспериментальных планов тем, что не предполагают рандомизации. *Рандомизацией называется перемешивание, в результате которого множество объектов располагаются в случайном порядке.* Отсутствие рандомизации затрудняет анализ. Например, при анализе по *плану неэквивалентных групп* необходима ковариационная модель анализа коррекции надежности, предполагающая, что претестовые баллы скорректированы с поправкой на ошибку в измерениях. *План перегиба регрессии* обращает особое внимание на криволинейность и отказ от детализации. Соответственно, мы будем придерживаться консервативного аналитического подхода, основанного на полиномной регрессии, который начинается с анализа истинной функции и ведет к редуцированию модели на основе данных проведенного анализа. *План точечного замещения регрессии* имеет лишь один компонент анализа. При этом план точечного замещения регрессии непосредственно основан на традиционной модели ковариационного анализа (ANCOVA).

Перечисленные различные модели анализа относятся к одной семье – к общей линейной модели. Овладение этой моделью сопряжено со многими нюансами анализа данных в прикладном контексте исследований.

## **t-КРИТЕРИЙ**

При помощи t-критерия оценивается – отличаются ли средние значения по двум группам друг от друга *статистически* достоверным образом. Такого рода анализ уместен в тех случаях, когда необходимо сравнить между собой средние значения по двум группам, а в особенности – при проведении анализа по *посттестовому двухгрупповому рандомизированному экспериментальному плану.*

На рис. 1 представлены распределения баллов в экспериментальной и контрольной группах. На рисунке видны идеализированные распределения – действительные распределения обычно изображаются при помощи гистограммы. Пунктирная линия показывает расположение средних значений для экспериментальной группы и для контрольной группы. t-критерий призван ответить на вопрос – существуют ли статистические различия между средними значениями?



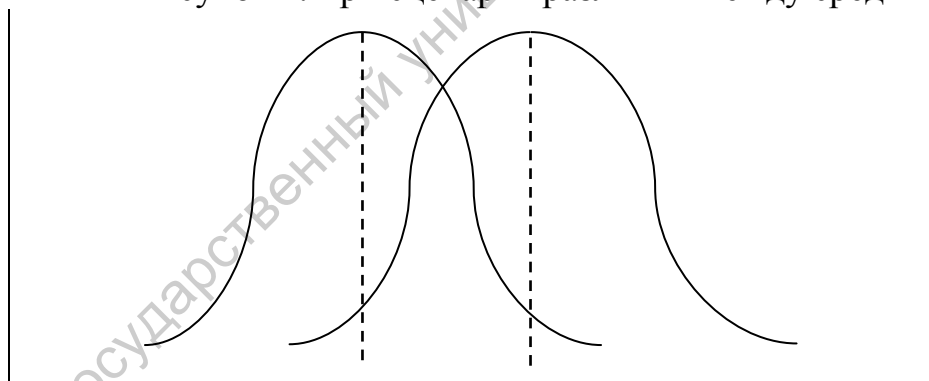
↑  
среднее для контрольной группы

↑  
среднее для экспериментальной группы

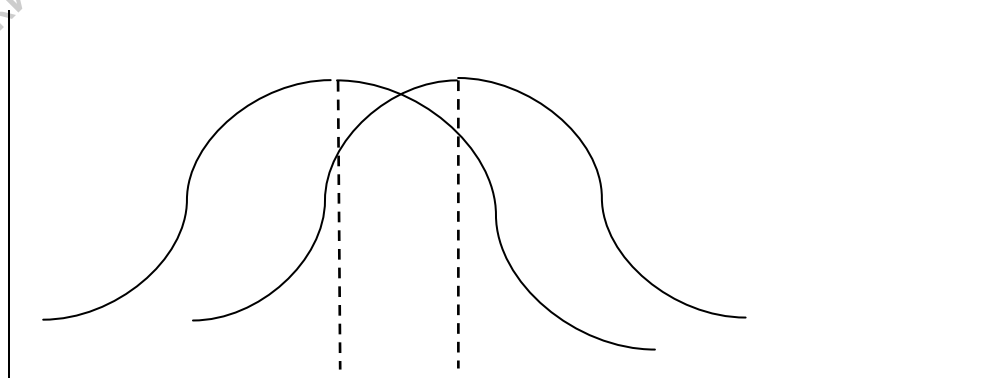
Рис. 1 Идеализированное распределение посттестовых значений в экспериментальной и контрольной группах

Что означает выражение «статистическое различие между группами по средним значениям»? Рассмотрим три ситуации, показанные на рисунке 2.

Рисунок 2. Три сценария различий между средними

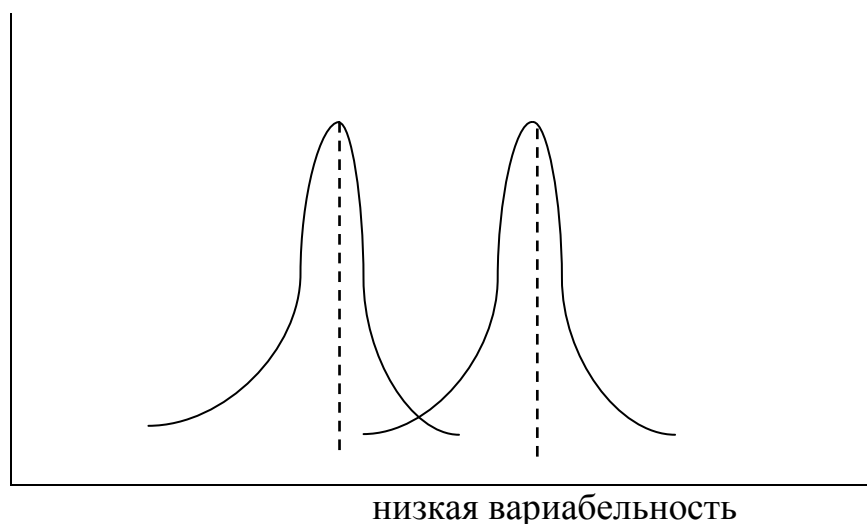


средняя вариабельность



высокая вариабельность

Саратовский государственный университет имени Н. Г. Чернышевского



Первое, на что следует обратить внимание во всех трех показанных случаях – это то, что **во всех трех случаях различие между средними одно и то же**. Однако, следует также обратить внимание на то, что эти три ситуации выглядят различно – в них отражены три разных случая.

На верхнем рисунке изображена умеренная вариабельность баллов в каждой из групп. На втором рисунке изображена высокая вариабельность. На третьем рисунке изображена низкая вариабельность. Понятно, что мы склонны сделать вывод о том, что наибольшие сходства либо наибольшие различия представлены в третьем из рассматриваемых случаев. Почему? Потому что две колоколообразные кривые почти не накладываются друг на друга. При высокой вариабельности групповые различия представляются наименее заметными, потому что обе колоколообразные кривые сильно накладываются друг на друга.

Из сказанного можно сделать важный вывод: когда мы рассматриваем различия в баллах между двумя группами, необходимо учитывать различия между средними значениями относительно распределения или вариабельности баллов. Эту задачу решает t-критерий.

### Статистический анализ с применением t-критерия

Формула для расчета t-критерия представляет собой дробь. В числителе дроби – различие между двумя средними. В знаменателе – значение вариабельности или дисперсия баллов. Данная формула представляет собой еще один метафорический пример отношения «сигнала к шуму»: различие между средними – это сигнал того, что мы включаем в наши данные; в знаменателе формулы – мера варьирования или шум, затрудняющий поиск межгрупповых различий.

Рисунок 3 иллюстрирует формулу для расчета t-критерия через соотношение числителя и знаменателя в распределении данных:

Числитель формулы нетрудно рассчитать – всего лишь определить разность между средними значениями. Знаменатель называется **стандартной ошибкой разности**. Для того, чтобы рассчитать стандартную ошибку, нужно

взять дисперсию каждой группы и разделить её на количество людей в этой группе минус 1. Затем два значения складывают, а из суммы извлекают квадратный корень. Точная формула имеет следующий вид:

### Формула стандартной ошибки различия между средними

$$\text{стандартная ошибка } (X_T - X_C) = \sqrt{\frac{\text{var}_1}{(n_1 - 1)} + \frac{\text{var}_2}{(n_2 - 1)}}$$

где **var** – сокращение от английского variance, что значит **дисперсия**. Необходимо помнить о том, что дисперсия – это всего лишь **квадрат стандартного отклонения**.

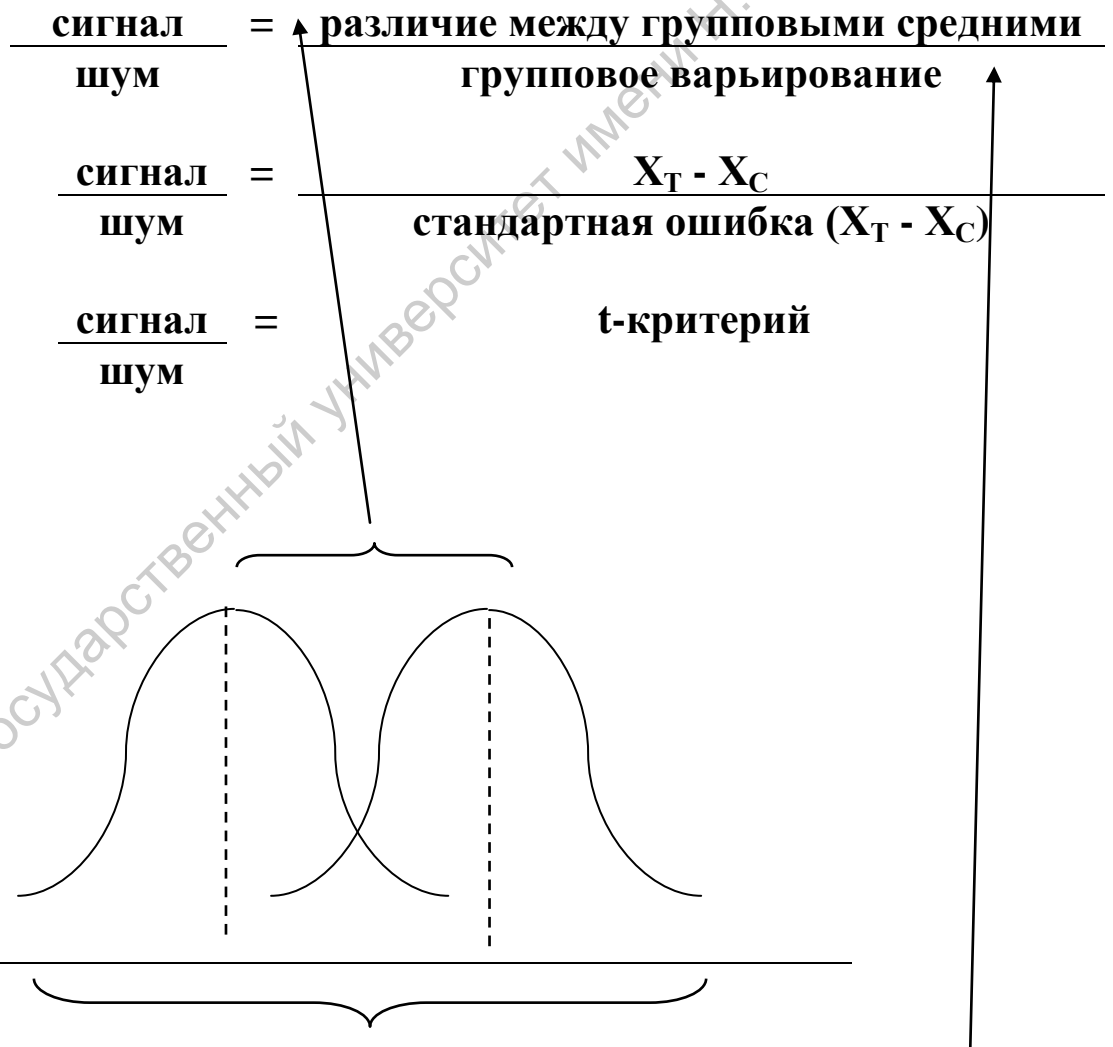


Рисунок 3. Формула t-критерия

Окончательный вид формулы для t-критерия:

$$t = \frac{X_T - X_C}{\sqrt{\frac{\text{var}_1}{(n_1 - 1)} + \frac{\text{var}_1}{(n_1 - 1)}}$$

t-критерий будет положительным, если первое среднее будет больше, чем второе среднее, и t-критерий будет отрицательным, если первое среднее меньше, чем второе среднее. После того, как вычислено значение t-критерия, нужно проверить его по таблице критических значений – является ли частное дроби достаточно большим для того, чтобы утверждать, что различие между группами не случайно. Для проверки значимости t-критерия, нужно задать уровень риска (который называется альфа-уровень). В большинстве психологических исследований эмпирически принято задавать уровень альфа равным 0,05. Это означает, что пять раз из ста мы обнаруживаем статистически значимое различие между средними значениями даже если статистически значимого различия не существует (т. е. «случайно»). Необходимо также определить степени свободы (df) t-критерия. Название «степени свободы» означает число оцениваемых по данной выборке параметров. В общем случае, количество степеней свободы равно объему выборки минус число параметров. Количество степеней свободы для t-критерия равно сумме человек в обеих группах минус 2.

Зная альфа-уровень, количество степеней свободы и значение t-критерия, можно найти t-значение в стандартных таблицах значимости (которые обычно находятся в приложении к большинству книг по статистике) и определить – достаточно ли велико t-значение для того, чтобы считать его значимым. Если ответ «да», тогда делается вывод о том, что различие между средними двух групп существует (даже с учетом варьирования). К счастью, в статистических программах для компьютеров заложено автоматическое определение уровня значимости, освобождающее от необходимости обращаться к таблицам.

t-критерий, дисперсионный анализ (ANOVA) с одним входом и регрессионный анализ являются математически эквивалентными и обеспечивают одинаковые результаты.

### **Вопросы для самоконтроля по теме «t-критерий»**

1. Какой критерий полезно применять для сравнения средней результативности в двух группах?
2. В семью каких статистических моделей входит этот критерий?



3. Что выступает основой для понимания того, каким образом производится сравнение групп?
4. На какие два основных типа подразделяются планы исследований?
5. В чем состоит отличие квазиэкспериментальных планов от экспериментальных планов?
6. Что такое рандомизация?
7. В каких случаях уместно применять t-критерий?
8. Какие распределения обычно изображаются при помощи гистограммы?
9. Что означает выражение «статистическое различие между группами по средним значениям»?
10. Какой вид имеет формула для расчета t-критерия?
11. Как рассчитать стандартную ошибку?
12. Чем является дисперсия?
13. При каком условии t-критерий будет положительным? Отрицательным?
14. При помощи чего проверяют вычисленное значение t-критерия?
15. Что нужно задать для проверки значимости t-критерия?
16. Какое значение альфа-уровня принято задавать в большинстве психологических исследований? Что оно означает?
17. Что означает название «степени свободы»?
18. Чему равно количество степеней свободы в общем случае?
19. Что следует знать для того, чтобы найти t-значение в стандартных таблицах значимости?
20. Чему математически эквивалентен t-критерий?

## Тестовые задания к теме 9

1. Когда на основании имеющихся данных необходимо сделать обобщенные заключения, мы прибегаем
  - а) к статистическим оценкам
  - б) к статистическим выводам
  - в) к статистическим тенденциям
  - г) к статистическим отклонениям
  - д) к статистическим погрешностям
2. При сравнении средней результативности в двух группах, полезно применять
  - а) корреляционную матрицу
  - б) медиану
  - в) моду
  - г) t-критерий
  - д) квотную выборку

3. В семью статистических моделей, известной под именем общей линейной модели, входит

- а) корреляционная матрица
- б) медиана
- в) мода
- г) t-критерий
- д) квотная выборка

4. Индикаторной (фиктивной) переменной называется такая переменная, которая применяет дискретные значения

- а) 1 и 2
- б) 0 и 2
- в) 0 и 1
- г) 0 и 10
- д) 2 и 1

5. Перемешивание, в результате которого множество объектов располагаются в случайном порядке, называется

- а) операционализация
- б) структуризация
- в) систематизация
- г) реструктуризация
- д) рандомизация

6. Ковариационная модель анализа коррекции надежности необходима при анализе

- а) по факторному экспериментальному плану
- б) по плану посттестового рандомизированного эксперимента
- в) по экспериментальному плану ковариационного анализа
- г) по плану точечного замещения регрессии
- д) по плану неэквивалентных групп

7. На традиционной модели ковариационного анализа (ANCOVA) непосредственно основан анализ

- а) по факторному экспериментальному плану
- б) по плану посттестового рандомизированного эксперимента
- в) по экспериментальному плану ковариационного анализа
- г) по плану точечного замещения регрессии
- д) по плану неэквивалентных групп

8. Когда необходимо сравнить между собой средние значения по двум группам, а в особенности – при проведении анализа по посттестовому двухгрупповому рандомизированному экспериментальному плану, уместно применение

- а) корреляционной матрицы
- б) медианы
- в) моды
- г) t-критерия
- д) квотной выборки

9. Выражение «статистическое различие между группами по средним значениям» означает, что две колоколообразные кривые
- а) накладываются друг на друга
  - б) не накладываются друг на друга
  - в) параллельны друг другу
  - г) пересекают ось  $x$
  - д) приближаются к оси  $y$
10. Выражение «статистическое различие между группами по средним значениям» означает
- а) высокую вариабельность
  - б) низкую вариабельность
  - в) приемлемую вариабельность
  - г) среднюю вариабельность
  - д) нулевую вариабельность
11. Когда необходимо учитывать различия между средними значениями относительно распределения или вариабельности баллов, уместно применение
- а) корреляционной матрицы
  - б) медианы
  - в) моды
  - г)  $t$ -критерия
  - д) квотной выборки
12. Формула для расчета  $t$ -критерия представляет собой
- а) интеграл
  - б) радикал
  - в) квадратный корень
  - г) дробь
  - д) матрицу
13. Формула для расчета  $t$ -критерия представляет собой
- а) отношение медианы к моде
  - б) произведение валидности и надежности
  - в) разность сомножителей
  - г) отношение сигнала к шуму
  - д) сумму отклонений от среднего
14. Если первое среднее будет больше, чем второе среднее, тогда  $t$ -критерий будет
- а) нулевым
  - б) рациональным
  - в) отрицательным
  - г) положительным
  - д) случайным
15. Если первое среднее меньше, чем второе среднее, тогда  $t$ -критерий будет
- а) нулевым
  - б) рациональным
  - в) отрицательным
  - г) положительным

- д) случайным
16. Для проверки значимости t-критерия, нужно задать
- а) доверительный интервал
  - б) уровень значимости
  - в) уровень случайности
  - г) интервал вероятности
  - д) уровень риска
17. В большинстве психологических исследований эмпирически принято задавать уровень альфа равным
- а) 3,14
  - б) 1,0
  - в) 0,9
  - г) 0,05
  - д) 0
18. Уровень альфа, равный 0,05, означает, что, даже если статистически значимого различия не существует, мы обнаруживаем статистически значимое различие между средними значениями
- а) сто раз из ста
  - б) 99 раз из ста
  - в) 95 раз из ста
  - г) 65 раз из ста
  - д) пять раз из ста
19. Название «степени свободы» означает число оцениваемых по данной выборке
- а) характеристик
  - б) ответов
  - в) статистик
  - г) параметров
  - д) габаритов
20. Количество степеней свободы для t-критерия равно сумме человек в обеих группах
- а) плюс один
  - б) минус один
  - в) плюс два
  - г) минус два
  - д) плюс-минус один-два
21. Зная альфа-уровень, количество степеней свободы и значение t-критерия, можно найти t-значение в стандартных таблицах значимости и определить – достаточно ли велико t-значение для того, чтобы считать его значимым. Если ответ «да», тогда делается вывод о том, что различие между средними двух групп
- а) велико
  - б) мало
  - в) положительно
  - г) отрицательно

д) существует

## **Семинар 2. Меры центральной тенденции и показатели вариации.**

Вопросы для обсуждения:

1. Мода. Соглашение об использовании моды.
2. Медиана. Вычисление медианы.
3. Квантили.
4. Вычисление и интерпретация моды, медианы и среднего.
5. Средняя арифметическая, средняя гармоническая, средняя квадратическая.
6. Средняя кубическая, средняя геометрическая.

Темы для рефератов:

1. Выбор меры центральной тенденции.
2. Размах вариации.
3. Дисперсия и ее свойства.
4. Среднее квадратическое отклонение. Поправка Шеппарда.
5. Коэффициент вариации.

Литература к семинару 2:

1. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., 1976. С. 58 – 72
2. Лакин Г. Ф. Биометрия. М., 1990. С. 38 – 64
3. Сочивко Д. В., Якунин В. А. Математические модели в психолого-педагогических исследованиях. Л., 1988. С. 5 – 40
4. Битинас Б. Многомерный анализ в педагогике и педагогической психологии. Вильнюс, 1971. С. 44 – 67; 129 – 148

## **МОДУЛЬ 3**

### **ТЕМА 10. ИНДИКАТОРНЫЕ (ФИКТИВНЫЕ) ПЕРЕМЕННЫЕ**

Индикаторной (фиктивной) переменной называется количественная переменная, используемая в регрессионном анализе для представления выборочных подгрупп. В научном плане исследования, индикаторная (фиктивная) переменная часто применяется для разграничения экспериментальных групп. В простейшем случае, применяется индикаторная (фиктивная) переменная со значениями 0 и 1, где испытуемому из контрольной группы присваивается значение 0, а испытуемому из экспериментальной группы присваивается значение 1. Полезность индикаторных (фиктивных) переменных объясняется тем, что при их помощи мы можем представить несколько групп испытуемых в виде одного уравнения. Это означает, что отсутствует необходимость расписывать отдельные математические выражения для каждой из подгрупп. Индикаторные (фиктивные) переменные можно сравнить с переключателями,

которые включают и выключают из уравнения различные исследуемые параметры.

Другое преимущество индикаторных (фиктивных) переменных, кодируемых нулем и единицей, заключается в том, что, несмотря на то, что они являются переменными номинального уровня (шкала наименований), но статистически с ними можно обращаться как с переменными интервального уровня (шкала интервалов). К примеру, возьмем среднее по переменной, кодируемой нулем и единицей, и получим следующее распределение единиц:

$$y_i = \beta_0 + \beta_1 Z_i + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена
- $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси
- $\beta_1$  = коэффициент наклона
- $Z_i$  = 1, если  $i$ -тый член относится к экспериментальной группе  
0, если  $i$ -тый член относится к контрольной группе
- $e_i$  = остаток для  $i$ -того члена после вычитания

Для иллюстрации индикаторных (фиктивных) переменных, рассмотрим простую регрессионную модель двухгруппового посттестового рандомизированного эксперимента. Это, по существу, такая же модель, что и при проведении проверки с помощью  $t$ -критерия по посттестовым средним для двух групп либо при проведении дисперсионного анализа (ANOVA) с одним входом. Ключевым элементом рассматриваемой модели выступает  $\beta_1$  или величина различия между группами.

Для того, чтобы увидеть то, как работают индикаторные (фиктивные) переменные, применим простую модель, которая продемонстрирует способ составления отдельных подуравнений для каждой подгруппы. Далее рассмотрим то, как вычисляются различия между подгруппами путем вычитания соответствующих уравнений. Мы увидим, что большой объем информации можно представить в виде единого уравнения с применением индикаторных (фиктивных) переменных. Главное заключается в том, чтобы понять, что  $\beta_1$  выражает различие между экспериментальной группой и контрольной группой.

Для того, чтобы это понять, первый шаг состоит в вычислении уравнения для каждой из двух групп по отдельности. Для контрольной группы  $Z = 0$ . Если подставить это значение в уравнение и допустить, что величина ошибки равна нулю, тогда прогнозируемое значение для контрольной группы равно  $\beta_0$ .

Теперь, для того, чтобы рассчитать уравнение для контрольной группы, вместо  $Z$  нужно подставить 1. Допустим, что ошибка снова равна нулю. Тогда уравнение для контрольной группы будет равно сумме двух значений  $\beta$ .

В алгебраическом выражении получаем следующее:

$$y_i = \beta_0 + \beta_1 Z_i + e_i$$

Сначала определяем эффективность каждой группы:

Для контрольной группы ( $Z_i = 0$ ):

$$y_C = \beta_0 + \beta_1(0) + 0 \quad \leftarrow \quad e_i \text{ в группе усредняется до } 0$$

$$y_C = \beta_0$$

Для экспериментальной группы ( $Z_i = 1$ ):

$$y_T = \beta_0 + \beta_1(1) + 0$$

$$y_T = \beta_0 + \beta_1$$

Теперь можно переходить ко второму действию – к расчету различий между группами. Как мы это определили? Различие между группами должно быть различием между уравнениями, которое мы вывели выше. Иными словами, для того, чтобы найти различие между группами, мы должны найти различие между уравнениями для двух групп! Очевидно, что различие выражается значением  $\beta_1$ . Повторим еще раз – различие между группами в данной модели равно  $\beta_1$ !

Теперь найдем различие между группами:

экспериментальная

группа

$$y_T = \beta_0 + \beta_1$$

контрольная

группа

$$y_C = \beta_0$$

$$y_T - y_C = (\beta_0 + \beta_1) - \beta_0$$

$$y_T - y_C = \cancel{\beta_0} + \beta_1 - \cancel{\beta_0}$$

$$y_T - y_C = \beta_1$$

В каждом случае регрессионной модели с индикаторными (фиктивными) переменными всегда можно увидеть то, каким образом используемые переменные отображают уравнения для множества подгрупп, если выполнять последовательность из двух указанных действий:

- ❖ создать отдельные уравнения для каждой подгруппы с применением индикаторных (фиктивных) переменных
- ❖ найти различие между группами через определение различия между их уравнениями

### Вопросы для самоконтроля

#### по теме «Индикаторные (фиктивные) переменные»

1. Что называют индикаторной (фиктивной) переменной?
2. Где применяют индикаторные (фиктивные) переменные?
3. Чем объясняется полезность индикаторных (фиктивных) переменных? Что это означает?
4. С чем можно сравнить индикаторные (фиктивные) переменные?
5. К какому уровню относятся индикаторные (фиктивные) переменные? Как можно обращаться с ними в статистическом смысле? Приведите пример распределения.
6. Какой элемент модели является ключевым?
7. В чем состоит первый шаг представления большого объема информации с помощью индикаторных (фиктивных) переменных?
8. Чему равно уравнение для контрольной группы? Как выразить его алгебраически?
9. Как определяется эффективность каждой группы?
10. Как рассчитывается различие между группами? Чему оно равно?
11. Какую последовательность действий следует выполнять для того, чтобы с помощью индикаторных (фиктивных) переменных отобразить уравнения для множества подгрупп?

### Тестовые задания к теме 10

1. Количественная переменная, используемая в регрессионном анализе для представления выборочных подгрупп, называется
  - а) независимая переменная
  - б) зависимая переменная
  - в) индикаторная (фиктивная) переменная
  - г) шум
  - д) константа
2. В простейшем случае применяется индикаторная (фиктивная) переменная со значениями
  - а) 1 и 2
  - б) 0 и 2
  - в) 0 и 1
  - г) 0 и 10
  - д) 2 и 1



3. Полезность индикаторных (фиктивных) переменных объясняется тем, что при их помощи мы можем представить несколько групп испытуемых в виде
- а) гистограммы
  - б) выборки
  - в) колоколообразной кривой
  - г) патерна
  - д) уравнения
4. Несмотря на то, что индикаторные (фиктивные) переменные являются переменными номинального уровня (шкала наименований), но статистически с ними можно обращаться как с переменными
- а) нулевого уровня
  - б) порядкового уровня
  - в) интервального уровня
  - г) уровня отношений
  - д) в принципе любого уровня
5. Ключевым элементом простой регрессионной модели двухгруппового посттестового рандомизированного эксперимента  $y_i = \beta_0 + \beta_1 Z_i + e_i$  оказывается
- а) полученное значение для  $i$ -того члена
  - б) коэффициент для отрезка, отсекаемого на координатной оси
  - в)  $i$ -тый член экспериментальной группы
  - г)  $i$ -тый член контрольной группы
  - д) величина различия между экспериментальной и контрольной группами
6. Различие между экспериментальной и контрольной группами должно быть различием между
- а) гистограммами
  - б) выборками
  - в) колоколообразными кривыми
  - г) патернами
  - д) уравнениями

## ТЕМА 11. ОБЩАЯ ЛИНЕЙНАЯ МОДЕЛЬ

Общая линейная модель лежит в основе большей части разновидностей статистического анализа, применяемых в психологических исследованиях. Она лежит в основе  $t$ -критерия, дисперсионного анализа (ANOVA), ковариационного анализа (ANCOVA), регрессионного анализа и многих многомерных методов, включая факторный анализ, кластерный анализ, многомерное шкалирование, анализ дискриминант функций, канонические корреляции и т. п. По причине своего всеобъемлющего характера, данная модель играет важную роль для студентов, изучающих психологию. Глубокое понимание общей линейной модели требует определенной подготовки в области математической статистики. Поэтому мы рассмотрим само понятие общей линейной модели и её характеристики.

## Двумерная линейная модель

Случай с двумя переменными наиболее доступен для понимания общей линейной модели. На рис. 1 представлена двумерная диаграмма для двух переменных. Это могут быть любые две непрерывные переменные, допустим, одна переменная – претестовая (на оси  $x$ ), а другая переменная – посттестовая (на оси  $y$ ). Каждая точка на графике соответствует балам, полученным испытуемым до проведения эксперимента и после проведения эксперимента. В данной модели ясно видна позитивная взаимосвязь, т. е. испытуемые, балл которых до эксперимента был выше, получили более высокий балл после эксперимента, и наоборот.

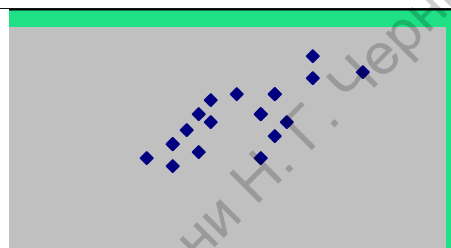


Рисунок 1. Двумерный график

Цель нашего анализа данных заключается в обобщении или в точном описании того, что происходит с данными. Данные представлены на двумерном графике. Каков лучший способ обобщения этих данных? На рисунке 2 показана прямая линия, проходящая через «облако» точечных данных. Несмотря на то, что эта линия не вполне точно характеризует конкретные значения (потому что ни одна точка не попадает на эту линию), она точным образом отражает образ наших данных. Когда мы проводим среди данных линию – мы создаем **линейную модель**. Термин «линейная» относится к тому факту, что мы провели линию. Слово «модель» относится к уравнению, которым описывается проведенная нами линия. Линия типа той, которую мы видим на рисунке 2, часто называется **линией регрессии**, а анализ, который связан с этой линией, часто называют **регрессионным анализом**.

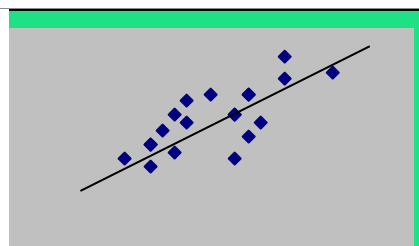


Рисунок 2. Прямолинейное обобщение данных

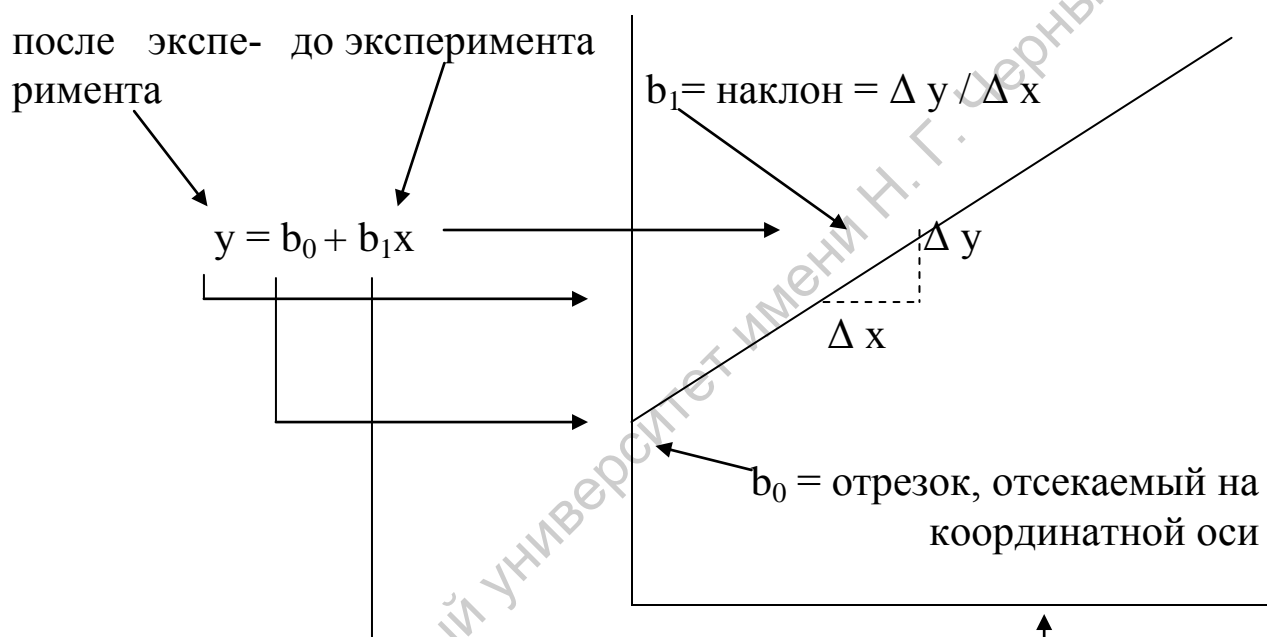
Для того, чтобы запомнить уравнение, которым описывается прямая линия, можно обратиться к школьному курсу алгебры, где часто говорилось о том, что уравнение прямой имеет вид  $y = mx + b$ . В этом уравнении присутствуют следующие компоненты:

$y$  = переменная оси  $y$ , или результат после эксперимента

$x$  = переменная оси  $x$ , или показатели до эксперимента

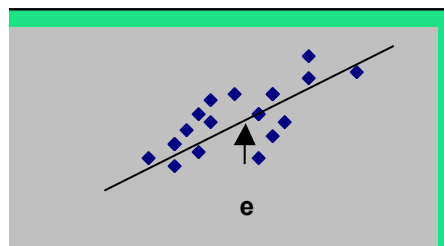
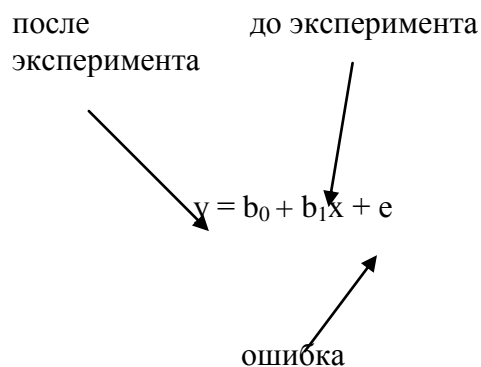
$b_0$  = коэффициент для отрезка, отсекаемого на координатной оси

$b_1$  = наклон линии

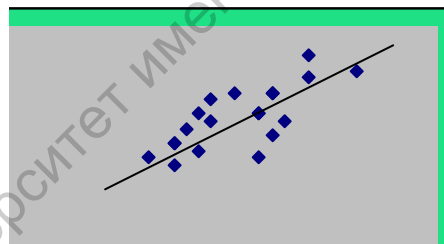
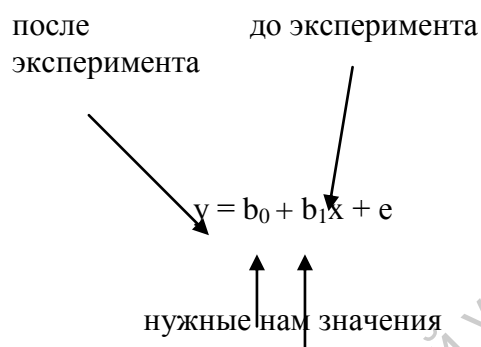


Наклон линии — это изменение показателей после эксперимента по отношению к тому, что было до эксперимента. Как отмечалось выше, этим уравнением туманность из точек на рисунке 1 описывается не полностью. Если бы уравнение описывало облако из точек совершенным образом, тогда бы каждая точка попала на линию. Для того, чтобы описать линию, проходящую через двумерный график, нам потребуется еще один компонент.

Необходимый компонент называется ошибкой. Ошибка — это величина  $e$ , описывающая расстояние по вертикали от прямой линии до каждой точки. Термин «ошибка» выбран потому, что эта величина указывает на степень ошибочности описания каждой точки на графике с помощью уравнения для прямой.



Когда мы применяем к нашим данным двумерную линейную модель, мы имеем по каждому испытуемому количество баллов  $x$  и количество баллов  $y$ . Мы вводим эти пары значений в компьютерную программу. Программа рассчитывает значения  $b_0$  и  $b_1$ . В результате, компьютер выдает нам два числа, соответствующие этим двум значениям.



Линию двумерной регрессии можно представить себе так же, как и любую другую дескриптивную статистику – она всего лишь описывает взаимосвязь между двумя переменными, во многом сходно с тем, каким образом среднее значение описывает центральную тенденцию для одной переменной. И точно так же, как среднее значение не отражает точным образом каждое значение в распределении, линия регрессии тоже не отражает точно каждое значение в двумерном распределении. Мы пользуемся этими показателями лишь потому, что они показывают общее направление в наших данных и позволяют описывать эти общие направления более кратким образом, по сравнению с тем, что нам позволяет делать полное распределение значений.

### Общая линейная модель

После такого краткого представления двумерного случая, мы можем распространить нашу модель на более общие случаи. Общая линейная модель представляет собой примерно то, что мы видели в двумерном варианте этой модели – т. е. уравнение. Отличие от двумерного варианта состоит в том, что каждый из четырех компонентов уравнения может представлять собой не

одну переменную, а несколько переменных. Итак, общая линейная модель записывается следующим образом:

$$y = b_0 + b_1x + e$$

где

$y$  = набор переменных, показывающих состояние после эксперимента

$x$  = набор переменных, показывающих состояние до эксперимента, или ковариаты

$b_0$  = набор отрезков, отсекаемых на координатной оси (каждое значение  $y$ , когда  $x = 0$ )

$b_1$  = набор коэффициентов, по одному для каждого  $x$

Вам необходимо понять, что в данную модель можно включить огромный объем информации. При экспериментальном или при квазиэкспериментальном исследовании, переменная, соответствующая программе либо терапии, соответствует одной или нескольким индикаторным (фиктивным) переменным, каждая из которых представляется в уравнении в виде дополнительного значения  $x$  (обычно для обозначения индикаторных (фиктивных) переменных, которые кодируются символом  $x$ , употребляется символ  $z$ ). Если состояние после нашего эксперимента описывается несколькими переменными, то их включают в комплекс переменных  $y$ . Если состояние до эксперимента описывается несколькими переменными, то их включают в комплекс переменных  $x$ . Для каждого значения  $x$  (и для каждого значения  $z$ ) рассчитывается значение  $b$ , соответствующее взаимосвязи между  $x$  и  $y$ . Числовые соответствия этих значений  $b$ , а также их статистическая проверка, позволяют нам проверять конкретные исследовательские гипотезы относительно взаимосвязей между переменными либо относительно различий между группами.

Общая линейная модель позволяет обобщать широкое многообразие исследовательских результатов. Основная проблема, с которой сталкивается исследователь, применяющий общую линейную модель, состоит в **конкретизации модели**. Исследователь отвечает за приведение общего уравнения к такому конкретному виду, который обеспечивает наилучшее обобщение данных исследования. Если модель не конкретизирована, то возникает вероятность того, что коэффициенты (значения  $b$ ) окажутся искаженными (т. е. неверными), а составленное в результате уравнение не будет точным образом описывать полученные данные. В сложных случаях, проблема конкретизации модели становится серьезной и трудной.

Общая линейная модель – это важнейший инструмент статистического анализа данных. В ней выражается главное достижение психологических исследований в двадцатом веке.

### Вопросы для самоконтроля по теме «Общая линейная модель»

1. В основе чего лежит общая линейная модель?
2. Чему соответствует каждая точка на двумерном графике?
3. Что такое позитивная взаимосвязь?
4. Как создать линейную модель?
5. К чему относится термин «линейная»?
6. Что такое линия регрессии?
7. Какой анализ называется регрессионным?
8. Какой вид имеет уравнение прямой?
9. Что показывает наклон линии регрессии?
10. Что описывает ошибка?
11. Какой величиной обозначается ошибка?
12. Почему был выбран термин «ошибка»?
13. Когда мы применяем к нашим данным двумерную линейную модель, то какие данные по каждому испытываемому нам известны? Что они означают?
14. Что описывает линия двумерной регрессии?
15. Что показывает линия двумерной регрессии?
16. В чем отличие линии регрессии от полного распределения значений?
17. Что представляет собой общая линейная модель?
18. В каком алгебраическом виде записывается общая линейная модель?
19. С чем соотносится переменная, соответствующая программе либо терапии, при экспериментальном или при квазиэкспериментальном исследовании?
20. Каким символом обозначаются в уравнении общей линейной модели индикаторные (фиктивные) переменные, которые кодируются символом  $x$ ?
21. Что включается в комплекс переменных  $y$  уравнения общей линейной модели?
22. Что включается в комплекс переменных  $x$  уравнения общей линейной модели?
23. Какое значение соответствует в общей линейной модели представлению взаимосвязи  $x$  и  $y$ ?
24. Числовые соответствия каких значений следует знать для проверки конкретных исследовательских гипотез относительно взаимосвязей между переменными либо относительно различий между группами?
25. С какой основной проблемой сталкивается исследователь, применяющий общую линейную модель?

## Тестовые задания к теме 11

1. В основе  $t$ -критерия, дисперсионного анализа (ANOVA), ковариационного анализа (ANCOVA), регрессионного анализа и многих многомерных методов, включая факторный анализ, кластерный анализ, многомерное шкалирование, анализ дискриминант функций, канонические корреляции и т. п. заложена

- а) ковариационная модель  
б) теоретическая модель  
в) алгебраическая модель  
г) общая линейная модель  
д) модель мысленного эксперимента
2. Взаимосвязь, при которой испытуемые, балл которых до эксперимента был выше, получили более высокий балл после эксперимента, и наоборот, называется
- а) нулевая взаимосвязь  
б) двусторонняя взаимосвязь  
в) односторонняя взаимосвязь  
г) позитивная взаимосвязь  
д) негативная взаимосвязь
3. Когда мы проводим среди данных линию – мы создаем
- а) ковариационную модель  
б) теоретическую модель  
в) алгебраическую модель  
г) общую линейную модель  
д) модель мысленного эксперимента
4. Изменение показателей после эксперимента по отношению к тому, что было до эксперимента, выражается
- а) сдвигом линии  
б) разворотом линии  
в) наклоном линии  
г) искривлением линии  
д) колоколообразной кривой
5. Если бы уравнение описывало облако из точек совершенным образом, тогда бы каждая точка
- а) попала мимо линии  
б) попала на линию  
в) попала на окружность  
г) попала на периметр  
д) попала на колоколообразную кривую
6. Для того, чтобы описать линию, проходящую через двумерный график, необходим компонент, который называется
- а) интервал  
б) отклонение  
в) оценка  
г) ошибка  
д) наклон
7. Величина, описывающая расстояние по вертикали от прямой линии до каждой точки, называется
- а) интервал  
б) отклонение  
в) оценка

- г) ошибка  
д) наклон
8. На степень ошибочности описания каждой точки на графике с помощью уравнения для прямой, указывает величина, которая называется
- а) интервал  
б) отклонение  
в) оценка  
г) ошибка  
д) наклон
9. Линию двумерной регрессии можно представить себе так же, как и любой другой
- а) коэффициент корреляции  
б) коэффициент отклонения  
в) статистический коэффициент  
г) статистический вывод  
д) дескриптивную статистику
10. Каким образом линия регрессии отражает каждое значение в двумерном распределении
- а) отражает точно  
б) отражает неточно  
в) не отражает совсем  
г) в принципе неважно  
д) затрудняюсь ответить
11. Каким образом среднее значение отражает каждое значение в распределении
- а) отражает точно  
б) отражает неточно  
в) не отражает совсем  
г) в принципе неважно  
д) затрудняюсь ответить
12. Каким образом линия регрессии отражает каждое значение в двумерном распределении
- а) отражает полно  
б) отражает кратко  
в) не отражает совсем  
г) в принципе неважно  
д) затрудняюсь ответить
13. Каким образом среднее значение отражает каждое значение в распределении
- а) отражает точно  
б) отражает кратко  
в) не отражает совсем  
г) в принципе неважно  
д) затрудняюсь ответить
14. Общая линейная модель представляет собой

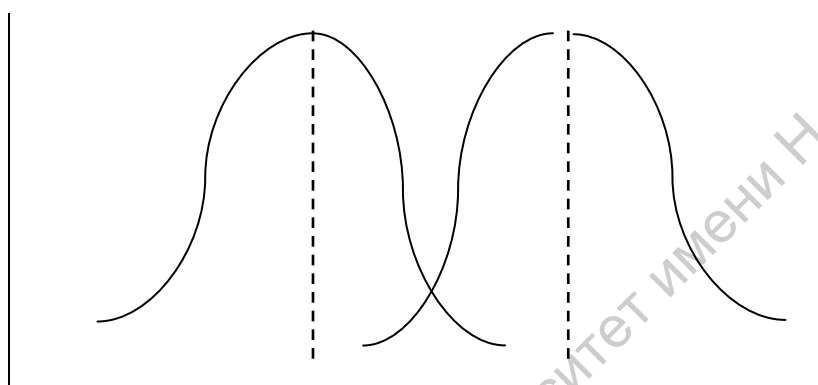


- а) гистограмму
  - б) выборку
  - в) колоколообразную кривую
  - г) паттерн
  - д) уравнение
15. При экспериментальном или при квазиэкспериментальном исследовании, переменная, соответствующая программе либо терапии, соответствует
- а) независимой переменной
  - б) зависимой переменной
  - в) индикаторной (фиктивной) переменной
  - г) шуму
  - д) константе
16. Если состояние до эксперимента описывается несколькими переменными, то их включают в комплекс
- а) переменных  $x$
  - б) переменных  $y$
  - в) переменных  $z$
  - г) индикаторной (фиктивной) переменной
  - д) паттернов
17. Что позволяет совершать с исследовательскими результатами общая линейная модель?
- а) упрощать
  - б) искажать
  - в) уточнять
  - г) обобщать
  - д) вычислять
18. Основная проблема, с которой сталкивается исследователь, применяющий общую линейную модель, состоит в
- а) упрощении модели
  - б) обобщении модели
  - в) искажении модели
  - г) уточнении модели
  - д) объяснении модели
19. Исследователь отвечает за приведение общего уравнения к такому конкретному виду, который обеспечивает наилучшее
- а) упрощение
  - б) искажение
  - в) уточнение
  - г) обобщение
  - д) вычисление

## ТЕМА 12. ПОСТЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ

Для проведения постэкспериментального анализа для двух групп по рандомизированному экспериментальному плану, необходимо соблюдать следующие условия:

- ❖ наличие двух групп
- ❖ проведение постэкспериментальных измерений
- ❖ наличие двух распределений, для каждого из которых следует знать среднее значение и варьирование
- ❖ оценить терапевтический эффект = статистическое (т. е. не случайное) различие между двумя группами

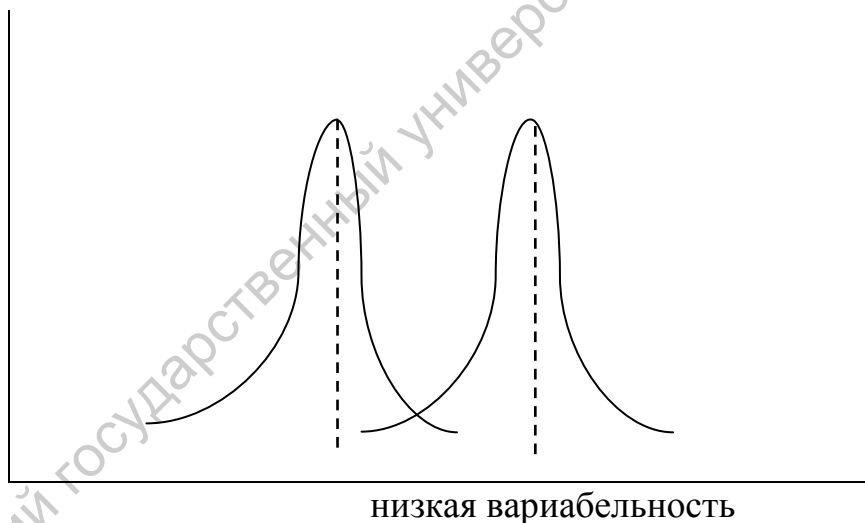
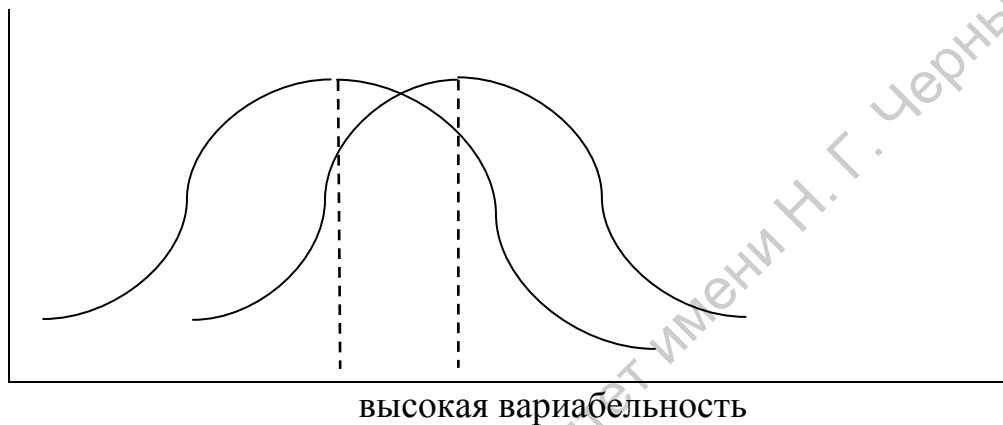
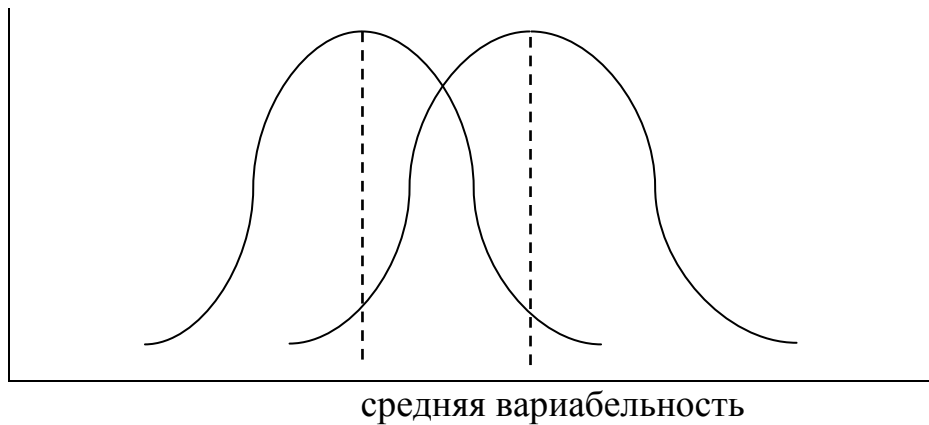


среднее для контрольной группы

среднее для экспериментальной группы

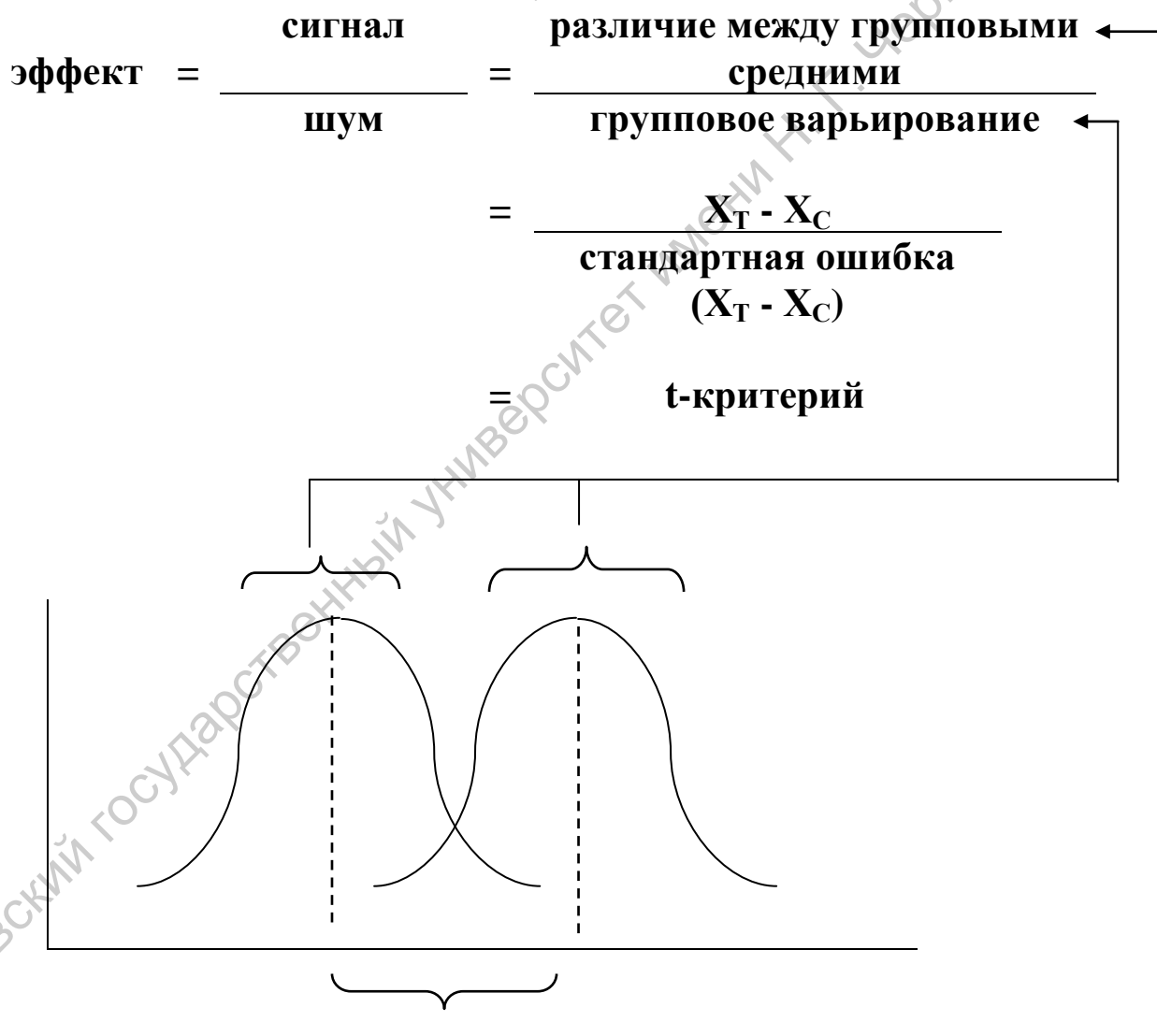
**Существует ли различие?**

Прежде, чем приступить к собственно анализу, полезно понять, что означает термин «различие» в вопросе: «Существует ли различие?» Каждую группу можно представить в виде колоколообразной кривой, описывающей распределение в группе по одной переменной. Колоколообразную кривую можно вообразить в виде сглаженной гистограммы, описывающей частоту каждого возможного значения реакции. Средние значения для каждой группы изображены пунктирными линиями. Различие между средними значениями – это разность по горизонтали между проекциями средних значений в экспериментальной группе и в контрольной группе на горизонтальной оси.



Теперь давайте рассмотрим три возможных результата, которые названы: средняя вариабельность, высокая вариабельность и низкая вариабельность. Обратите внимание на то, что различие между средними во всех трех ситуациях совершенно одинаково. Единственное различие между этими тремя случаями состоит в вариабельности или «растекании» значений вокруг среднего. В каком из этих трех случаев легче всего сделать вывод о том, что средние значения двух групп не одинаковы? Правильный ответ: при низкой вариабельности! Почему легче всего сделать вывод о том, что в случае низкой вариабельности существуют различия между группами? Потому что при низкой вариабельности взаимное наложение двух колоколообразных кривых

оказывается наименьшим. Если посмотреть на пример высокой вариабельности, то приходит в голову, что не столь много существует контрольных групп, в которых размах значений идентичен размаху в экспериментальной группе, и наоборот. Почему это так важно? Потому что, если мы хотим найти различия между двумя группами, то мало будет пользы от простого вычитания одного среднего значения из другого среднего значения. Необходимо учитывать вариабельность вокруг средних значений! Небольшое различие между средними трудно обнаружить, если велика вариабельность (= велик шум). Большое различие между средними значениями удастся легко обнаружить, если вариабельность низкая. Данный подход к рассмотрению межгрупповых различий напрямую соотносится с метафорой «сигнал/шум»: различия более очевидны, когда сигнал сильный, а шум слабый.



Имея это в виду, теперь рассмотрим – как оценивать межгрупповые различия, которые часто называются **размером эффекта**. В числителе дроби – действительное различие между средними значениями. В знаменателе дроби – оценка вариабельности вокруг средних значений. В нашем контексте мы рассчитываем то, что принято называть стандартной ошибкой различия между средними. Стандартная ошибка несет в себе информацию о стандартном отклонении (вариабельности) по каждой из двух групп. Вычисляя значение

доби, мы определяем t-значение, которое описывает различия между группами относительно вариабельности всех значений в группах.

При анализе по постэкспериментальному рандомизированному плану, существуют три различных способа определения терапевтического эффекта. Все три обеспечивают математически эквивалентные результаты, что дает полное основание считать их одним и тем же результатом. Тогда для чего нужны три различных способа? Эти три подхода сформировались во многом независимо друг от друга, и только впоследствии стало очевидно, что мы тремя способами приходим к одному и то же. Что это за способы?

- ❖ Во-первых, мы можем рассчитать независимый t-критерий, как было описано выше.
- ❖ Во-вторых, мы можем рассчитать дисперсионный анализ (ANOVA) для двух независимых групп с одним входом.
- ❖ Наконец, мы можем провести регрессионный анализ, позволяющий выполнять регрессию значений, измеренных после эксперимента, к индикаторным (фиктивным) переменным.

Наиболее общим среди этих трех способов оказывается регрессионный анализ. Со временем, вы придете к пониманию того, что мы здесь рассматриваем все статистические модели для любых экспериментальных и квазиэкспериментальных планов с точки зрения регрессионной модели. Пока просто примите к сведению, что результаты всех трех способов будут одинаковыми.

$$y_i = \beta_0 + \beta_1 Z_i + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена
- $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси
- $\beta_1$  = коэффициент наклона
- $Z_i$  = 1, если  $i$ -тый член относится к экспериментальной группе  
0, если  $i$ -тый член относится к контрольной группе
- $e_i$  = остаток для  $i$ -того члена после вычитания

Итак, у нас имеется статистическая модель, записанная в алгебраическом виде. Вы можете этого не сознавать, но данная формула представляет собой уравнение прямой линии, в которое включена случайная ошибка ( $e_i$ ).

Вы помните школьную алгебру? А школу помните? Для тех, у кого проблемы с памятью, напомним, что уравнение прямой нам преподавали в таком виде:

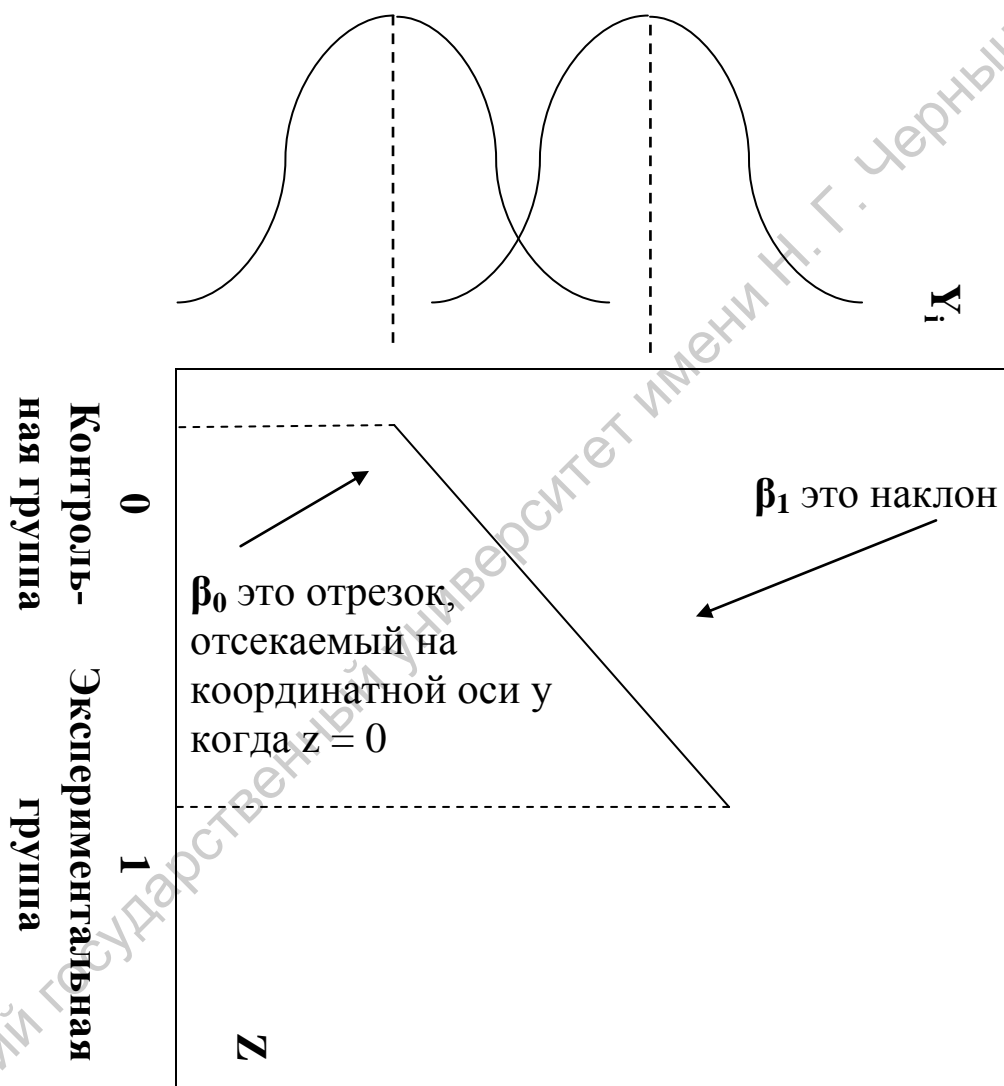
$$y = mx + b$$

Если его переформулировать, тогда получим:

$$y = b + mx$$

Вас нервирует свойство коммутативности? Если вам кажется, что здесь какая-то хитрость, то лучше сделайте перерыв. Съешьте что-нибудь, выпейте кофе, прогуляйте собаку.

Вам должно быть понятно, что  $y_i$  в статистической модели – то же самое, что  $y$  в формуле прямой;  $\beta_0$  – то же самое, что  $b$ ;  $\beta_1$  – то же, что  $m$ ; а  $Z_i$  – то же, что  $x$ . Иначе говоря,  $\beta_0$  – отрезок, отсекаемый на координатной оси, а  $\beta_1$  – наклон.



Важно понять, что наклон  $\beta_1$  – это то же самое, что постэкспериментальное различие между средними для двух групп. Каким образом может наклон оказаться различием между средними? Для того, чтобы это понять, следует взглянуть на график. На графике постэкспериментальные значения представлены на вертикальной оси ( $y$ ). Две колоколообразные кривые, представленные на этом графике, точно такие же, как и на предыдущем графике, за исключением того, что они лежат на левом боку. На горизонтальной оси представлена переменная  $Z$ . Эта переменная имеет всего два значения: 0 – если испытуемый входит в контрольную группу, и 1 – если испытуемый во-

дит в экспериментальную группу. Переменная такого рода называется индикаторной (фиктивной) переменной. Две точки на графике указывают средние постэкспериментальные значения для контрольной группы ( $Z = 0$ ) и для экспериментальной группы ( $Z = 1$ ). Линия, соединяющая эти две точки, нарисована лишь в целях лучшей наглядности – поскольку между 0 и 1 нет других значений, то нет их и на нашем отрезке прямой. Тем не менее, имеет смысл говорить о наклоне нарисованной линии, соединяющей постэкспериментальные значения переменной  $Z$ . Вы помните, как определяется наклон? (Снова приходится вспоминать среднюю школу!) Наклон – это изменение по оси  $y$  при изменении по оси  $x$  (в нашем случае – по оси  $Z$ ). Но мы помним, что изменение по оси  $Z$  для двух групп всегда равняется 1 (т. е.  $1 - 0 = 1$ ). Итак, наклон линии должен быть равен различию между средними значениями по оси  $y$  для двух групп. Именно это мы и хотели изобразить на нашем графике.

$\beta_1$  – это то же самое значение, которое мы бы получили, если бы просто вычли одно среднее значение из другого среднего значения. В нашем случае, поскольку мы задали для экспериментальной группы значение 1, это означает, что мы вычитаем контрольную группу из экспериментальной группы. Положительное значение говорит о том, что среднее значение для экспериментальной группы больше среднего значения для контрольной группы, а отрицательное значение говорит нам об обратном.

Однако, вспомните, что в самом начале мы говорили о том, что в знании различия между средними значениями мало пользы с точки зрения оценивания терапевтического эффекта, потому что необходимо еще принимать во внимание вариабельность или «растекание» значений выборки. Так что нам теперь делать с этим? Любая компьютерная программа регрессионного анализа выдаст, в дополнение к значениям бета, отчет о том, являются ли значения бета статистически значимыми. В отчете будет указано  $t$ -значение, определяющее, отличаются ли значения бета от нуля. И оказывается, что  $t$ -значение для коэффициента  $\beta_1$  – совершенно то же самое число, которое мы бы получили, рассчитывая  $t$ -критерий для двух независимых групп. Оно точно такое же, как квадратный корень из значения  $F$  при дисперсионном анализе (ANOVA) для двух независимых групп с одним входом, потому что  $t^2 = F$ .

Из всего рассмотренного следуют несколько выводов:

- ❖  $t$ -критерий, дисперсионный анализ (ANOVA) с одним входом и регрессионный анализ приводят к одинаковому результату
- ❖ метод регрессионного анализа предполагает применение индикаторной (фиктивной) переменной ( $Z$ ) в отношении экспериментальной группы
- ❖ регрессионный анализ оказывается наиболее общей моделью из трех рассмотренных моделей

### **Вопросы для самоконтроля по теме «Постэкспериментальный анализ»**

1. Какие условия необходимо соблюдать для проведения постэкспериментального анализа двух групп по рандомизированному экспериментальному плану?
2. В каком виде можно представить каждую группу испытуемых?
3. Как выглядят различия между средними значениями?
4. Почему в случае низкой вариабельности легче всего сделать вывод о том, что различия между группами существуют?
5. Почему необходимо учитывать вариабельность вокруг средних значений?
6. Чему соответствует большая вариабельность?
7. Когда различия между группами более очевидны?
8. Как называются межгрупповые различия?
9. Как оценивают размер эффекта?
10. Сколько способов определения терапевтического эффекта существует при анализе по постэкспериментальному рандомизированному плану?
11. Что это за способы? Назовите каждый способ. Что между ними общего?
12. Какой из трех способов является наиболее общим?
13. Какая алгебраическая модель применяется при постэкспериментальном анализе?
14. Что представляет собой данное уравнение?
15. В чем заключается свойство коммутативности?
16. Что такое наклон  $\beta_1$ ?
17. Каким образом может наклон оказаться различием между средними?
18. Сколько значений имеет переменная  $Z$ ?
19. Какие это значения?
20. Как называется такая переменная?
21. Как определяется наклон?
22. Чему равен наклон линии регрессии?
23. Как проще получить значение наклона  $\beta_1$ ?
24. О чем говорит положительное (отрицательное) значение наклона  $\beta_1$ ?
25. Почему необходимо принимать во внимание вариабельность значений выборки?
26. Чему равно  $t$ -значение для коэффициента  $\beta_1$ ?

### Тестовые задания к теме 12

1. Разность по горизонтали между проекциями средних значений в экспериментальной группе и в контрольной группе на горизонтальной оси, называется  
а) сходство между средними значениями



- б) различие между средними значениями  
в) близость между средними значениями  
г) тождественность средних значений  
д) ошибка в средних значениях
2. Колоколообразную кривую можно вообразить в виде  
а) в виде матрицы корреляций  
б) в виде уравнения  
в) в виде гистограммы  
г) в виде доверительного интервала  
д) в виде континуума
3. В каком случае легче всего сделать вывод о том, что средние значения двух групп не одинаковы?  
а) при низкой вариабельности  
б) при высокой вариабельности  
в) при равной вариабельности  
г) при нулевой вариабельности  
д) при искаженной вариабельности
4. При низкой вариабельности взаимное наложение двух колоколообразных кривых оказывается  
а) наибольшим  
б) наименьшим  
в) равным  
г) нулевым  
д) смещенным
5. Небольшое различие между средними трудно обнаружить, если вариабельность  
а) велика  
б) мала  
в) положительна  
г) отрицательна  
д) существует
6. Большая вариабельность означает большой  
а) сигнал  
б) шум  
в) гам  
г) кошмар  
д) валидность
7. Межгрупповые различия более очевидны, когда велик  
а) сигнал  
б) шум  
в) гам  
г) кошмар  
д) валидность
8. Межгрупповые различия более очевидны при слабом  
а) сигнале

- б) шуме  
в) гаме  
г) кошмаре  
д) валидности
9. Стандартная ошибка несет в себе информацию о  
а) выборочном среднем  
б) среднем арифметическом  
в) стандартном отклонении  
г) выборочной совокупности  
д) генеральной совокупности
10. Описывает различия между группами относительно варибельности всех значений в группах  
а) корреляционная матрица  
б) медиана  
в) мода  
г) t-критерий  
д) квотная выборка
11. Наиболее общим способом определения терапевтического эффекта выступает  
а) дисперсионный анализ (ANOVA)  
б) ковариационный анализ (ANCOVA)  
в) регрессионный анализ  
г) факторный анализ  
д) кластерный анализ
12. Формула  $y_i = \beta_0 + \beta_1 Z_i + e_i$  представляет собой  
а) уравнение параболы  
б) уравнение гиперболы  
в) уравнение прямой  
г) уравнение окружности  
д) уравнение химической реакции
13. Если уравнение  $y = mx + b$  переформулировать в виде  $y = b + mx$ , это свойство называется  
а) свойство ассоциативности  
б) свойство валентности  
в) свойство кратности  
г) свойство делимости  
д) свойство коммутативности
14. Постэкспериментальное различие между средними для двух групп выражается  
а) значением для  $i$ -того члена  
б) остатком для  $i$ -того члена после вычитания  
в)  $i$ -тым членом экспериментальной группы  
г)  $i$ -тым членом контрольной группы  
д) коэффициентом наклона

## ТЕМА 13. АНАЛИЗ ПО ФАКТОРНОМУ ПЛАНУ

$$y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{1i} Z_{2i} + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена  
 $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси  
 $\beta_1$  = среднее различие по фактору 1  
 $\beta_2$  = среднее различие по фактору 2  
 $\beta_3$  = среднее различие по фактору 3  
 $Z_{1i}$  = индикаторная (фиктивная) переменная для фактора 1 (0 = 1 час в неделю; 1 = 4 часа в неделю)  
 $Z_{2i}$  = индикаторная (фиктивная) переменная для фактора 2 (0 = аудиторные занятия; 1 = самостоятельная работа)  
 $e_i$  = остаток для  $i$ -того члена после вычитания

Перед вами – формулировка регрессионной модели простого факторного плана  $2 \times 2$ . В этом плане представлены: один фактор, соответствующий времени, затраченному на обучение (1 час в неделю либо 4 часа в неделю) и второй фактор, соответствующий условиям обучения (аудиторные занятия и самостоятельная работа). В данной модели применяются индикаторные (фиктивные) переменные для каждого фактора (обозначенные буквой  $Z$ ). Факторные планы с двумя входами, наподобие представленного здесь плана, учитывают два главных эффекта и одно взаимодействие.

В данной модели, основные эффекты выражаются через статистики, в которых значения бета выступают смежными с переменной  $Z$ . Эффект взаимодействия оказывается статистически связан с  $\beta_3$  (т. е. с  $t$ -значением для данного коэффициента), поскольку он в формуле оказывается смежным с произведением (т. е. с взаимодействием) индикаторных (фиктивных) переменных для обоих факторов, обозначенных буквой  $Z$ . Поскольку здесь имеются две индикаторные (фиктивные) переменные, у каждой из которых существуют два значения, мы можем на основе данной общей модели составить  $2 \times 2 = 4$  отдельных уравнения. Вы можете попробовать расписать уравнения для каждой из четырех ячеек матрицы самостоятельно. Далее, давайте взглянем на различия между группами. Мы можем расписать два уравнения для каждой переменной  $Z$ . В этих уравнениях будут выражаться два главных эффекта. Для того, чтобы увидеть различия между уровнями одного фактора, вычтем эти уравнения одно из другого. Если возникают затруднения, то обратитесь к разделу «Индикаторные (фиктивные) переменные».

### Вопросы для самоконтроля

#### по теме «Анализ по факторному плану»

1. В каком алгебраическом виде формулируется регрессионная модель простого факторного плана  $2 \times 2$ ?

2. Сколько факторов представлено в такой модели?
3. Какие это факторы?
4. Что учитывают подобные факторные планы?
5. При помощи чего в данной модели выражаются основные эффекты?
6. С чем статистически связан эффект взаимодействия? Почему?
7. Самостоятельно распишите уравнения для каждой из четырех ячеек матрицы.

### Тестовые задания к теме 13

1. В регрессионной модели простого факторного плана  $2 \times 2$  для каждого фактора применяются
  - а) независимые переменные
  - б) зависимые переменные
  - в) индикаторные (фиктивные) переменные
  - г) паттерны
  - д) коэффициенты наклона
2. В данной модели, основные эффекты выражаются через статистики, в которых значения бета выступают смежными
  - а) с переменной  $x$
  - б) с переменной  $y$
  - в) с переменной  $z$
  - г) с индикаторной (фиктивной) переменной
  - д) с коэффициенты наклона
3. Эффект взаимодействия оказывается смежным с произведением
  - а) независимых переменных
  - б) зависимых переменных
  - в) индикаторных (фиктивных) переменных
  - г) паттернов
  - д) коэффициентов наклона
4. Поскольку в общей модели имеются две индикаторные (фиктивные) переменные, у каждой из которых существуют два значения, то мы можем на основе данной общей модели составить
  - а) 1 отдельное уравнение
  - б) 2 отдельных уравнения
  - в) 4 отдельных уравнения
  - г) много отдельных уравнения
  - д) ни одного отдельного уравнения
5. Для того, чтобы увидеть различия между уровнями одного фактора, необходимо эти уравнения
  - а) суммировать
  - б) перемножить
  - в) поделить друг на друга
  - г) вычесть одно из другого
  - д) приравнять к нулю

### Семинар 3. Дисперсионный анализ и факторный анализ.

Вопросы для обсуждения:

1. Сущность метода дисперсионного анализа.
2. Основные понятия и символы дисперсионного анализа.
3. Классификация факторов и связь между отдельными видами факторов.
4. Схема решения и основные проблемы факторного анализа.

Темы для рефератов:

1. Условия образования и виды дисперсионных комплексов.
2. История развития факторного анализа.
3. Два вводных примера факторного анализа.
4. Основные уравнения и концепции факторного анализа в алгебраической форме.
5. Схема решения и основные проблемы факторного анализа.
6. Пример расчета факторного анализа.
7. Геометрическая интерпретация модели факторного анализа.
8. Геометрическая интерпретация выделения факторов и метод вращения.
9. Частные коэффициенты корреляции как исходные элементы факторного анализа.

Литература к семинару 3:

1. Иберла К. Факторный анализ. М., 1980. С. 12 – 20; 55 – 93
2. Лакин Г. Ф. Биометрия. М., 1990. С. 155 – 159
3. Аренс Х., Лейтер Ю. Многомерный дисперсионный анализ. М., 1985. С. 16 – 25
4. Павлов Ю. В. Статистическая обработка результатов педагогического эксперимента. М., 1972. С. 4 – 30

## МОДУЛЬ 4

### ТЕМА 14. РАНДОМИЗИРОВАННЫЙ АНАЛИЗ БЛОКОВ

Представим статистическую модель рандомизированного анализа блоков в алгебраической записи для регрессионного анализа. Вот выражение для четырех блоков или четырех гомогенных подгрупп:

$$y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена  
 $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси  
 $\beta_1$  = среднее различие в эксперименте

- $\beta_2$  = коэффициент блокирования для второго блока
- $\beta_3$  = коэффициент блокирования для третьего блока
- $Z_{1i}$  = индикаторная (фиктивная) переменная для эксперимента (0 = контрольная группа; 1 = экспериментальная группа)
- $Z_{2i}$  = индикаторная (фиктивная) переменная для блока 2 (1 = испытуемый принадлежит к блоку 2; 0 = иное)
- $Z_{3i}$  = индикаторная (фиктивная) переменная для блока 3 (1 = испытуемый принадлежит к блоку 3; 0 = иное)
- $Z_{4i}$  = индикаторная (фиктивная) переменная для блока 4 (1 = испытуемый принадлежит к блоку 4; 0 = иное)
- $e_i$  = остаток для  $i$ -того члена после вычитания

Обратите внимание на то, что при конкретизации данной модели мы используем множество индикаторных (фиктивных) переменных. Экспериментальная группа обозначена переменной  $Z_1$ . Блоки второй, третий и четвертый обозначены соответственно переменными  $Z_2$ ,  $Z_3$  и  $Z_4$ . Аналогичным образом, значения бета ( $\beta$ ) отражают экспериментальную группу и блоки второй, третий и четвертый. Что стало с первым блоком в данной модели? Для того, чтобы увидеть вид уравнения для первого блока сравниваемой группы, необходимо подставить индикаторные (фиктивные) переменные и произвести умножение. В нашем случае, все четыре  $Z$  равны нулю, тогда коэффициент для отрезка, отсекаемого на координатной оси ( $\beta_0$ ) будет соответствовать первому блоку контрольной группы. Для первого блока экспериментальной группы,  $Z_1 = 1$ , тогда искомое значение равно  $\beta_0 + \beta_1$ . Подставляя соответствующие индикаторные (фиктивные) переменные «переключатели», удастся вычислить уравнение для любого из блоков экспериментальной группы.

Матрица данных в данном плане анализа состоит из пяти столбцов, а количество строк соответствует количеству сторон, участвующих в эксперименте: постэкспериментальные данные. Один столбец состоит из нулей или единиц для каждой из четырех индикаторных (фиктивных) переменных.

### **Вопросы для самоконтроля по теме «Рандомизированный анализ блоков»**

1. Как представить статистическую модель рандомизированного анализа блоков в алгебраической записи для регрессионного анализа для четырех блоков?
2. Каким символом обозначена экспериментальная группа?
3. Какими переменными обозначены второй, третий и четвертый блоки?
4. Что означают значения бета ( $\beta$ )?
5. Как увидеть вид уравнения для первого блока?
6. Из чего образуется матрица данных в данном плане анализа?

## Тестовые задания к теме 14

1. При конкретизации статистической модели рандомизированного анализа блоков используется множество
  - а) независимых переменных
  - б) зависимых переменных
  - в) индикаторных (фиктивных) переменных
  - г) паттернов
  - д) коэффициентов наклона
2. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $e_i$  означает
  - а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания
  - д) индикаторную (фиктивную) переменную
3. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $\beta_0$  означает
  - а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания
  - д) индикаторную (фиктивную) переменную
4. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $Z$  означает
  - а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания
  - д) индикаторную (фиктивную) переменную
5. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $y$  означает
  - а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания
  - д) индикаторную (фиктивную) переменную

## ТЕМА 15. АНАЛИЗ КОВАРИАЦИЙ

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена  
 $\beta_0$  = коэффициент для отрезка, отсекаемого на координат-

- ной оси
- $\beta_1$  = преэксperimentalный коэффициент
  - $\beta_2$  = среднее различие в эксперименте
  - $X_i$  = ковариация
  - $Z_i$  = индикаторная (фиктивная) переменная для эксперимента (0 = контрольная группа; 1 = экспериментальная группа)
  - $e_i$  = остаток для  $i$ -того члена после вычитания

Статистическая модель плана ковариационного анализа представлена в алгебраической записи для регрессионного анализа. Показанная здесь модель соответствует случаю, при котором имеется одна ковариата и по одной контрольной и экспериментальной группе.

При конкретизации продемонстрированной модели применяются индикаторные (фиктивные) переменные. Индикаторная (фиктивная) переменная  $Z_i$  соответствует экспериментальной группе. Бета-значения ( $\beta$ ) соответствуют оцениваемым параметрам. Значение  $\beta_0$  соответствует отрезку, отсекаемому на координатной оси. В нашей модели, это прогнозируемое постэкспериментальное значение в контрольной группе для заданного значения  $X$ . Если  $X = 0$ , то это отрезок линии регрессии для контрольной группы. Почему? Поскольку в случае контрольной группы  $Z = 0$ , при том, что переменная  $Z$  умножается на  $\beta_2$ , все произведение полностью выпадает.

Применяемая в анализе матрица данных состоит из трех столбцов, а количество строк соответствует количеству сторон, участвующих в эксперименте: постэкспериментальные данные. Один столбец состоит из нулей или единиц, обозначая то, к какой экспериментальной группе принадлежит испытуемый. В отдельной строке отмечается количество ковариат.

Данная модель предполагает, что данные по двум группам хорошо ложатся на прямые линии, а наклон прямых линий одинаков. Если эти условия не выполняются, то модель следует соответствующим образом видоизменить.

### **Вопросы для самоконтроля по теме «Анализ ковариаций»**

1. В каком алгебраическом виде записывается статистическая модель плана ковариационного анализа для одной ковариаты и одной контрольной и экспериментальной группы?
2. Чему в данной модели соответствует индикаторная (фиктивная) переменная  $Z_i$ ?
3. Чему в данной модели соответствует значение  $\beta_0$ ?
4. Из чего состоит применяемая в анализе матрица данных?
5. Что предполагает данная модель?

### **Тестовые задания к теме 15**

1. При конкретизации статистической модели плана ковариационного анализа, применяются



- а) независимые переменные
  - б) зависимые переменные
  - в) индикаторные (фиктивные) переменные
  - г) паттерны
  - д) коэффициенты наклона
2. Экспериментальной группе соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $y_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
3. Ковариации соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $y_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
4. Среднему различию в эксперименте соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $y_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
5. Остатку для  $i$ -того члена после вычитания соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $y_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
6. Значению  $i$ -того члена соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $y_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$

## ТЕМА 16. АНАЛИЗ НЕЭКВИВАЛЕНТНЫХ ГРУПП

### Требования к анализу

**N O X O**

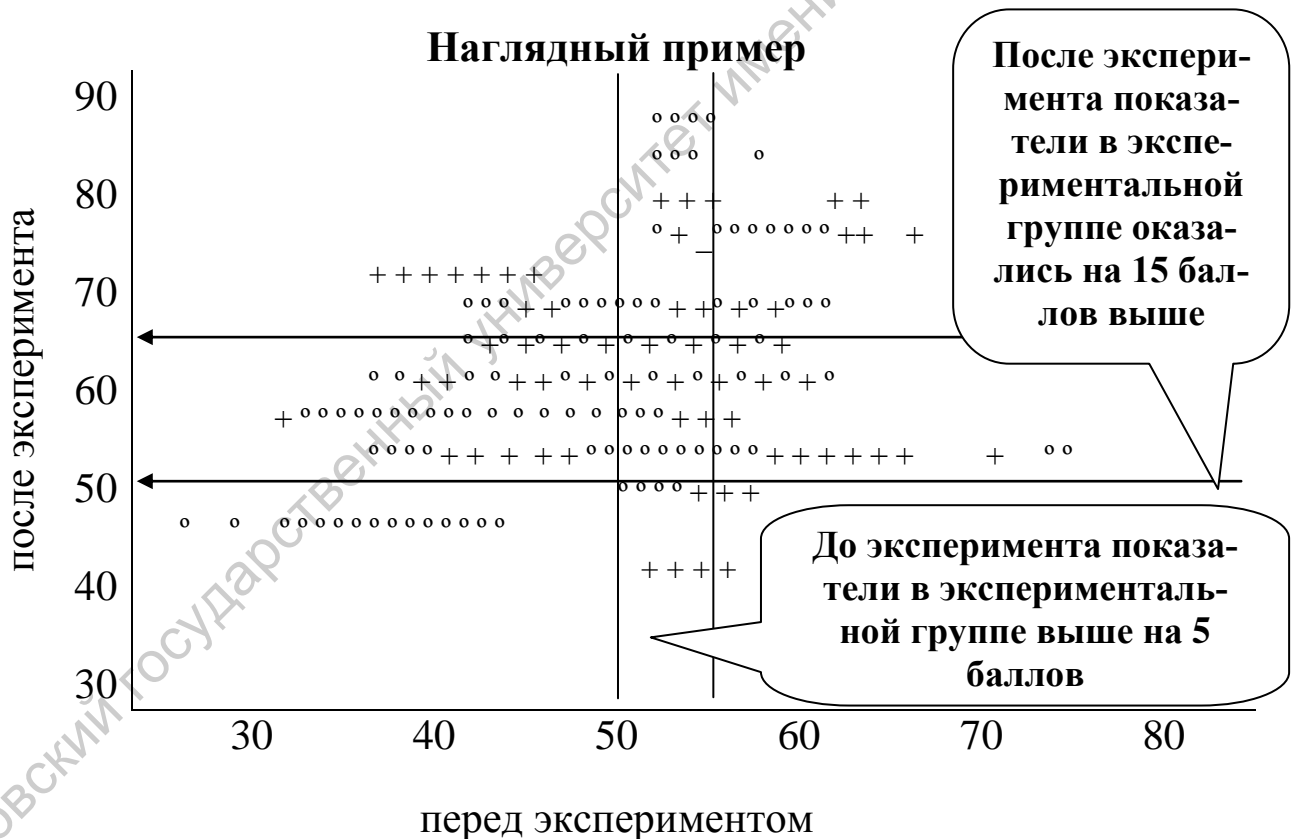
**N O O**

Алгебраическая запись плана для неэквивалентных групп показывает, что у нас имеются две группы: экспериментальная группа и контрольная

группа, а также имеются измеренные для них значения до эксперимента и после эксперимента. Следует подумать о том, что статистическая модель, применимая в данной ситуации, будет включать в себя

- ❖ преэкспериментальную переменную,
- ❖ постэкспериментальную переменную и
- ❖ индикаторную (фиктивную) переменную, описывающую принадлежность испытуемого к той или иной группе.

Мы будем в начале нашего анализа исходить из этих трех переменных. Нас будет интересовать оценивание различий между группами после эксперимента. Сначала нужно откорректировать преэкспериментальные различия. Именно этим занимается ковариационный анализ (ANCOVA), модель которого была описана в связи с рандомизированными экспериментами. При обращении к плану для неэквивалентных групп в связи с описанной моделью возникает лишь одна главная проблема – эта модель не действует! Мы узнаем, почему модель ковариационного анализа (ANCOVA) нас подводит, и как её скорректировать, чтобы она работала верно.



Чтобы понять, что происходит, когда мы применяем ковариационный анализ (ANCOVA) к данным, полученным по плану для неэквивалентных групп, обратимся к гипотетическим данным, сгенерированным с помощью компьютера. Допустим, у нас есть 500 гипотетических испытуемых, 250 человек в экспериментальной группе и 250 человек в контрольной группе. Поскольку у нас план неэквивалентного исследования, пусть наши группы будут неэквивалентны – мы добавим каждому испытуемому в экспериментальной группе 5 баллов к значениям, полученным до эксперимента. Затем мы добавим каждому испытуемому в экспериментальной группе 15 баллов к

значениям, полученным после эксперимента. Когда мы рассматриваем исходное преимущество в пять баллов, мы должны найти эффективность эксперимента, равную десяти баллам.

Далее будем анализировать данные с применением модели ковариационного анализа (ANCOVA). Будем помнить о том, что ковариационный анализ (ANCOVA) верен в том случае, если удастся установить эффект равный 10 баллам. Результаты приводятся в таблице:

предиктор (независимая переменная)	коэффициент	стандартная ошибка	t	p
константа	18,714	1,969	9,50	0,000
до экспери- мента	0,62600	0,03864	16,20	0,000
группа	11,2818	0,5682	19,85	0,000

В нашем анализе, для каждого испытуемого имеются три значения: значение до эксперимента (X), значение после эксперимента (Y) и либо 0 либо 1, обозначающие принадлежность испытуемого либо к контрольной группе (Z = 0) либо к экспериментальной группе (Z = 1). В таблице представлено уравнение, которое рассчитывается в модели ковариационного анализа (ANCOVA). Это уравнение включает в себя три перечисленные переменные (X, Y и Z), а также три коэффициента, которые вычисляются при расчетах по избранной модели.

$$y_i = 18,7 + 0,626X_i + 11,3Z_i$$

Самый важный коэффициент тот, который находится ближе всего к переменной Z. Этот коэффициент отражает средние различия между экспериментальной группой и контрольной группой, потому что этот коэффициент спарен с индикаторной (фиктивной) переменной, показывающей, в какую группу входит испытуемый). Значение должно быть равно 10, потому что было задано различие в десять баллов. При проведении анализа мы находим, что значение равно 11,3 (либо 11,2818, если говорить более точно).

Можно сказать, что это неплохой результат. Он почти равен 10 баллам. Нужно лишь определить, есть ли статистически значимое различие между полученным значением, равным 11,2818 и истинным значением, равным 10. Для того, чтобы это проверить, нужно задать доверительный интервал вокруг рассчитанного нами значения и проверить различие между 11,2818 и 10 в связи с вариабельностью данных. К счастью, компьютер делает это за нас автоматически. Если взглянуть в таблицу, то мы увидим, что в третьей строке имеются

- коэффициент, относящийся к различиям между группами,

- стандартная ошибка для этого коэффициента (показатель вариабельности),
- t-значение и
- значение вероятной погрешности p.

Все t-значения показывают, что коэффициент 11,2818 статистически отличен от нуля. Однако, нам нужно узнать, отличается ли он от истинного значения эффекта, равного 10. Для того, чтобы это выяснить, зададим доверительный интервал вокруг t-значения, используя стандартную ошибку.

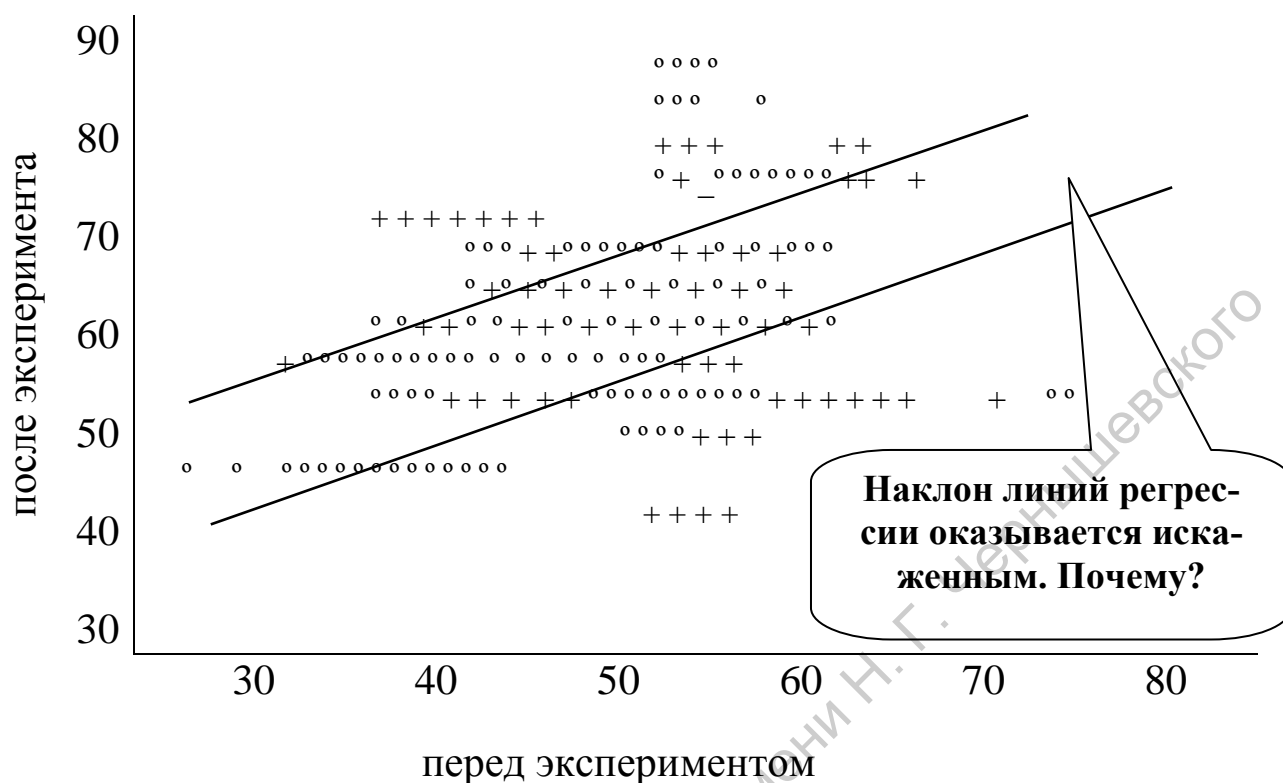
$$\begin{aligned}
 * \text{ДИ}_{0,95(\beta_2=10)} &= \beta_2 \pm 2 \text{СО}(\beta_2) \\
 &= 11,2818 \pm 2(0,5682) \\
 &= 11,2818 \pm 1,1364
 \end{aligned}$$

$$* \text{ДИ} = 10,1454 \text{ до } 12,4182$$

Мы знаем, что 95%-ный доверительный интервал равен коэффициенту плюс/минус два значения стандартной ошибки. Расчеты показывают, что 95%-ный доверительный интервал для нашего коэффициента 11,2818 имеет границы от 10,1454 до 12,4182. Любое значение, попадающее в эти границы, нельзя с вероятностью более 95% считать отличающимся от полученного нами значения 11,2818. Однако, истинное значение, равное 10, выходит за эти границы. Иначе говоря, полученное нами значение 11,2818 статистически отличается от истинного значения. Говоря еще иначе, результаты проведенного анализа оказываются искаженными – мы получили ложный ответ. В нашем примере, оцениваемый нами эффект оказался значительно больше истинного эффекта (даже при том, что различие между 10 и 11,2818 не представляется слишком большим, оно превышает случайный уровень). Поэтому, при применении той модели анализа, которую нам подсказывает наша интуиция, к плану неэквивалентных групп, возникает проблема. Для того, чтобы понять, почему возникает искажение, следует поглубже взглянуть на то, как работает статистический анализ применительно к плану неэквивалентных групп.

### Проблема

Почему ковариационный анализ (ANCOVA) оказывается искаженным при обращении к плану неэквивалентных групп? И почему он не бывает искаженным при обращении к рандомизированному преэксperimentalному/постэксperimentalному плану?



В действительности, происходят несколько вещей, вызывающих искажения, которые не совсем легко поддаются пониманию. Вот две причины, по которым возникает искажение:

- ❖ ошибка в преэкспериментальных измерениях, приводящая к сглаживанию наклонов линий регрессии
- ❖ неэквивалентность групп

Первая проблема возникает также в рандомизированных планах исследований, но не приводит к искажениям экспериментального эффекта, поскольку группы в эксперименте эквивалентны (по крайней мере, есть вероятность их эквивалентности). Проблему вызывает сочетание двух условий. Понимание причин проблемы ведет к её решению.

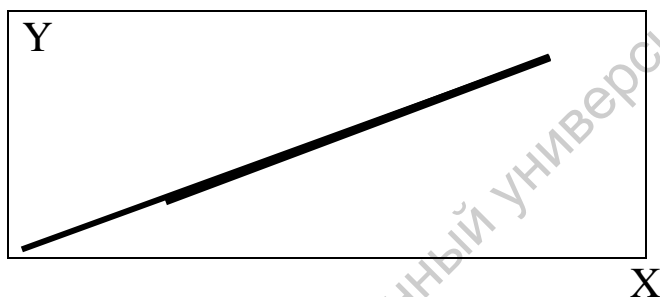
**Регрессия и ошибка измерения.** Начнем наши попытки понять источник искажения с рассмотрения того, каким образом ошибка измерения отражается на регрессионном анализе. Рассмотрим три различных сценария ошибочного измерения для того, чтобы посмотреть, к чему приводит ошибка. Во всех трех сценариях будем предполагать, что истинный экспериментальный эффект отсутствует, т. е. верна нулевая гипотеза.

- ❖ Первый сценарий – это случай, когда ошибки измерения нет вообще. В таком гипотетическом случае все точки расположатся прямо на линиях регрессии.
- ❖ Второй сценарий – отсутствует ошибка измерения до эксперимента, но возникает ошибка измерения после эксперимента. В этом случае точки на графике будут смещены вертикально – вверх и вниз от линий регрессии. Представьте себе особый случай – в нашем исследовании принимает участие всего один испытуемый. Если бы не было ошибки измерения вообще, то баллы нашего испытуемого попали

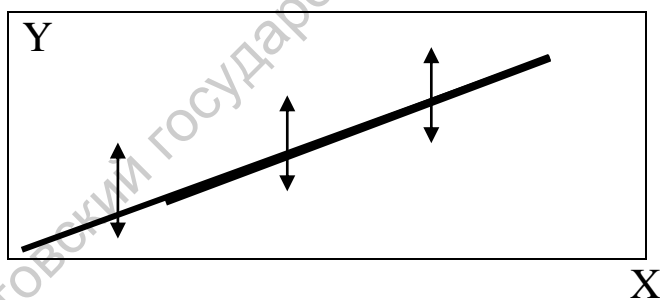
бы прямо на линию регрессии. Если имеется ошибка в измерении после эксперимента, то баллы нашего испытуемого будут отличаться от линии регрессии либо в лучшую сторону, либо в худшую сторону относительно истинного измеренного значения. А это приводит к смещению значений на графике по вертикали.

- ❖ При третьем сценарии ошибка в измерении существует только перед экспериментом. В этом случае смещение возникает по горизонтали – влево либо вправо относительно линий регрессии.

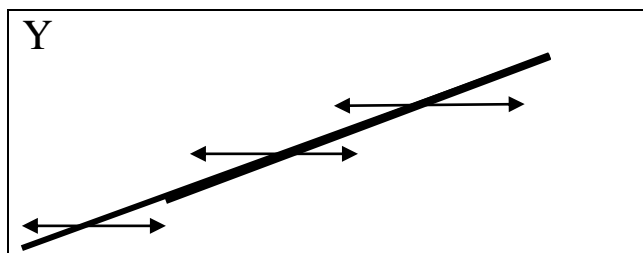
Все три этих гипотетических случая, ни один из которых не имеет места в реальности, схематически представлены на графиках:



**Ошибки измерения нет**



**Ошибка измерения только после эксперимента**

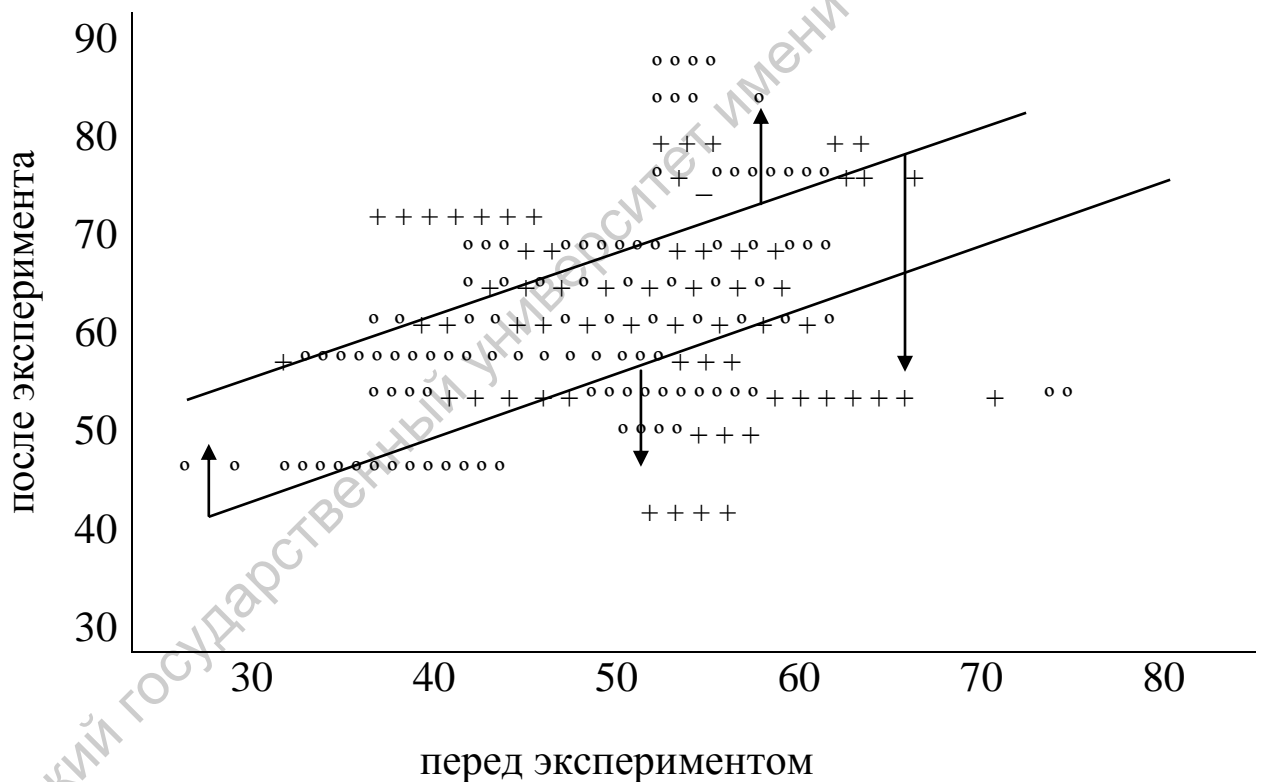


**Ошибка измерения только до эксперимента**

## Х

**Каким образом регрессия вписывается в линии.** Регрессионный анализ – это процедура с *наименьшим количеством квадратов*. Действующий критерий для того, чтобы вписаться в линии, заключается в сведении к минимуму суммы квадратов разностей, остающихся при вычитании из линии регрессии. Давайте немного упростим эту фразу. Здесь ключевое слово – «остающиеся». Остатком от вычитания называется вертикальное расстояние от линии регрессии до каждой точки.

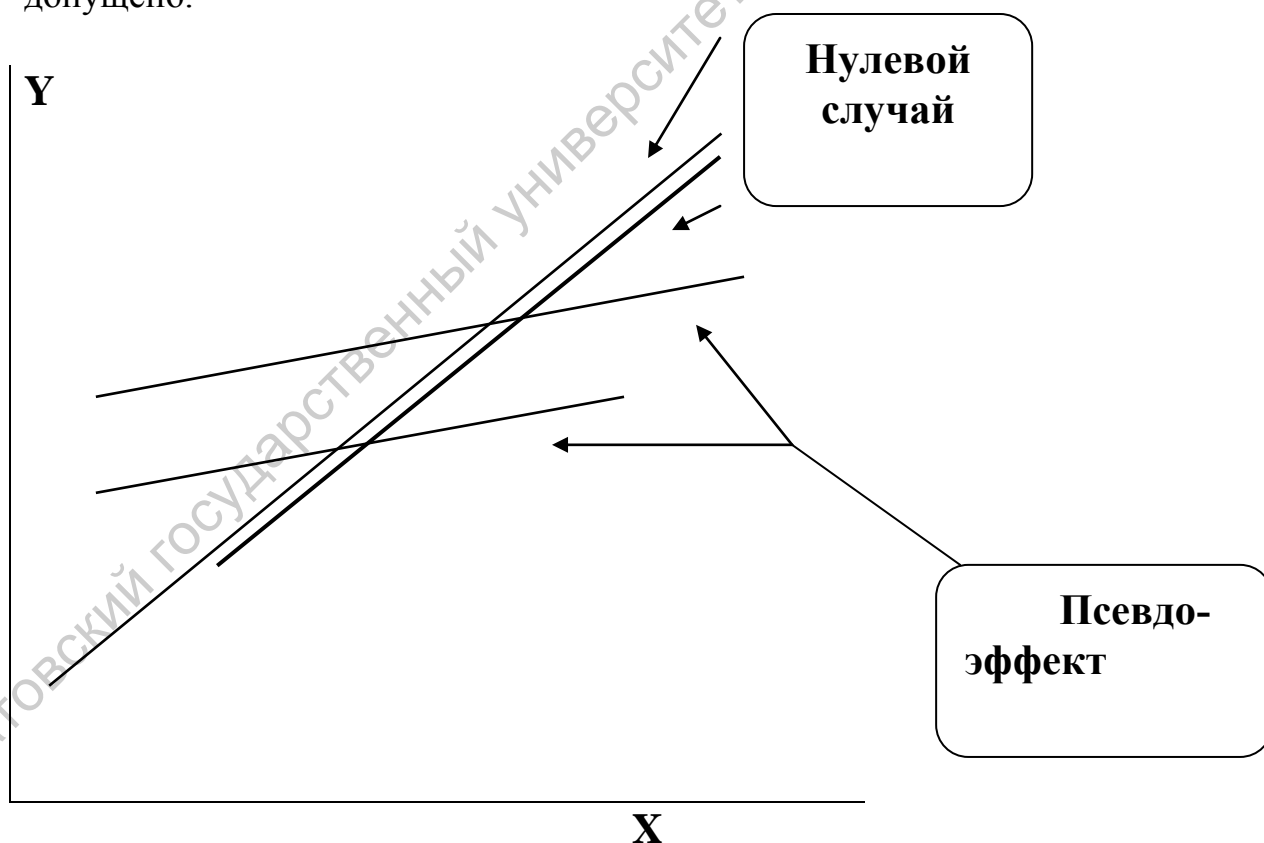
На графике изображены четыре таких остаточные разности, по две для каждой группы. Две остаточные разности оказываются ниже линии регрессии, и две оказываются выше линии регрессии. Каков критерий проведения линии через облако из точек? Возьмем все остаточные разности в группе (будем проводить отдельно линию для контрольной группы и отдельно линию для экспериментальной группы).



Если они окажутся выше линии, то это положительные значения, а если они окажутся ниже линии, то это отрицательные значения. Возведем все остатки разностей группы в квадрат. Вычислим сумму квадратов остаточных разностей – просто сложим их. Вот и все. Регрессионный анализ предполагает проведение линии через те данные, которые образуют наименьшую сумму квадратов остатков разностей. Каким образом это делается – другой вопрос. Важно понять, *что* делается. Важно заметить, что *линия регрессии проводится через остаточные разности, а остаточные разности представляют собой вертикальные отклонения от линии регрессии.*

**Каким образом ошибка измерения влияет на наклон.** Теперь мы готовы объединить вместе идеи, высказанные в двух первых подразделах. Снова обратимся к трем сценариям ошибочного измерения, описанным выше. Там, где ошибка измерения отсутствует, ничто не влияет на наклон линий регрессии. Это видно на рисунке. Обратим внимание на то, что ни на одном из трёх рисунков, иллюстрирующих сценарии, не просматривается экспериментальный эффект. О каком-либо эффекте можно говорить лишь в том случае, если между двумя линиями наблюдается вертикальное смещение.

Теперь рассмотрим случай, когда ошибка измерения допущена после эксперимента. Повлияет ли он на наклон линий регрессии? Ответ отрицательный. Почему? Потому что при регрессионном анализе линии регрессии проводятся с учетом смещения точек по вертикали. Ошибка измерения, допущенная после эксперимента, влияет на вертикальный параметр и, если ошибки рандомизированы, то мы получим столько остатков разностей, смещающих наклон линии одновременно и вверх и вниз, что в среднем наклон линии останется прежним, как и в случае нулевого сценария. В случае сценария ошибки измерения после эксперимента, мы имеем более существенную вариабельность данных вокруг линии регрессии, но сама линия регрессии останется на прежнем месте, как если бы никакой ошибки измерения не было допущено.

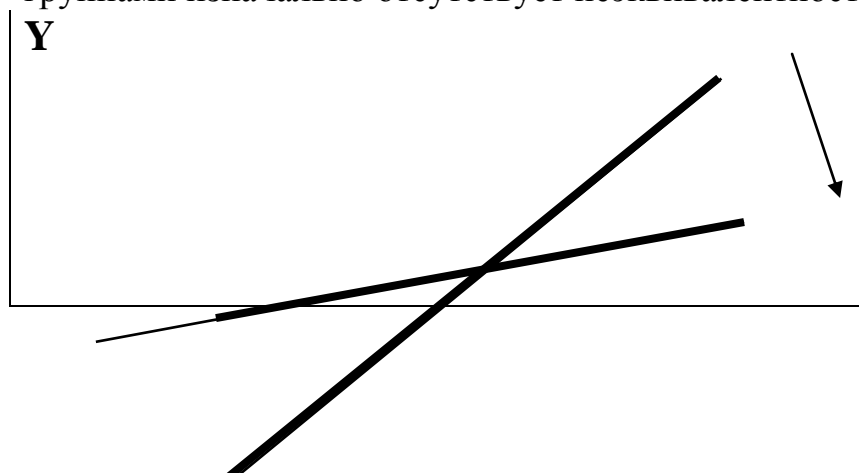


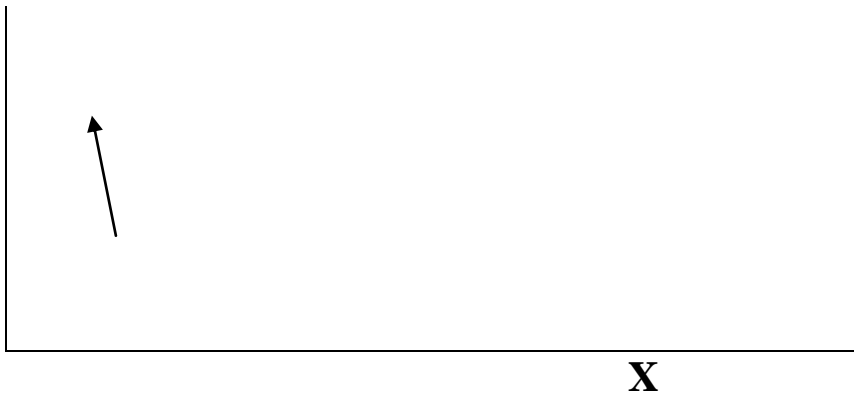
Теперь давайте рассмотрим случай, когда ошибка измерения допущена до эксперимента. При данном сценарии ошибки добавляются по горизонтальному параметру. Однако, регрессионный анализ предполагает проведение линий в соответствии с вертикальным смещением. Как это отразится на наклоне линий регрессии? То, что будет происходить, проиллюстрировано на



рисунке. Если бы ошибки измерения не было, то линии регрессии накладывались бы друг на друга – это так называемый «нулевой случай». Когда мы прибавляем ошибку измерения, допущенную до эксперимента, то мы фактически удлиняем горизонтальное измерение, не влияя при этом на вертикальное измерение. Поскольку регрессионный анализ соотносится с вертикальным измерением, это заставит линию регрессии вытянуться в соответствии с удлиненным горизонтальным измерением. Единственный способ совершить такое растяжение – совершить вращение вокруг центральной точки. В результате линия становится более плоской и вытянутой – наклон линии уменьшается там, где допущена ошибка измерения до эксперимента, по сравнению с истинным значением. Постарайтесь увидеть то, как линия каждой группы становится более плоской за счет её поворота вокруг собственного центра, в результате между двумя линиями возникает смещение, которого изначально между ними не было. Несмотря на то, что в оригинале никакого эффекта не существовало, возник фальшивый эффект или псевдоэффект. Смещенная величина наклона, вызываемая ошибкой измерения, допущенной до эксперимента, образует ложный экспериментальный эффект. В данном примере мы наблюдаем эффект там, где никакого эффекта нет. Когда мы производили расчеты, то мы видели, что псевдоэффект превысил истинный эффект.

**Почему эта проблема не возникает при рандомизированных планах?** Итак, почему такой псевдоэффект не возникает при рандомизированном плане ковариационного анализа? На следующем рисунке показано, что даже при рандомизированном плане ошибка измерения, допущенная до эксперимента, на самом деле вызывает сглаживание наклона линий. Однако, в рандомизированном случае псевдоэффекта не возникает, даже несмотря на сглаживание линий. Почему? Потому что в рандомизированном случае две группы до эксперимента эквивалентны – между линиями отсутствуют различия по горизонтали. В нулевом случае линии превосходно накладываются друг на друга. Поэтому когда возникает сглаживание, оно будет одинаковым для обеих линий, и вертикального смещения между линиями не возникнет. Сравните этот рисунок с предыдущим. Вы должны понять, что различие состоит в том, что в первом случае при плане исследования неэквивалентных групп существует изначальная неэквивалентность между группами, вследствие чего возникает сглаживание наклонов. При этих обстоятельствах сглаживание линий приводит к искажению. В рандомизированном случае сглаживание также имеет место, однако не возникает искажения, потому что между группами изначально отсутствует неэквивалентность.





**Обобщение проблемы.** Итак, как нам справиться с проблемой? Статистическая модель ковариационного анализа (ANCOVA) на первый взгляд представляется содержащей все верные компоненты для верного моделирования данных плана исследования неэквивалентных групп. Однако, мы установили, что она не срабатывает верным образом – величина эффекта оказывается искаженной. Когда мы исследовали, почему это так, мы увидели, что искажение вызывается двумя главными факторами:

- ❖ сглаживанием наклона в результате ошибки в измерениях до эксперимента и
- ❖ изначальной неэквивалентностью между группами.

Причина проблемы не в ошибке измерения после эксперимента, что объясняется тем критерием, на основании которого проводится линия регрессии при регрессионном анализе. Проблемы не возникает в рандомизированных экспериментах, поскольку в них отсутствует неэквивалентность перед экспериментом.

Исходя из этих соображений, мы можем предположить, что чем больше неэквивалентность между группами, тем сильнее окажется искажение – чем менее группы подобны друг другу, тем больше окажется проблема. При реальном, а не сконструированном исследовании, ошибка измерения возникает и до эксперимента, и после эксперимента – точности измерений добиться трудно. Итак, мы теперь знаем, что при плане исследования неэквивалентных групп ковариационный анализ (ANCOVA) приводит к искаженным результатам.

### **Решение проблемы**

Теперь, когда нам понятна проблема, возникающая при плане исследования неэквивалентных групп, постараемся эту проблему решить. Поскольку проблема вызывается ошибкой измерения, один из вариантов решения проблемы состоит в преодолении ошибочности наших измерений. Если бы нам удалось устранить ошибку измерения и приблизиться к тому случаю, когда ошибка в измерении до эксперимента отсутствует, тогда бы не происходило сглаживания линий регрессии и не возникал бы псевдоэффект. Для того, чтобы понять, каким образом можно корректировать ошибку в измерениях до

эксперимента, нам потребуется вспомнить всё, что нам известно об ошибках в измерениях и их отношении к надежности измерения.

Надежность может быть выражена следующей дробью:

$$\frac{\text{var} (T)}{\text{var} (T) + \text{var} (e)}$$

где  $T$  – истинная способность либо уровень измерения, а  $e$  - ошибка измерения. Отсюда следует, что надежность измерения до эксперимента прямо соотносится с объемом ошибок в измерении.

- ❖ Если ошибка в измерении до эксперимента отсутствует, то величина  $\text{var} (e)$  в знаменателе равна нулю, а надежность измерения равна 1.
- ❖ Если в измерениях до эксперимента содержатся только исключительно ошибки, тогда величина  $\text{var} (T)$  равна нулю, а надежность измерения нулевая. Т. е., если в измерении нет ничего, кроме ошибок, тогда измерение полностью ненадежно.
- ❖ Если половина измерений истинна, а другая половина измерений ошибочна, тогда надежность составляет 0,5.

Итак, мы видим, что существует прямая взаимосвязь между ошибкой измерения и надежностью измерения – в надежности отражается доля ошибок в вашем измерении.

Поскольку ошибка измерения до эксперимента – это необходимое условие искажения при исследовании по плану неэквивалентных групп (если отсутствует ошибка измерения до эксперимента, то отсутствует искажение даже при плане исследования неэквивалентных групп), если мы скорректируем ошибку измерения, то мы скорректируем и искажение. Однако, мы не можем увидеть ошибку измерения непосредственно из наших данных – никто не знает, каково должно быть истинное значение и какова в связи с ним ошибка. Тем не менее, мы в состоянии рассчитать надежность измерения. Поскольку надежность измерения напрямую связана с ошибкой измерения, мы можем при помощи расчета надежности вычислить – насколько в измерении присутствует ошибка. И мы в состоянии скорректировать баллы, измеренные до эксперимента, зная величину надежности, что позволяет нам скорректировать сглаживание линий регрессии, и благодаря этому мы в состоянии устранить искажение из исследования по плану для неэквивалентных групп.

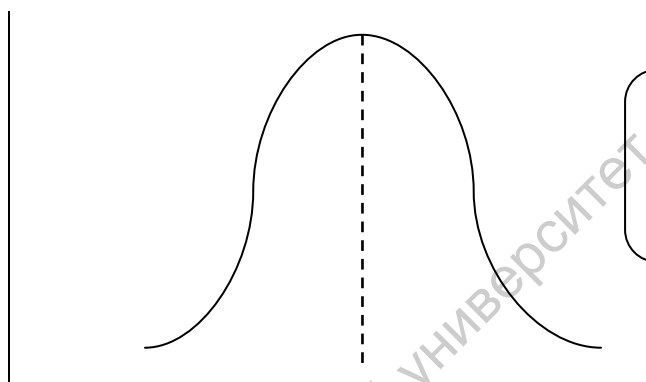
**Ковариационный анализ (ANCOVA) со скорректированной надежностью.** Мы пытаемся устранить эффект искажения ковариационного анализа (ANCOVA) для неэквивалентных групп, применяя коррекцию «надежности», что обеспечивает коррекцию ошибки измерения, допущенной до проведения эксперимента. На рисунке изображено, как выглядит коррекция надежности. На верхнем графике показано наблюдаемое распределение значений перед экспериментом, в которое включена ошибка измерения. Вспомним,

выше было сказано, что добавление ошибки измерения расширяет или удлиняет горизонтальное измерение в двумерном распределении.

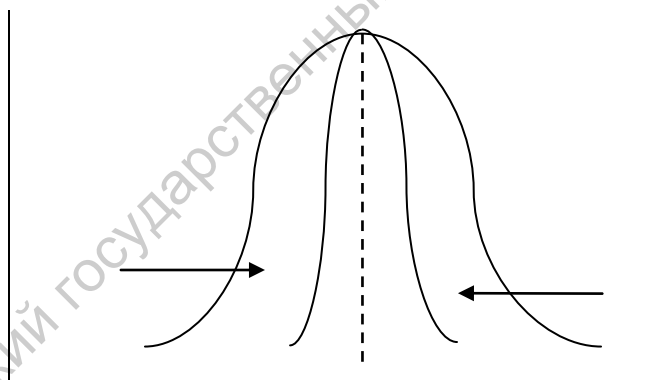
На показанном на верхнем графике частотном распределении мы видим, что распределение оказывается шире по сравнению с тем, каким оно должно быть при отсутствии ошибки измерения.

На втором графике то, чего мы хотим добиться при коррекции баллов, измеренных до эксперимента – мы хотим сдвинуть распределение значений внутрь за счет величины, пропорциональной ошибке измерения, на которую растянулось распределение. Коррекцию значений проводят отдельно для экспериментальной группы и отдельно для контрольной группы.

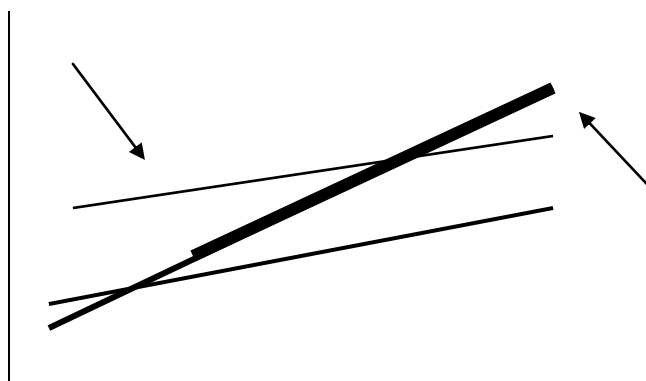
На третьем графике показано, какое воздействие «сдавливание» оказывает на линии регрессии – наклон линий увеличивается, линии поворачиваются в обратном направлении, как они и должны располагаться, устраняется искажение, вызванное ошибкой измерения. В общем, выполняется всё то, что противоположно действию ошибки измерения.



**Исходное распределение значений до эксперимента**



**Скорректированное распределение значений до эксперимента**



**Скорректированная регрессия**

Всё, что нам необходимо знать – насколько следует сжать распределений значений, измеренных до эксперимента, чтобы верно скорректировать ошибку измерения. Поскольку *надежность измерения представляет собой величину той доли, в которой измеренные значения являются истинными по отношению к ошибочным значениям*, она должна подсказать нам насколько нужно «сжимать». В действительности, формула для коррекции ошибки измерения весьма проста:

$$X_{\text{кор}} = \bar{X} + r (X - \bar{X})$$

где

$X_{\text{кор}}$  = скорректированное значение, измеренное до эксперимента

$\bar{X}$  = оригинальное значение, измеренное до эксперимента

$r$  = надежность

Идея данной формулы заключается в том, что мы должны создать новые значения для доэкспериментальной ситуации по каждому испытуемому. Эти новые значения должны быть «скорректированы» с поправкой на ненадежность измерения до эксперимента на величину, пропорциональную надежности. Балл каждого испытуемого должен стать ближе к среднему значению в группе. Данная формула показывает нам – насколько именно ближе. Рассмотрим несколько примеров.

Во-первых, рассмотрим случай, когда ошибка измерения до эксперимента отсутствует. В этом случае надежность равна 1. В этом случае нам совсем не требуется производить корректировку данных измерения. Допустим, что у нас имеется испытуемый, балл которого до эксперимента равен 40, тогда как среднее значение для групп до эксперимента равно 50. Тогда вероятна следующая корректировка:

$$X_{\text{кор}} = 50 + 1 (40 - 50)$$

$$X_{\text{кор}} = 40$$

Либо, иными словами, никакой корректировки не производится. В этом особенность случая с нулевой ошибкой измерения до эксперимента.

Теперь, предположим, что надежность относительно низкая, скажем 0,5. Если у испытуемого измеренный до эксперимента балл равен 40, тогда как средний балл в группе равен 50, тогда получим:

$$X_{\text{кор}} = 50 + 0,5 (40 - 50)$$

$$X_{\text{кор}} = 45$$

Либо, когда надежность составляет 0,5, то нам следует передвинуться в направлении среднего значения на половину отделяющего нас расстояния (на половину от оригинального значения в направлении среднего, или на 45).

Наконец, предположим, что в том же самом случае надежность оказалась сильнее и равняется 0,8. Тогда корректировка надежности выглядит следующим образом:

$$X_{\text{кор}} = 50 + 0,8 (40 - 50)$$

$$X_{\text{кор}} = 42$$

То есть, при надежности равной 0,8, нам следует скорректировать измеренное значение на 20% в направлении медианы (поскольку, если надежность равна 0,8, то объем влияния ошибки измерения на измеренное значение составляет  $1 - 0,8 = 0,2$ ).

Важно понять, что, если применить подобную корректировку ко всем баллам в группе, измеренным до эксперимента, тогда распределение доэкспериментальных значений окажется «сжатым» на величину, пропорциональную ошибке измерения (1 минус надежность). Важно отметить, что такого рода корректировка проводится отдельно для контрольной группы и отдельно для экспериментальной группы.

Теперь мы готовы к тому, чтобы взять скорректированные значения и заменить ими оригинальные доэкспериментальные значения в модели ковариационного анализа (ANCOVA):

$$y_i = \beta_0 + \beta_1 X_{\text{кор}} + \beta_2 Z_i + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена
- $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси
- $\beta_1$  = преэкспериментальный коэффициент
- $\beta_2$  = среднее различие в эксперименте
- $X_{\text{кор}}$  = ковариата, скорректированная на надежность
- $Z_i$  = индикаторная (фиктивная) переменная для эксперимента (0 = контрольная группа; 1 = экспериментальная группа)
- $e_i$  = остаток для  $i$ -того члена после вычитания

Обратим внимание на то, что единственным отличием является то, что в оригинальной формуле ковариационного анализа (ANCOVA) мы заменили  $X_i$  символом  $X_{\text{кор}}$ .

### Повторное рассмотрение примера

Итак, давайте посмотрим на то, насколько хорошо работает наша корректировка. Используем данные из нашего примера. Результаты следующие:

предиктор (независимая переменная)	коэффициент	стандартная ошибка	t	p
константа	-3,141	3,300	-0,95	0,342
до экспери- мента	1,06316	0,06557	16,21	0,000
группа	9,3048	0,6166	15,09	0,000

$$y_i = -3,14 + 1,06X_{\text{кор}} + 9,03Z_i$$

$$\begin{aligned} * \text{ДИ}_{0,95(\beta_2=10)} &= \beta_2 \pm 2 \text{CO}(\beta_2) \\ &= 9,3048 \pm 2(0,6166) \\ &= 9,3048 \pm 1,2332 \end{aligned}$$

$$* \text{ДИ} = 8,0716 \text{ до } 10,5380$$

На сей раз, мы получаем значение экспериментального эффекта равное 9,3048 (вместо 11,2818). Это значение оказывается ближе к истинному значению, равному 10 баллам. Когда мы задаем интервал доверительности для нашего скорректированного значения на уровне 95%, мы видим, что истинное значение, равное 10 баллам, попадает в этот интервал. То есть, наш анализ позволил определить экспериментальный эффект, не отличающийся от истинного эффекта – это неискаженное значение.

Следует еще раз сравнить наклон линии скорректированной модели с оригинальным наклоном. Сейчас наклон стал равен почти 1, т. е. 1,06316, тогда как ранее он был равен 0,626 – значительно ниже или «более плоский» наклон. Наклон в нашей скорректированной модели приближается к расчетному истинному наклону линий (который есть 1). Оригинальный наклон продемонстрировал сглаживание, вызванное ошибкой измерения до эксперимента.

Итак, модель ковариационного анализа (ANCOVA) со скорректированной надежностью применяется в статистическом анализе по плану неэквивалентных групп с целью исправления смещения, вызываемого ошибкой измерения до эксперимента.

### Какую надежность предпочесть?

Для того, чтобы завершить данный раздел, нам осталось решить ещё один серьезный вопрос. Из теории надежности нам известно, что истинную надежность вычислить невозможно, её можно только оценить. Существует множество оценок надежности, и все они приводят к разным значениям. Альфа Кронбаха обеспечивает верхнюю границу оценки надежности. Тест-ретестовая надежность – нижнюю границ оценки надежности. Так что же предпочесть для нашей формулы коррекции надежности? Ответ: и то, и дру-

гое! Когда проводится анализ по плану неэквивалентных групп, то, чтобы обезопасить себя, следует выполнить оба анализа – определить и верхнюю, и нижнюю границы оценки надежности. Если удастся обнаружить значительный экспериментальный эффект с помощью обеих оценок, то можно быть совершенно уверенным в том, что действительно удалось обнаружить значительный эффект на основании данных, которые не содержат ошибки измерения.

Если такое завершение нашей головоломной истории про анализ неэквивалентных групп не выглядит удовлетворяющим, то давайте взглянем на неё как на цену, которую приходится платить при проведении рандомизированного исследования – наш анализ становится более комплексным, поскольку мы вынуждены производить необходимую коррекцию по причине неэквивалентности групп. Тем не менее, обращение к неэквивалентным группам обладает своими собственными преимуществами. Пусть исследователь сам решает, что для него более предпочтительно.

### **Вопросы для самоконтроля по теме «Анализ неэквивалентных групп»**

1. Что включает в себя статистическая модель анализа плана для неэквивалентных групп?
2. Что позволяет оценить данная модель?
3. Какие различия позволяет устанавливать ковариационный анализ (ANCOVA)?
4. Какая проблема возникает в связи с моделью ковариационного анализа (ANCOVA) при обращении к плану для неэквивалентных групп?
5. Какие переменные включает в себя уравнение модели ковариационного анализа (ANCOVA)?
6. Какой коэффициент этого уравнения самый важный? Что он отражает?
7. Что показывает индикаторная (фиктивная) переменная в этом уравнении?
8. Чему равен 95%-ный доверительный интервал? Что это означает?
9. По каким причинам возникают искажения при применении ковариационного анализа (ANCOVA) к плану неэквивалентных групп?
10. Какой процедурой является регрессионный анализ? Поясните.
11. Каков критерий проведения линии через облако из точек?
12. Какие значения являются положительными (отрицательными)?
13. Что такое остаточные разности?
14. Что влияет на наклон линий регрессии, когда ошибка измерения отсутствует?



15. О чем говорит вертикальное смещение двух линий?
16. Влияет ли на наклон линий регрессии ошибка измерения, допущенная после эксперимента? Почему?
17. Как отражается на наклоне линий регрессии ошибка измерения, допущенная до эксперимента?
18. Что происходило бы с линиями, если бы ошибки измерения не было?
19. Как называется такой случай?
20. С каким измерением соотносится регрессионный анализ?
21. Как оно влияет на линию регрессии?
22. Где уменьшается наклон линии?
23. В результате чего возникает псевдоэффект?
24. Почему такой псевдоэффект не возникает при рандомизированном плане ковариационного анализа?
25. Почему возникает сглаживание линий?
26. К чему приводит сглаживание линий?
27. Происходит ли сглаживание линий и искажение данных в рандомизированном случае? Почему?
28. Какими причинами вызывается искажение при применении модели ковариационного анализа (ANCOVA) к плану исследования неэквивалентных групп?
29. Верно ли, что чем больше неэквивалентность между группами, тем сильнее окажется искажение? Объясните, почему.
30. Каков возможный вариант решения проблемы, возникающей при плане исследования неэквивалентных групп?
31. Как устранить сглаживание линий регрессии?
32. С чем соотносится надежность измерения?
33. Если половина измерений истинна, а другая половина измерений ошибочна, тогда чему равна надежность?
34. Можем ли мы увидеть ошибку измерения непосредственно из наших данных? Почему?
35. Что представляет собой надежность?
36. Какова формула для коррекции ошибки измерения?
37. В чем заключается идея данной формулы?
38. На какую величину требуется корректировать значения по каждому испытуемому?
39. К чему должны приблизиться баллы каждого испытуемого?
40. Если ошибка измерения до эксперимента отсутствует, то чему равна надежность?
41. Как в этом случае проводится коррекция данных измерения?
42. Какой вид имеет корректировка надежности, равной 0,8?

43. Что означает «сжатие» распределения на величину, пропорциональную ошибке измерения?
44. Какой вид имеет после корректировки формула ковариационного анализа (ANCOVA)?
45. В чем её отличие от оригинальной формулы ковариационного анализа (ANCOVA)?
46. Какая модель ковариационного анализа (ANCOVA) применяется в статистическом анализе по плану неэквивалентных групп с целью исправления смещения, вызываемого ошибкой измерения до эксперимента?
47. При помощи чего оценивается верхняя граница надежности?
48. При помощи чего оценивается нижняя граница надежности?
49. Когда можно быть уверенным в том, что действительно удалось обнаружить значительный эффект на основании данных, которые не содержат ошибки измерения?

### Тестовые задания к теме 16

1. Средние различия между экспериментальной группой и контрольной группой отражают тот коэффициент, который связан с переменной
  - а) переменной  $X$
  - б) переменной  $Z$
  - в) переменной  $Y$
  - г) переменной  $e$
  - д) переменной  $\beta$
2. Для того, чтобы определить, есть ли статистически значимое различие между полученным значением и истинным значением, нужно задать
  - а) ширину и высоту
  - б) угол наклона
  - в) стандартное отклонение
  - г) стандартную ошибку
  - д) доверительный интервал
3. 95%-ный доверительный интервал равен коэффициенту плюс/минус
  - а) одно значение стандартной ошибки
  - б) два значения стандартной ошибки
  - в) три значения стандартной ошибки
  - г) девятю пять значений стандартной ошибки
  - д) среднее значение стандартной ошибки
4. Если бы не было ошибки измерения вообще, то баллы нашего испытуемого попали бы прямо

- а) на линию регрессии
  - б) на стену
  - в) на потолок
  - г) в книгу рекордов Гиннеса
  - д) под кривую нормального распределения
5. Регрессионный анализ – это процедура
- а) с наименьшим количеством треугольников
  - б) с наименьшим количеством квадратов
  - в) с наименьшим количеством кружков
  - г) с наименьшим количеством крестиков
  - д) с наименьшим количеством галочек
6. Вертикальное расстояние от линии регрессии до каждой точки называется
- а) остатком от деления
  - б) остатком от вычитания
  - в) сухим остатком
  - г) врожденными задатками
  - д) атмосферными осадками
7. Регрессионный анализ предполагает проведение линии через те данные, которые образуют наименьшую сумму квадратов
- а) остатков от деления
  - б) остатков от вычитания
  - в) сухих остатков
  - г) врожденных задатков
  - д) атмосферных осадков
8. Там, где ошибка измерения отсутствует, наклон линий регрессии влияет
- а) наблюдаемый паттерн
  - б) теоретический паттерн
  - в) психологический паттерн
  - г) ничто не влияет
  - д) затрудняюсь ответить
9. О каком-либо эффекте можно говорить лишь в том случае, если между двумя линиями наблюдается
- а) горизонтальное смещение
  - б) вертикальное смещение
  - в) прямой угол
  - г) выравнивание
  - д) параллакс
10. Повлияет ли на наклон линий регрессии ошибка измерения, которая допущена после эксперимента?
- а) разумеется, да
  - б) разумеется, нет
  - в) строгие правила отсутствуют
  - г) в принципе неважно
  - д) затрудняюсь ответить

11. При регрессионном анализе линии регрессии проводятся с учетом смещения точек
- а) влево
  - б) вправо
  - в) по вертикали
  - г) по горизонтали
  - д) по часовой стрелке
12. Если ошибки рандомизированы, то мы получим столько остатков разностей, смещающих наклон линии одновременно и вверх и вниз, что в среднем наклон линии
- а) сильно увеличится
  - б) сильно уменьшится
  - в) несильно увеличится
  - г) несильно уменьшится
  - д) не изменится
13. Когда ошибка измерения допущена до эксперимента, то ошибки добавляются
- а) влево
  - б) вправо
  - в) по вертикали
  - г) по горизонтали
  - д) по часовой стрелке
14. Регрессионный анализ предполагает проведение линий в соответствии
- а) с вертикальным смещением
  - б) с горизонтальным смещением
  - в) с смещением влево
  - г) с смещением вправо
  - д) с смещением по часовой стрелке
15. Если бы ошибки измерения не было, то линии регрессии
- а) были бы перпендикулярны друг другу
  - б) были бы параллельны друг другу
  - в) вращались бы относительно друг друга
  - г) накладывались бы друг на друга
  - д) исчезали бы и появлялись друг за другом
16. Когда мы прибавляем ошибку измерения, допущенную до эксперимента, то мы фактически удлиняем
- а) вертикальное измерение
  - б) горизонтальное измерение
  - в) четвертое измерение
  - г) порядковое измерение
  - д) интервальное измерение
17. В нулевом случае линии
- а) перпендикулярны друг другу
  - б) параллельны друг другу
  - в) вращаются относительно друг друга

- г) накладываются друг на друга  
д) исчезают и появляются друг за другом
18. Чем больше неэквивалентность между группами, тем искажение окажется
- а) невыносимее
  - б) непоправимее
  - в) слабее
  - г) сильнее
  - д) не изменяется
19. Если ошибка в измерении до эксперимента отсутствует, тогда надежность измерения равна
- а) минус 1
  - б) ноль
  - в) плюс 1
  - г) 3,14
  - д) 95%
20. Если в измерениях до эксперимента содержатся только исключительно ошибки, тогда надежность измерения равна
- а) минус 1
  - б) ноль
  - в) плюс 1
  - г) 3,14
  - д) 95%
21. Необходимым условием искажения при исследовании по плану неэквивалентных групп оказывается
- а) точность измерения
  - б) надежность измерения
  - в) валидность измерения
  - г) тщательность измерения
  - д) ошибка измерения
22. Поскольку надежность измерения представляет собой величину той доли, в которой измеренные значения являются истинными по отношению к ошибочным значениям, она должна подсказать нам, насколько нужно
- а) прижимать
  - б) выжимать
  - в) отжимать
  - г) сжимать
  - д) разжимать
23. Новые значения должны быть «скорректированы» с поправкой
- а) на надежность измерения
  - б) на внимательность измерения
  - в) на уместность измерения
  - г) на строгость измерения
  - д) на ненадежность измерения

24. Если применить подобную корректировку ко всем баллам в группе, измеренным до эксперимента, тогда распределение доэкспериментальных значений окажется «сжатым» на величину, пропорциональную

- а) точности измерения
- б) надежности измерения
- в) валидности измерения
- г) тщательности измерения
- д) ошибке измерения

## ТЕМА 17. АНАЛИЗ ПЕРЕГИБА ЛИНИИ РЕГРЕССИИ

### Требования к анализу

С О Х О

С О О

В основном виде план перегиба линии регрессии представляет собой доэкспериментальную/послеэкспериментальную модель для двух групп, что видно из предлагаемой алгебраической записи данной модели. Подобно другим версиям структуры данного плана (например, ковариационный анализ рандомизированного эксперимента, план неэквивалентных групп), нам будет необходима статистическая модель, включающая символы, обозначающие значения до эксперимента, значения после эксперимента и индикаторную (фиктивную) переменную для эксперимента.

### Допущения при анализе

Прежде, чем приступить к рассмотрению конкретной статистической модели, важно понять те допущения, с которыми придется иметь дело. Допускается, что мы рассматриваем общий вид плана перегиба линии регрессии. Отступления от общего вида будут представлены позже. Для того, чтобы аналитическая модель считалась адекватной, следует принять пять следующих допущений, каждое из которых будет поочередно рассмотрено:

1. **Отсекающий критерий.** Отсекающий критерий должен соблюдаться без исключений. Если возникает искажение, имеющее отношение к отсекающему значению (кроме тех случаев, когда достоверно известно, что это случайность), тогда вся выборка оказывается под угрозой, и высока вероятность того, что оценка эффекта окажется искаженной. Искажения, имеющие отношение к отсекающему значению, часто называются «пушистым» планом перегиба линии регрессии. «Пушистые» планы создают затруднения, которые у нас нет возможности рассматривать здесь.

2. **Распределение до и после эксперимента.** Допускается, что распределение до и после эксперимента может быть описано полиномиальной функцией (многочлен). Если истинная доэкспериментальная взаимосвязь оказывается логарифмической, экспоненциальной либо иной функцией, то рассматриваемая нами модель оказывается непригодной, а оценка экспериментального эффекта становится искаженной. Разумеется, если данные могут быть до начала анализа преобразованы таким образом, чтобы сформировать полиномиальное распределение, тогда рассматриваемая нами модель становится адекватной, однако, могут возникать затруднения при интерпретации результатов анализа. Иногда возможно, что, пусть истинная взаимосвязь не полиномиальная, однако, полином (многочлен) более высокого порядка способен адекватным образом объяснить существующую функцию. Однако, мало вероятно, что аналитик опознает именно такой случай.
3. **Дисперсия доэкспериментальных значений в контрольной группе.** В контрольной группе должно быть достаточное множество значений, позволяющее выполнить адекватную оценку истинной взаимосвязи (например, линия доэкспериментальной/послеэкспериментальной регрессии) для данной группы. Обычно бывает желательно располагать также вариабельностью значений в экспериментальной группе, хотя строгого требования на этот счет не существует, потому что линия регрессии для контрольной группы может быть приведена к одной точке в экспериментальной группе.
4. **Непрерывность доэкспериментального распределения.** Для обеих групп (контрольной группы и экспериментальной группы) должно существовать единое непрерывное распределение значений, измеренных до эксперимента, разделение групп на этом распределении зависит от отсекающего критерия. В некоторых случаях возможно создать незатронутые группы (например, две группы пациентов из двух различных географических местностей), существенно различающихся по какому-либо измеренному параметру настолько, что критерий отсекания оказывается очевидным. К таким группам с естественным перегибом следует подходить с осторожностью, потому что велика вероятность того, что, раз эти группы естественным образом различаются по критерию отсекания ещё до проведения эксперимента, то в подобном различии может отражаться искажение, допущенное при формировании выборки, образующее естественные доэкспериментальные/послеэкспериментальные перегибы в данной точке.
5. **Реализация эксперимента.** Допускается, что эксперимент проводится единообразно в отношении всех испытуемых, т. е. каждый испытуемый получает одинаковую дозировку, подвергается одинаковой продолжительности пребывания, усваивает одинаковый объем информации и т. п. Если это не так, то отступления от эксперимента следует отражать в модели анализа, что осложняет проведение анализа.

## Проблема криволинейности

Основная проблема, связанная с анализом данных по плану перегиба линии регрессии, заключается в неадекватности модели по причине её неверной конкретизации. Если конкретизация модели выполнена неверно, тогда оценки экспериментального эффекта, скорее всего, окажутся искаженными. Для того, чтобы понять эту мысль, для начала рассмотрим то, что происходит, если данные (например, двумерная доэкспериментальная/послеэкспериментальная взаимосвязь) криволинейны, а мы применяем к таким данным прямолинейную модель.

**Если истинная доэкспериментальная взаимосвязь не линейная...**

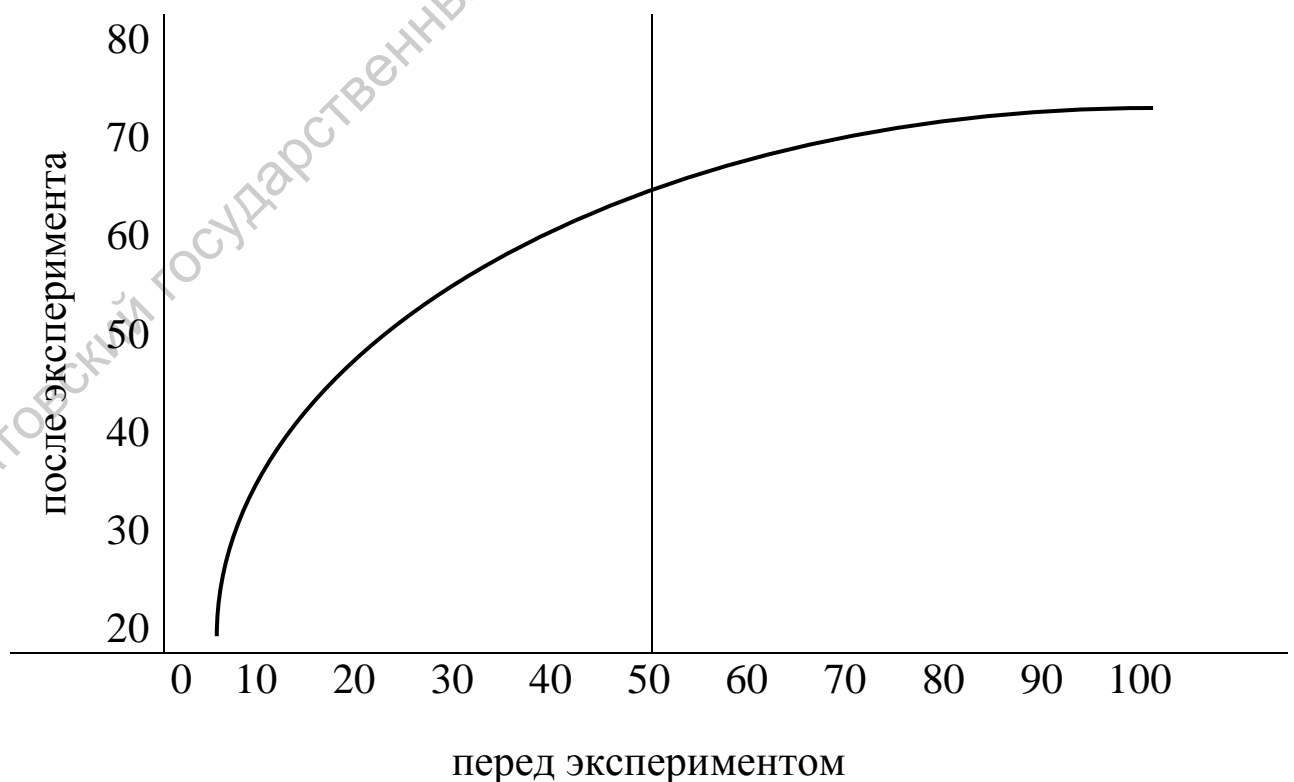


Рис. 1. Криволинейная взаимосвязь



На рис. 1. представлена простая криволинейная взаимозависимость. Если кривая линия на рис. 1 описывает взаимозависимость в ситуации до эксперимента, то нам потребуется принять во внимание данное обстоятельство применительно к нашей собственной статистической модели. Обратите внимание на то, что, несмотря на то, что значение отсекания на рисунке равно 50, в этой точке отсутствует прыжок или перегиб. Это говорит нам об отсутствии экспериментального эффекта.

**а мы применяем прямолинейную модель ...**

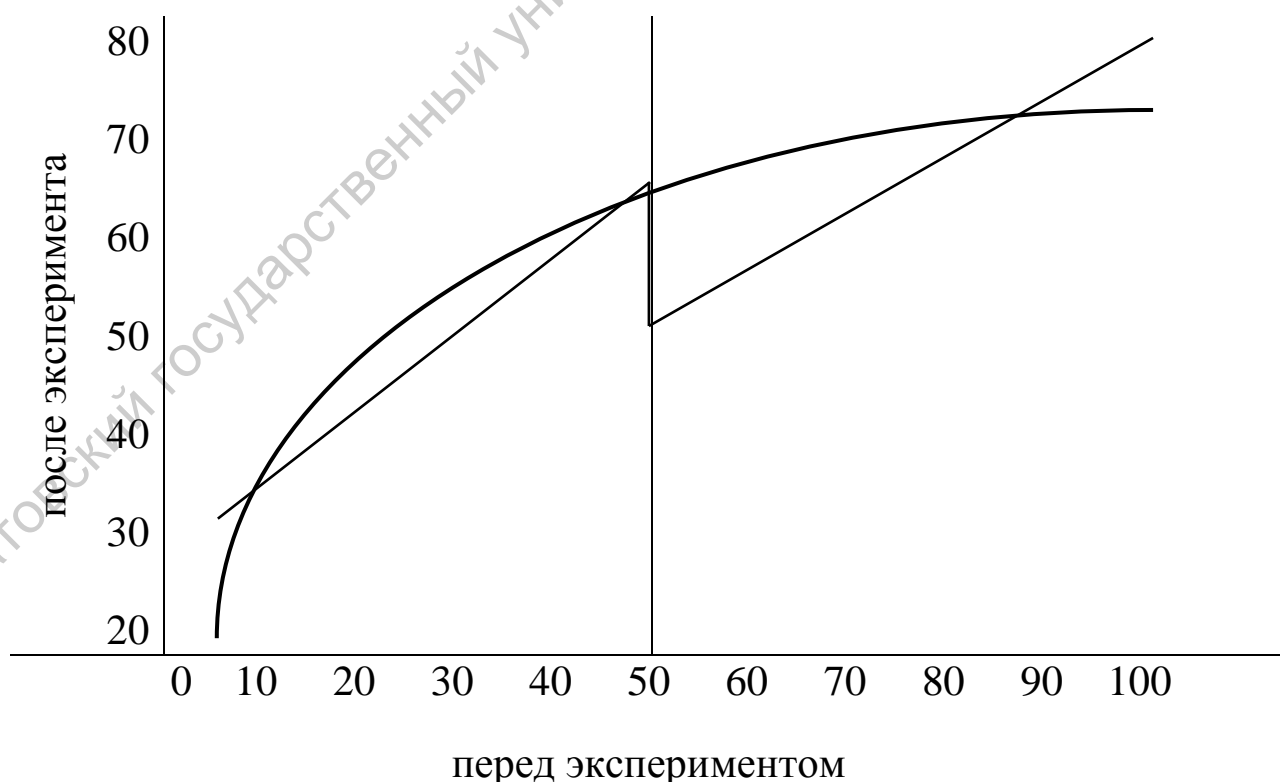


Рис. 2. Описание криволинейной взаимосвязи с помощью прямолинейной модели

Теперь взглянем на рисунок 2. На рисунке 2 изображено то, что происходит, когда мы применяем прямолинейную модель к криволинейной взаимосвязи, показанной на рисунке 1. Наклон обеих прямых линий одинаков, т. е. отсутствует взаимодействие между экспериментом и доэкспериментальными значениями. Вы видите, что прямолинейная модель допускает прыжок в точке отсекания, даже несмотря на то, что в истинной функции отсутствует перегиб.

**и даже если линии не параллельны (эффект взаимодействия) ...**

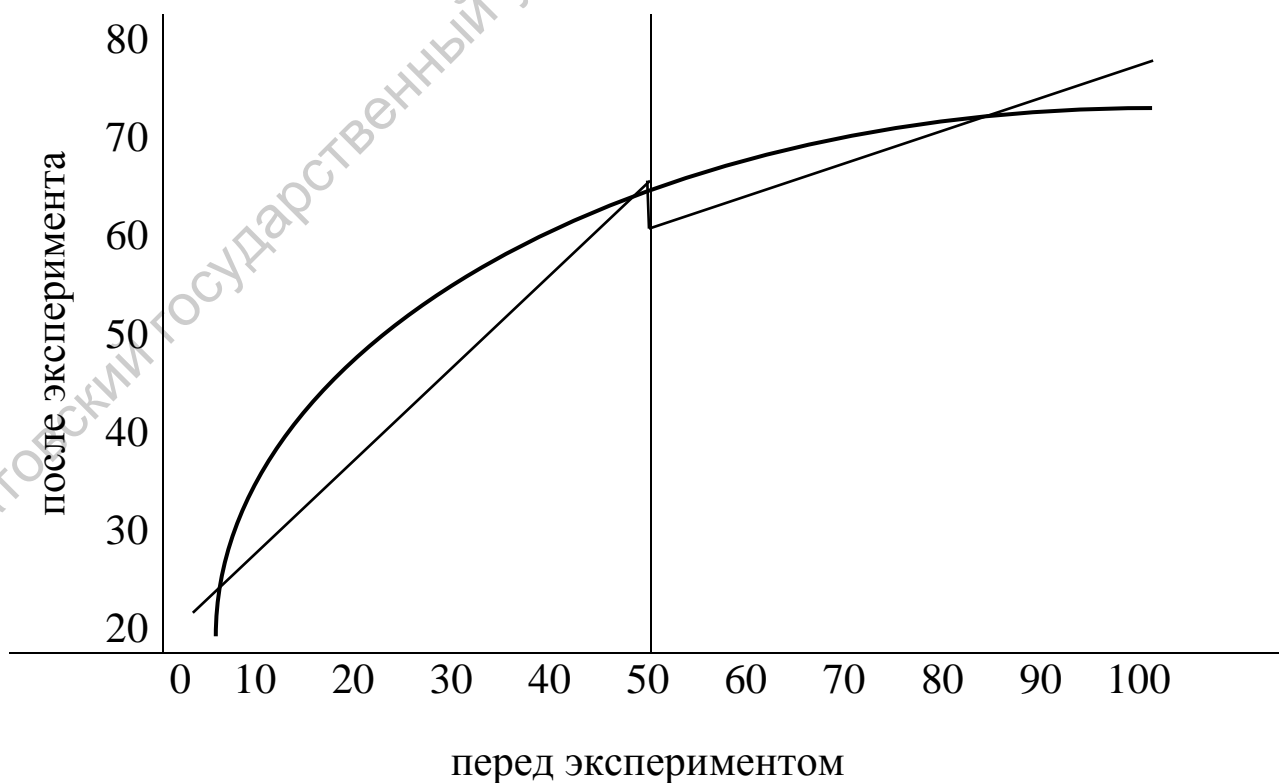


Рис. 3. Описание криволинейной взаимосвязи с помощью прямых линий с неодинаковым наклоном

(эффект взаимодействия)

Даже если прямые линии различаются своим наклоном, наша проблема не решается. На рисунке 3 показано то, что происходит в таком случае. Несмотря на то, что псевдоэффект в данном случае меньше, чем при одинаковом наклоне, псевдоэффект всё равно существует.

Вывод из этого простой. Если истинная модель криволинейная, а мы применяем к ней прямолинейную модель, то есть вероятность прийти к ошибочному заключению о наличии эффекта (различий до эксперимента и после эксперимента), когда эффекта нет. Это частный случай более широкой проблемы конкретизации общей линейной модели.

### Конкретизация модели

Для понимания проблемы конкретизации общей линейной модели и того, каким образом эта проблема соотносится с анализом перегиба линии регрессии, необходимо выделить три типа конкретизации.

Первый тип конкретизации общей линейной модели – **точная конкретизация** истинной модели. Что значит «точная конкретизация»?

Первым уравнением описывается «истинность» данных. Уравнением описывается простая линейная взаимосвязь до и после эксперимента с экспериментальным эффектом. Обратите внимание на то, что в уравнении представлены данные до эксперимента (переменная X), данные после эксперимента (переменная Y) и сам эксперимент (индикаторная (фиктивная) переменная Z).

Во втором уравнении представлена та модель, которую мы конкретизируем для нашего анализа. Во втором уравнении также имеются данные до эксперимента (переменная X), данные после эксперимента (переменная Y) и сам эксперимент (индикаторная (фиктивная) переменная Z). Это все, что есть в уравнении – в модели, которую мы конкретизируем, отсутствуют ненужные члены.

*Когда мы точно конкретизируем истинную модель, мы получаем неискаженную и действительную оценку экспериментального эффекта.*

Если истинная функция:

$$y_i = B_0 + B_1 X_i + B_2 Z_i$$

а мы применяем:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i$$

**Наша модель *точным* образом конкретизирована, и мы получаем неискаженную и действительную оценку**

Второй тип конкретизации общей линейной модели – **избыточная конкретизация** истинной модели, при которой в аналитическую модель включаются лишние и ненужные члены. Поскольку мы включили в уравнение все необходимые члены, то наша оценка экспериментального эффекта не будет искажена. Однако, нам придется заплатить за включение в наш анализ ненужных нам членов – наша оценка экспериментального эффекта не будет действительной. Что это означает? Это означает увеличение вероятности того, что мы придем к выводу об отсутствии эффекта, хотя на самом деле эффект будет существовать. Включение в анализ ненужных членов приравнивается к усилению шума в данных – шум затрудняет нахождение эффекта, даже когда эффект присутствует.

С одной стороны, если истинная функция:

$$y_i = B_0 + B_1 X_i + B_2 Z_i$$

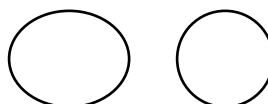
а мы применяем:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + (\beta_2 X_i Z_i) + e_i$$

**Наша модель *избыточным* образом конкретизирована, и мы получаем *недействительную* оценку**

Третий тип конкретизации общей линейной модели – **недостаточная конкретизация** истинной модели, при которой в аналитическую модель включаются не все нужные члены. В этом случае реальность оказывается богаче и многообразнее нашей модели. Оценка экспериментального эффекта будет и искаженной, и недействительной.

Наконец, если истинная функция:



$$y_i = B_0 + B_1 X_i + B_2 Z_i + B_2 X_i Z_i + B_2 Z_i^2$$

а мы применяем:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i$$

**Наша модель *недостаточным* образом конкретизирована, и мы получаем *искаженную* оценку**

### Анализ стратегии

После того, как мы рассмотрели варианты неправильной конкретизации общей линейной модели, мы можем разработать стратегию моделирования, спланированную таким образом, чтобы, во-первых, уберечься от искаженных оценок и, во-вторых, обеспечить максимальную действительность оценок. Наилучшим вариантом, без сомнения, стало бы конкретизировать истинную модель точным образом. Однако, на практике это бывает труднодостижимо, поскольку истинная модель зачастую омрачается ошибками в наших данных. Если мы вынуждены совершить ошибку – если мы вынуждены конкретизировать модель неверным образом – то обычно предпочитается избыточная конкретизация истинной модели перед недостаточной конкретизацией. Избыточная конкретизация гарантирует нам, что мы включили в уравнение все необходимые члены, пусть даже ценой включения ненужных членов. Мы получим неискаженную оценку эффекта, пусть даже она окажется недействительной. Недостаточной конкретизации модели чаще всего стараются избегать, потому что она чревата и искаженными, и недействительными оценками.

После того, как сделан выбор относительно варианта конкретизации модели, общая стратегия анализа начинается с такой конкретизации модели, относительно которой мы совершенно уверены – она избыточная. Оценка экспериментального эффекта в этой модели характеризуется вероятностью быть неискаженной, хотя она и будет недействительной. Далее, в процессе проведения анализа, постепенно производят сокращение членов высокого порядка до тех пор, пока не покажется, что оценка экспериментального эффекта отличается от состояния до эксперимента либо пока диагностика модели (например, графики, на которых остатки вычитания представлены в виде точек) не укажет на плохую адекватность модели.

### Этапы анализа

В своей основе анализ перегиба линии регрессии состоит из пяти этапов:

## 1. Преобразование доэкспериментальных данных.

$$\tilde{X}_i = X_i - X_c$$

Анализ начинается с вычитания отсекающего значения из каждого доэкспериментального значения. Это производится с целью приравнять коэффициент отрезка, отсекаемого на координатной оси, к значению отсекающего критерия. Какая от этого польза? Если вычесть значение отсекания из каждого доэкспериментального значения, тогда те доэкспериментальные значения, которые изначально были отсекающими значениями, окажутся равными нулю. Поскольку отсекаемый отрезок по определению является значением переменной  $y$ , если  $x = 0$ , то выполненная нами математическая операция приводит  $x$  к нулю, тем самым отсекающий критерий становится началом отрезка, отсекаемого на оси  $y$ .

## 2. Визуальное рассмотрение взаимосвязи.

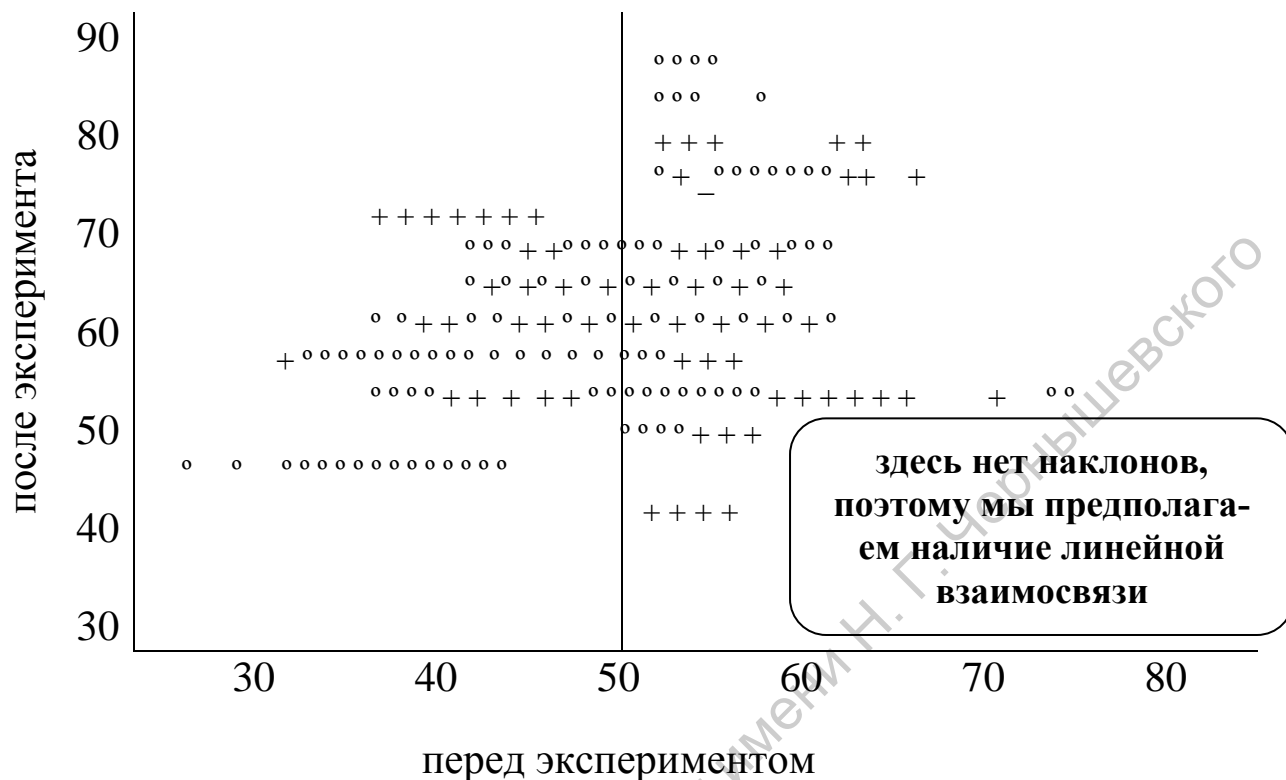
При рассмотрении графика доэкспериментальной взаимосвязи следует обращать внимание на два главных момента. Во-первых, важно определить, имеется ли какой-либо зрительно различимый перегиб взаимосвязи в точке отсекания. Перегибом может считаться изменение вертикального уровня (основной эффект), изменение наклона (эффект взаимодействия), либо и то, и другое. Если зрительно замечен перегиб в точке отсекания, тогда результаты анализа, свидетельствующие об отсутствии экспериментального эффекта, следует признать неудовлетворительными. Однако, если зрительно перегиб не замечен, то возможно, что вариабельность данных маскирует эффект, поэтому к результатам анализа следует подходить тщательно.

Второй момент, на который следует обращать внимание при рассмотрении двумерных графиков взаимосвязи, это степень многочлена, которая может оказаться необходимой в связи с двумерным наклоном распределения, в особенности в контрольной группе. Хорошим способом для этого является подсчет точек изгиба распределения (т. е. сколько раз распределение «изгибается»). Если взаимосвязь линейная, тогда точек изгиба нет. Единственный изгиб укажет на квадратичный многочлен или многочлен второго порядка. Эта информация понадобится при определении той модели, которую будут подвергать конкретизации.

## 3. Конкретизация членов высшего порядка и взаимодействий

С учетом точек изгиба, определенных на втором этапе анализа, на третьем этапе производится преобразование значений переменной  $X$ . Эмпирически установлено, что следует выбирать многочлен на два порядка превышающий количество точек изгиба. Например, если двумерная взаимосвязь оказалась линейной (т. е. точки изгиба отсутствуют), тогда придется подвергнуть преобразованию многочлен второго порядка ( $0 + 2$ ). Например, на следующем графике отсутствуют точки изгиба:

**Подсчитывают количество точек изгиба в обеих группах...**



Многочлен первого порядка уже существует в модели ( $X$ ), поэтому все, что необходимо сделать для создания многочлена второго порядка – возвести  $X$  во вторую степень, чтобы получить  $X^2$ . При каждом преобразовании  $X$  также создается член взаимодействия путем умножения на  $Z$ . В нашем примере появляются два члена взаимодействия:  $X_i Z_i$  и  $X_i^2 Z_i$ . Любые преобразования выполняются довольно просто на компьютере. Если на графике двумерного распределения обнаружены две точки изгиба, тогда следует выполнить трансформацию четвертой степени ( $2 + 2$ ) и создать для них члены взаимодействия.

Визуальный осмотр графика взаимосвязи – это не единственная опора при определении необходимой степени многочлена. Разумеется, следует принимать во внимание прежний опыт моделирования аналогичных данных. Приведенное здесь эмпирическое правило подразумевает, что лучше ошибаться в направлении избыточной оценки порядка истинного многочлена, по тем же причинам, по которым лучше произвести избыточную конкретизацию истинной модели. Какой бы порядок многочлена не был установлен при визуальном рассмотрении графика двумерного распределения, все преобразования и их взаимодействия следует выполнить в соответствии с этим порядком. Если выбран четвертый порядок, то все члены  $X$  следует привести к  $X^4$  и соответственно включить в модель их взаимодействия.

#### 4. Оценка исходной модели

На этом этапе всё готово к тому, чтобы приступить собственно к анализу. Для этого подойдет любая компьютерная программа множественной регрессии. Следует всего лишь выполнить регрессию послеэксперименталь-

ных значений, или  $Y$ , по известным доэкспериментальным значениям  $X$ , экспериментальной переменной  $Z$  и учесть все трансформации высшего порядка и их взаимодействия, выведенные на третьем этапе. Коэффициент регрессии, связанный с членом  $Z$  (т. е. переменная принадлежности к экспериментальной группе либо к контрольной группе) – это оценка основного экспериментального эффекта. Если в точке отрезания имеется вертикальный перегиб, то он будет выражаться этим коэффициентом. Значимость этого (или любого другого коэффициента) проверяется при помощи  $t$ -критерия с применением стандартной ошибки данного коэффициента, которую рассчитывает компьютерная программа.

Исходная модель для случая, в котором отсутствуют точки изгибов (конкретизация полной квадратичной модели):

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \beta_4 X_i^2 + \beta_5 X_i^2 Z_i + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена
- $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси
- $\beta_1$  = линейный преэкспериментальный коэффициент
- $\beta_2$  = среднее различие в эксперименте
- $\beta_3$  = линейное взаимодействие
- $\beta_4$  = квадратичный преэкспериментальный коэффициент
- $\beta_5$  = квадратичное взаимодействие
- $X_i$  = преобразованные доэкспериментальные данные
- $Z_i$  = индикаторная (фиктивная) переменная для эксперимента (0 = контрольная группа; 1 = экспериментальная группа)
- $e_i$  = остаток для  $i$ -того члена после вычитания

Если при выполнении третьего этапа анализа избыточная оценка функции многочлена была выполнена корректно, в соответствии с моделью распределения данных, тогда оценка экспериментального эффекта окажется, по крайней мере, неискаженной. Однако, по причине того, что в истинную модель оказались включены те члены, которые могут не быть необходимыми, имеется вероятность того, что оценка окажется недействительной, то есть члены, выражающие стандартную ошибку, окажутся менее значимыми, соответственно, значимость экспериментального эффекта станет ниже. Тем не менее, если на данном этапе анализа коэффициент сильно значим, то резонно было бы заключить, что экспериментальный эффект существует. Направление эффекта интерпретируется с учетом знака коэффициента и направления послеэкспериментальной шкалы. Например, значительный коэффициент регрессии для члена  $XZ$  предполагает линейное взаимодействие.

## 5. Усовершенствование модели



На основе результатов четвертого этапа анализа, можно попробовать удалить очевидно ненужные члены и более точным образом оценить величину экспериментального эффекта. Это хитроумная процедура, выполнять которую следует осторожно, если мы желаем свести вероятность искажения к минимуму. Для её выполнения, следует исследовать результат регрессионного анализа, полученный на четвертом этапе, обращая внимание на степень общего соответствия модели анализируемым данным, на присутствие незначимых коэффициентов и на характер остаточных значений от вычитания. Консервативный способ определения того, каким образом следует усовершенствовать модель, заключается в том, что сначала следует исследовать члены высшего порядка в имеющейся модели и их взаимодействие. Если оба коэффициента незначимы, но показатели соответствия модели исследуемым данным и характер остаточных значений от вычитания указывают на хорошее соответствие, то эти два члена можно сократить, а то, что останется – вычислить заново. Например, если был рассчитан многочлен четвертого порядка, и если было установлено, что коэффициенты  $X^4$  и  $X^4Z$  оказались незначимыми, то эти члены можно отбросить, и тогда останется уточненная модель третьего порядка. Эта процедура может продолжаться до тех пор, пока

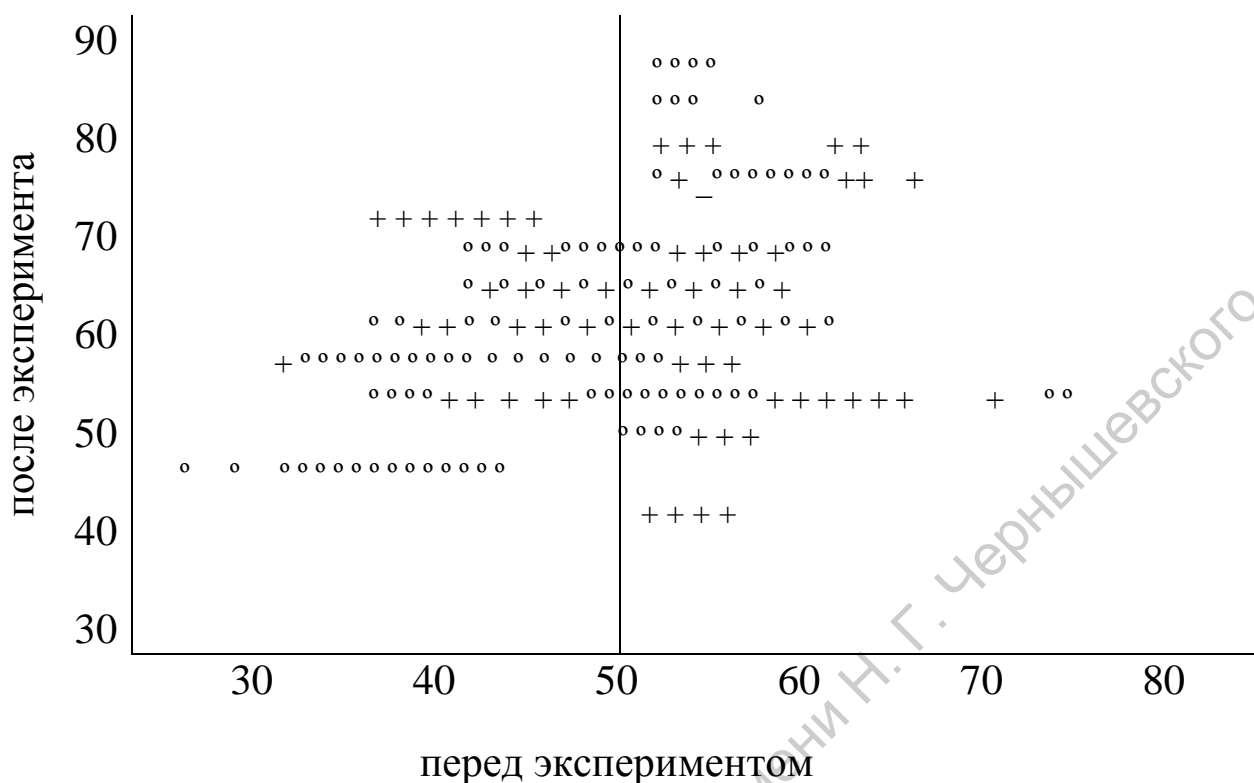
- а) один из коэффициентов не окажется значимым;
- б) показатель адекватности модели значительно не уменьшится;
- в) характер остатков от вычитания не укажет на плохое соответствие модели анализируемым данным.

Окончательный вид модели может по-прежнему включать некоторые избыточные члены, однако, имеется вероятность их уменьшения, соответственно, эффективность повысится.

Процедуры конкретизации модели, предполагающие отбрасывание отдельных членов на отдельных этапах анализа, являются более опасными и характеризуются большей вероятностью приводить к искаженным оценкам по причине значительной мультиколлинеарности между членами модели.

### **Пример анализа**

Анализ по плану перегиба линии регрессии легче будет понять, если рассмотреть пример. Данные для нашего примера представлены на рисунке:



При рассмотрении данного рисунка очевидны несколько вещей.

Во-первых, существует потрясающий экспериментальный эффект. Экспериментальный эффект на рисунке составляет 10 баллов.

Во-вторых, обе группы прекрасно описываются прямыми линиями – точки наклона не очевидны. Тем самым, в качестве исходной модели может быть применена полная квадратичная модель, уравнение которой приведено выше.

Результаты исходной конкретизации представлены в следующей таблице:

Уравнение регрессии имеет вид  
 постэкспериментальный эффект = 49,1 + 0,972\*отсекание +  
 10,2\*группа - 0,236\*линейный интервал - 0,00539\*квадратичный +  
 0,00276 квадратичный интервал

Предиктор (независимая переменная)	Коэффициент	Стандартное отклонение	t-отно- шение	p
Константа	49,1411	0,8964	54,82	0,000
отсекание	0,9716	0,1492	6,51	0,000
группа	10,231	1,248	8,20	0,000
линейный интер- вал	-0,2363	0,2162	-1,09	0,275
квадратичный	-0,005391	0,004994	-1,08	0,281
квадратичный интервал	0,002757	0,007475	0,37	0,712

s = 6,643

R - квадр. = 47,7%

R-квадр. (корр.)  
= 47,1%

Так выглядят результаты регрессии для полной квадратичной модели в нашем примере. Оценка экспериментального эффекта приведена следующей после «групповой» переменной. Исходная оценка составляет 10,231 (стандартная ошибка = 1,248) – очень близко к истинному значению в 10 баллов. Однако, нужно учесть, что имеются свидетельства того, что некоторые члены высшего порядка статистически незначимы и в модели не требуются. Говоря конкретно, член линейного взаимодействия, именуемый «линейным интервалом» (XZ), не является значимым.

Несмотря на то, что мы склонны (вероятно, оправданно) отбросить все три члена данной модели, однако, если мы будем следовать указаниям, приведенным в пятом этапе алгоритма анализа, то мы отбросим лишь два квадратичных члена, именуемых «квадратичным» и «квадратичным интервалом». Результаты уточненной модели представлены в следующей таблице:

Уравнение регрессии имеет вид

постэкспериментальный эффект = 49,8 + 0,824\*отсекание + 9,89\*группа - 0,0196\*линейный интервал

Предиктор (независимая переменная)	Коэффициент	Стандартное отклонение	t-отно- шение	p
Константа	49,7508	0,6957	71,52	0,000
отсекание	0,82371	0,05889	13,99	0,000
группа	9,8939	0,9528	10,38	0,000
линейный интервал	-0,01963	0,08284	-0,24	0,813

s = 6,639

R - квадр. = 47,5%

R-квадр. (корр.)  
= 47,2%

Так выглядят результаты регрессии для исходной модели без квадратичных членов. Мы видим, что в этой модели экспериментальный эффект уже равен 9,89 (стандартная ошибка = 0,95). Снова, мы получаем оценку, очень близкую к истинному экспериментальному эффекту, который составляет 10 баллов. Однако, обратим внимание на то, что стандартная ошибка здесь меньше, чем в первоначальной модели. Это то преимущество в эффективности, которое мы приобрели, когда отбросили два ненужных квадратичных члена. Мы также видим, что член линейного взаимодействия именуемый «линейным интервалом» также мало значим. Этот член был бы значимым, если бы различались линии наклона в двух группах. Рассмотрение графика показывает, что наклон одинаков, поэтому данный член действительно незначим.

Наконец, давайте отбросим незначимый член линейного взаимодействия и обратимся к уточненной модели:

Уравнение регрессии имеет вид

постэкспериментальный эффект = 49,8 + 0,814\*отсекание + 9,89\*группа

Предиктор (независимая)	Коэффициент	Стандартное отклонение	t-отно- шение	p
----------------------------	-------------	---------------------------	------------------	---

переменная)

Константа	49,8421	0,5786	86,14	0,000
отсекание	0,81379	0,04138	19,67	0,000
группа	9,8875	0,9515	10,39	0,000

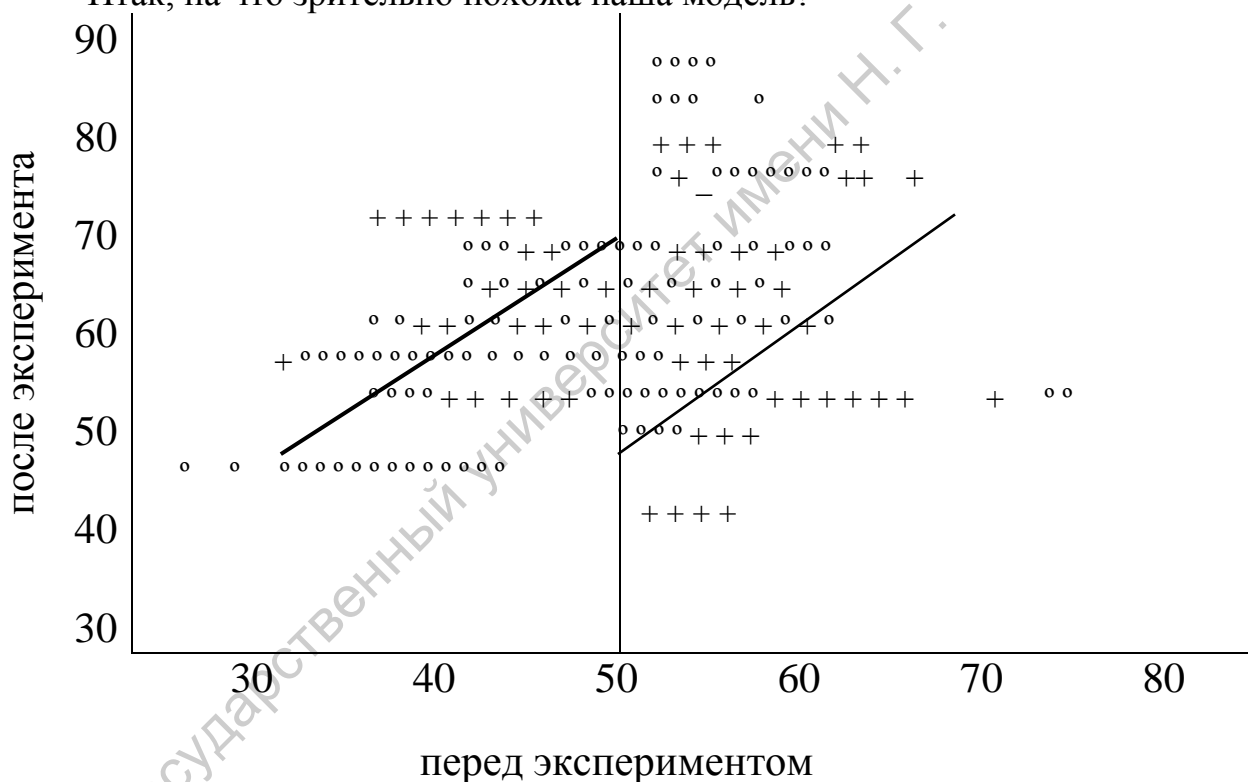
$s = 6,633$

R - квадр. = 47,5%

R-квадр. (корр.)  
= 47,3%

Так выглядят результаты регрессии в окончательной модели. Здесь мы видим, что экспериментальный эффект и стандартная ошибка почти тождественны предыдущей модели, и что оценка экспериментального эффекта почти неискаженно равна истинному эффекту, составляющему 10 баллов. Кроме того, видно, что все члены уравнения модели в окончательном виде статистически значимы, что дает основание полагать, что они необходимы для моделирования данных и их нельзя удалить.

Итак, на что зрительно похожа наша модель?



Ясно, что модель прекрасно соответствует данным, и статистически, и зрительно.

### Вопросы для самоконтроля по теме «Анализ перегиба линии регрессии»

1. Что представляет собой план перегиба линии регрессии в основном виде?
2. Что должна включать в себя статистическая модель структуры данного плана?
3. Что необходимо допустить прежде, чем приступить к рассмотрению конкретной статистической модели?

4. Какие допущения следует принять для того, чтобы аналитическая модель считалась адекватной? Назовите и охарактеризуйте эти допущения.
5. Какие планы называются «пушистыми»?
6. Какой функцией может быть описано распределение до и после эксперимента?
7. Какое распределение значений должно существовать для контрольной группы и экспериментальной группы?
8. Что означает, что эксперимент проводится единообразно в отношении всех испытуемых?
9. В чем заключается основная проблема, связанная с анализом данных по плану перегиба линии регрессии?
10. Что происходит в случае, если данные криволинейны, а мы применяем к таким данным прямолинейную модель?
11. Что значит «точная конкретизация» общей линейной модели?
12. Каким уравнением описывается точным образом конкретизированная модель?
13. Что значит «избыточная конкретизация» общей линейной модели?
14. Каким уравнением описывается избыточным образом конкретизированная модель?
15. Что значит «недостаточная конкретизация» общей линейной модели?
16. Каким уравнением описывается недостаточным образом конкретизированная модель?
17. Каким образом надлежит планировать стратегию моделирования?
18. В силу чего точная конкретизация истинной модели бывает труднодостижима?
19. Какой вид конкретизации более предпочтителен, если мы вынуждены конкретизировать модель неверным образом? Почему?
20. После того, как сделан выбор относительно варианта конкретизации модели, с чего начинается общая стратегия анализа?
21. До каких пор постепенно производят сокращение членов высокого порядка?
22. Из скольких этапов состоит анализ перегиба линии регрессии? Назовите и охарактеризуйте эти этапы.
23. С какой целью из каждого доэкспериментального значения вычитают отсекающие значения?
24. На что следует обращать внимание при рассмотрении графика доэкспериментальной взаимосвязи?
25. Что такое перегиб?

26. Как связана степень многочлена с количеством точек изгиба?
27. Что выражает коэффициент регрессии, связанный с членом  $Z$ ?
28. Какой алгебраический вид имеет конкретизация полной квадратичной модели? Охарактеризуйте каждый компонент этого выражения.
29. Что предполагает значительный коэффициент регрессии для члена  $XZ$ ?
30. В чём состоит консервативный способ определения того, каким образом следует усовершенствовать модель?
31. До каких пор следует продолжать процедуру усовершенствования модели?
32. В чем состоит опасность процедуры конкретизации модели, предполагающей отбрасывание отдельных членов на отдельных этапах анализа?

### Тестовые задания к теме 17

1. Отсекающий критерий должен соблюдаться
  - а) по мере возможности
  - б) по мере надобности
  - в) по желанию
  - г) от случая к случаю
  - д) без исключений
2. Искажения, имеющие отношение к отсекающему значению, часто называются
  - а) статистическими
  - б) противными
  - в) вредными
  - г) белыми
  - д) пушистыми
3. Распределение до и после эксперимента может быть описано при помощи
  - а) гистограммы
  - б) паттерна
  - в) колоколообразной кривой
  - г) многочлена
  - д) выборки
4. Отсутствие прыжка или перегиба кривой линии говорит нам о том, что экспериментальный эффект
  - а) велик
  - б) мал
  - в) не велик и не мал
  - г) отсутствует
  - д) затрудняюсь ответить
5. Включение в анализ ненужных членов приравнивается к усилению
  - а) сигнала

- б) шума
  - в) гама
  - г) кошмара
  - д) наблюдаемого паттерна
6. Если вычесть значение отсекаания из каждого доэкспериментального значения, тогда те доэкспериментальные значения, которые изначально были отсекающими значениями, окажутся равными
- а) минус 1
  - б) ноль
  - в) плюс 1
  - г) 3,14
  - д) 95%
7. Изменение вертикального уровня (основной эффект), изменение наклона (эффект взаимодействия), либо и то, и другое указывает на то, что экспериментальный эффект
- а) велик
  - б) мал
  - в) не велик и не мал
  - г) существует
  - д) затрудняюсь ответить
8. Основной экспериментальный эффект оценивается при помощи
- а) стандартной ошибки
  - б) стандартного отклонения
  - в) теоретического паттерна
  - г) коэффициента регрессии
  - д) коэффициента корреляции
9. Если в точке отрезания имеется вертикальный перегиб, то он будет выражаться
- а) стандартной ошибкой
  - б) стандартным отклонением
  - в) теоретическим паттерном
  - г) коэффициентом регрессии
  - д) коэффициентом корреляции
10. Если избыточная оценка функции многочлена была выполнена корректно, в соответствии с моделью распределения данных, тогда оценка экспериментального эффекта окажется
- а) существующей
  - б) несуществующей
  - в) искаженной
  - г) неискаженной
  - д) равной нулю
11. По причине того, что в истинную модель оказались включены те члены, которые могут не быть необходимы, имеется вероятность того, что оценка окажется
- а) равной нулю

- б) отличной от нуля
  - в) действительной
  - г) мнимой
  - д) удивительной
12. Направление эффекта интерпретируется с учетом знака
- а) стандартной ошибки
  - б) стандартного отклонения
  - в) теоретического паттерна
  - г) коэффициента регрессии
  - д) коэффициента корреляции
13. Если был рассчитан многочлен четвертого порядка, и если было установлено, что коэффициенты  $X^4$  и  $X^4Z$  оказались незначимыми, то эти члены
- а) можно перемножить
  - б) можно поделить
  - в) можно суммировать
  - г) можно вычесть
  - д) можно сократить

## ТЕМА 18. АНАЛИЗ ВЫТЕСНЕНИЯ ТОЧЕК РЕГРЕССИИ

### Статистические требования

$N_{(n=1)}$     0    X    0

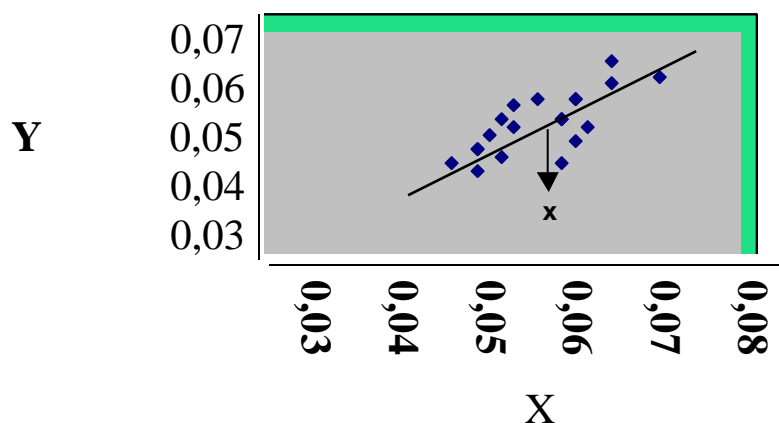
N        0        0

Алгебраическая запись плана анализа вытеснения точек регрессии показывает, что для статистического анализа требуются:

- ❖ доэкспериментальные значения
- ❖ послеэкспериментальные значения
- ❖ переменная, представляющая группы в эксперименте (где 0 = контрольной группе, а 1 = экспериментальной группе)

Это те же самые требования, которые предъявляет модель ковариационного анализа (ANCOVA). Единственное различие заключается в том, что при анализе по плану вытеснения точек регрессии, рассматриваются значения только одной из групп, участвующих в проведении эксперимента.





На рисунке показано двумерное (до и после эксперимента) распределение значений в соответствии с планом вытеснения точек регрессии для образовательной программы, посвященной СПИДу. Новая образовательная программа, посвященная СПИДу, действует в одном из районов губернии, остальные районы могут рассматриваться в качестве контрольных групп. Значения переменной  $x$  соответствуют количеству лиц с позитивной ВИЧ-реакцией на 1000 человек в течение года, предшествовавшего началу проведения учебной программы по СПИДу, значения переменной  $y$  соответствуют этому же показателю в течение года после начала данной учебной программы. Наша цель состоит в том, чтоб оценить величину вертикального вытеснения в экспериментальной группе по отношению к линии регрессии, которая проведена для всех контрольных групп. Вытеснение показано на графике стрелкой. Мы будем применять модель ковариационного анализа (ANCOVA) в следующем виде модели регрессии:

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i$$

где

- $y_i$  = полученное значение для  $i$ -того члена
- $\beta_0$  = коэффициент для отрезка, отсекаемого на координатной оси
- $\beta_1$  = преэксperimentalный коэффициент
- $\beta_2$  = среднее различие в эксперименте
- $X_i$  = ковариация
- $Z_i$  = индикаторная (фиктивная) переменная для эксперимента (0 = контрольная группа; 1 = экспериментальная группа)
- $e_i$  = остаток для  $i$ -того члена после вычитания

Когда мы конкретизировали данную модель применительно к имеющимся у нас данным, то получили следующую таблицу регрессии:

Уравнение регрессии имеет вид  
 $Y = 0,0120 + 0,784 X - 0,0199 Z$

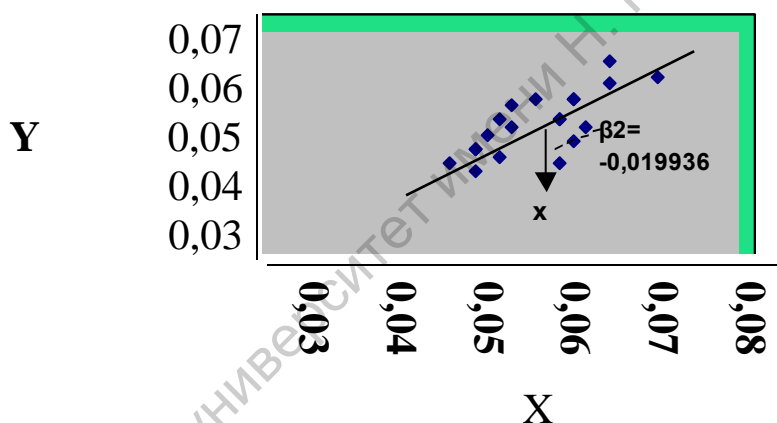
Предиктор (независимая переменная)	Коэффициент	Стандартное отклонение	t-отно- шение	p
Константа	0,011956	0,004965	2,41	0,023
X	0,78365	0,09864	7,94	0,000
Y	-0,019936	0,005800	-3,44	0,002

$s = 0,005689$

R - квадр. = 72,6%

R-квадр. (корр.)  
= 70,6%

Коэффициент, ассоциированный с дихотомической экспериментальной переменной (Z), выражает оценку вертикального вытеснения по отношению к линии регрессии. В нашем примере, результат показывает, что за время действия учебной программы по СПИДу, уменьшение количества лиц с позитивной ВИЧ-реакцией выражается показателем 0,0199, и это значение оказывается статистически значимым. Это вытеснение отражено на следующем графике:



### Вопросы для самоконтроля

#### по теме «Анализ вытеснения точек регрессии»

1. Что необходимо для статистического анализа вытеснения точек регрессии?
2. К какой ещё модели анализа предъявляются те же самые требования?
3. В чём состоит различие моделей анализа?
4. В чём заключается цель анализа вытеснения точек регрессии?
5. Какой алгебраический вид принимает модель ковариационного анализа (ANCOVA)? Охарактеризуйте каждый компонент этого уравнения.
6. Что выражает коэффициент, ассоциированный с дихотомической экспериментальной переменной Z?

### Тестовые задания к теме 18

1. Оценку вертикального вытеснения по отношению к линии регрессии выражает коэффициент, ассоциированный с

- а) переменной  $x$   
б) переменной  $y$   
в) переменной  $z$   
г) индикаторной (фиктивной) переменной  
д) паттерном
2. В модели ковариационного анализа вытеснения точек регрессии переменная  $e_i$  выражает
- а) значение  $i$ -того члена  
б) преэкспериментальный коэффициент  
в) среднее различие в эксперименте  
г) индикаторную (фиктивную) переменную  
д) остаток для  $i$ -того члена после вычитания
3. В модели ковариационного анализа вытеснения точек регрессии переменная  $\beta_1$  выражает
- а) значение  $i$ -того члена  
б) преэкспериментальный коэффициент  
в) среднее различие в эксперименте  
г) индикаторную (фиктивную) переменную  
д) остаток для  $i$ -того члена после вычитания
4. В модели ковариационного анализа вытеснения точек регрессии переменная  $y_i$  выражает
- а) значение  $i$ -того члена  
б) преэкспериментальный коэффициент  
в) среднее различие в эксперименте  
г) индикаторную (фиктивную) переменную  
д) остаток для  $i$ -того члена после вычитания
5. В модели ковариационного анализа вытеснения точек регрессии переменная  $Z_i$  выражает
- а) значение  $i$ -того члена  
б) преэкспериментальный коэффициент  
в) среднее различие в эксперименте  
г) индикаторную (фиктивную) переменную  
д) остаток для  $i$ -того члена после вычитания

#### Семинар 4. Алгебраические модели психических процессов.

Вопросы для обсуждения:

1. Множества и отношения.
2. Отображения и функции.
3. Операции и алгебры.

Темы для рефератов:

1. Линейные пространства.
2. Метрические пространства.
3. Топологические пространства.
4. Композиции моделей.
5. Современное состояние измерения индивидуальных различий.

б. Измерение в связи с воспитанием учащихся.

Литература к семинару 6:

1. Сочивко Д. В., Якунин В. А. Математические модели в психолого-педагогических исследованиях. Л., 1988. С. 5 – 40
2. Битинас Б. Многомерный анализ в педагогике и педагогической психологии. Вильнюс, 1971. С. 44 – 67; 129 – 148

## МОДУЛЬ 5

### Контрольные задания по дисциплине «Математические методы в психологии»

#### Вариант I

1. Сформулируйте ответы на следующие вопросы:

1. Сформулируйте определения понятий «ответ», «статистика», «параметр».
2. Когда значение среднего будет являться параметром?
3. Как составляется простая случайная выборка?
4. Как называется комбинирование выборочных методов?
5. Как называется неоднородная выборка?
6. В чем заключается отличительная особенность гетерогенных выборок?
7. Что такое теоретический паттерн и наблюдаемый паттерн?
8. Что позволяет проверить критериальная валидность?
9. В чем состоит «реляционалистское» направление представления понятия конструктивной валидности?
10. Что необходимо показать для подтверждения конвергентной валидности?
11. Охарактеризуйте «социальные» угрозы валидности конструкторов.
12. Каковы основные преимущества и недостатки матрицы мультисвойств-мультиметодов?
13. В чем заключается задача концептуализации?
14. Что такое корреляция паттерна согласования?
15. Какое качество предполагает надежность?
16. Что такое систематическая ошибка?
17. В каком виде можно себе представить надежность?
18. Насколько велика оценка надежности?
19. Сколько общих классов оценок надежности существует?
20. Что необходимо для оценки надежности параллельных форм?
21. Как рассчитать среднюю корреляцию по общему баллу?
22. С чем соотносится уровень измерения? Что это означает?
23. Что такое интервальное измерение?
24. Как возникло шкалирование?
25. Приведите пример дихотомической и интервальной шкалы ответов.

26. Каким может быть трехмерный конструкт? Приведите пример.
27. Как называется шкалирование по Гуттману?
28. Что такое внутренняя валидность?
29. Какие причины угрожают внутренней валидности? Перечислите и охарактеризуйте.
30. Что понимается под «единственной группой»? Приведите пример.
31. Чем меньше корреляция двух переменных, тем сильнее регрессия к среднему? Почему?
32. Что такое искаженный отбор?
33. Что называется валидностью выводов?
34. В чем заключается проблема степени ошибки?
35. В чем состоит ошибка типа «пропуск цели»?
36. С конструированием каких двух взаимно исключающих гипотез связаны все статистические выводы? Как называются эти гипотезы?
37. Как связаны альфа-уровень и мощность?
38. Из чего складывается подготовка данных для анализа?
39. Назовите три основные характеристики каждой отдельно взятой переменной.
40. Что такое медиана?
41. Что такое корреляция?
42. Что означает критическое значение корреляции?
43. Какой критерий полезно применять для сравнения средней результативности в двух группах?
44. Что такое рандомизация?
45. При каком условии t-критерий будет положительным? Отрицательным?
46. Чему равно количество степеней свободы в общем случае?
47. Что называют индикаторной (фиктивной) переменной?
48. К какому уровню относятся индикаторные (фиктивные) переменные? Как можно обращаться с ними в статистическом смысле? Приведите пример распределения.
49. В основе чего лежит общая линейная модель?
50. Что такое линия регрессии?
51. Какой величиной обозначается ошибка?
52. В чем отличие линии регрессии от полного распределения значений?
53. Что включается в комплекс переменных у уравнения общей линейной модели?
54. С какой основной проблемой сталкивается исследователь, применяющий общую линейную модель?
55. Какие условия необходимо соблюдать для проведения постэкспериментального анализа двух групп по рандомизированному экспериментальному плану?
56. Почему необходимо учитывать вариабельность вокруг средних значений?
57. Сколько способов определения терапевтического эффекта существует при анализе по постэкспериментальному рандомизированному плану?
58. В чем заключается свойство коммутативности?

59. Каким образом может наклон оказаться различием между средними?
60. В каком алгебраическом виде формулируется регрессионная модель простого факторного плана  $2 \times 2$ ?
61. Как представить статистическую модель рандомизированного анализа блоков в алгебраической записи для регрессионного анализа для четырех блоков?
62. В каком алгебраическом виде записывается статистическая модель плана ковариационного анализа для одной ковариаты и одной контрольной и экспериментальной группы?
63. Что включает в себя статистическая модель анализа плана для неэквивалентных групп?
64. Какой коэффициент уравнения модели ковариационного анализа (ANCOVA) самый важный? Что он отражает?
65. Каков критерий проведения линии через облако из точек?
66. Влияет ли на наклон линий регрессии ошибка измерения, допущенная после эксперимента? Почему?
67. С каким измерением соотносится регрессионный анализ?
68. К чему приводит сглаживание линий?
69. Как устранить сглаживание линий регрессии?
70. Какова формула для коррекции ошибки измерения?
71. Если ошибка измерения до эксперимента отсутствует, то как в этом случае проводится коррекция данных измерения?
72. Какая модель ковариационного анализа (ANCOVA) применяется в статистическом анализе по плану неэквивалентных групп с целью исправления смещения, вызываемого ошибкой измерения до эксперимента?
73. Что представляет собой план перегиба линии регрессии в основном виде?
74. Какой функцией может быть описано распределение до и после эксперимента?
75. Что значит «точная конкретизация» общей линейной модели?
76. Каким уравнением описывается недостаточным образом конкретизированная модель?
77. До каких пор постепенно производят сокращение членов высокого порядка?
78. Как связана степень многочлена с количеством точек изгиба?
79. Что необходимо для статистического анализа вытеснения точек регрессии?

## 2. Выполните следующие тестовые задания:

1. Конкретное измеряемое значение, получаемое нами от единицы (респондента), входящей в выборку, называется:
- а) характеристика
  - б) ответ
  - в) статистика

- г) параметр  
д) габарит
2. Распределение бесконечного количества выборок одного и того же объема называют
- а) матрица  
б) частное  
в) многочлен  
г) константа  
д) выборочное распределение
3. Чем больше объем выборки, тем ближе выборка
- а) к нулю  
б) к выборочному среднему  
в) к стандартной ошибке  
г) к стандартному отклонению  
д) к генеральной совокупности
4. Вероятностным выборочным методом является любой метод составления выборки, в котором применяется какая-то форма
- а) случайного отбора  
б) естественного отбора  
в) мысленного эксперимента  
г) нормального распределения  
д) алгебраической модели
5. Когда нам известны шансы на то, что генеральная совокупность окажется представлена репрезентативно, такая выборка называется
- а) кластерной выборкой  
б) многоэтапной выборкой  
в) экспертной выборкой  
г) простой случайной выборкой  
д) выборкой «снежный ком»
6. Степень, в которой возможны обоснованные теоретические выводы на основе применяющейся в исследовании операционализации, называется
- а) мысленный эксперимент  
б) алгебраическая модель  
в) внутренняя валидность  
г) валидность выводов  
д) валидность конструкторов
7. Когда концепция или понятие транслируется в реальное функционирование, это называется
- а) рандомизация  
б) триангуляция  
в) схематизация  
г) популяция  
д) операционализация

8. На оценку способности нашей операционализации проводить различие между теми группами, между которыми она должна теоретически быть способна проводить различие, направлена
- а) прогностическая валидность
  - б) конвергентная валидность
  - в) дискриминантная валидность
  - г) одновременная валидность
  - д) валидность конструкта
9. Для подтверждения дискриминантной валидности, необходимо показать, что те измерения, которые не должны быть взаимосвязаны
- а) реально надежны
  - б) реально не взаимосвязаны
  - в) реально взаимосвязаны
  - г) в принципе неважно
  - д) затрудняюсь ответить
10. Матрица или таблица корреляций, устроенная таким образом, чтобы облегчить интерпретацию оценок валидности конструктов, называется
- а) номологическая сеть
  - б) линейная матрица
  - в) симметричная матрица
  - г) выборочное распределение
  - д) матрица мультисвойств-мультиметодов
11. В матрице мультисвойств-мультиметодов коэффициенты на диагоналях валидности должны быть
- а) отличаться от нуля
  - б) равны нулю
  - в) равны между собой
  - г) неравны между собой
  - д) в принципе неважно
12. Идея об одновременном рассмотрении паттерна конвергенции и дискриминации аналогична по своей цели
- а) матрице корреляций
  - б) матрице мультисвойств-мультиметодов
  - в) доверительному интервалу
  - г) выборочному распределению
  - д) номологической сети
13. Данные, применяемые для исследования теоретической модели, образуют
- а) наглядный паттерн
  - б) фактический паттерн
  - в) теоретический паттерн
  - г) наблюдаемый паттерн
  - д) психологический паттерн
14. Надежность предполагает качество
- а) отношения
  - б) измерения



- в) рассуждения
  - г) вычисления
  - д) применения
15. Случайная ошибка выталкивает наблюдаемые значения
- а) куда подальше
  - б) куда поближе
  - в) куда поглубже
  - г) туда, где надежнее
  - д) то вверх, то вниз
16. Когда мы говорим о заслуживающем доверия измерении, мы имеем в виду, что это измерение
- а) не надежно, но валидно
  - б) и надежно, и валидно
  - в) надежно, но не валидно
  - г) не надежно, и не валидно
  - д) в принципе не важно
17. Единственная вещь, которая является общей для двух наблюдений, это
- а) их истинные значения
  - б) их повторные значения
  - в) их первичные значения
  - г) их точные значения
  - д) их искаженные значения
18. Если измерение абсолютно надежно, то ошибка измерения
- а) равна 0
  - б) равна 1
  - в) равна 3,14
  - г) равна стандартному отклонению
  - д) равна стандартной ошибке
19. При оценке согласованности измерения в различные моменты времени, применяется
- а) надежность внешних наблюдений
  - б) тест-ретестовая надежность
  - в) надежность параллельных форм
  - г) внутренне согласованная надежность
  - д) старый проверенный способ
20. Оценка тест-ретестовой надежности подразумевает, что в перерыве между двумя измерениями измеряемый конструкт
- а) не претерпевает значительных изменений
  - б) претерпевает незначительные изменения
  - в) претерпевает значительные изменения
  - г) не претерпевает значительных искажений
  - д) претерпевает значительные искажения
21. Когда мы последовательно и систематическим образом измеряем неверное значение для всех испытуемых, то такое измерение
- а) валидно и надежно

- б) валидно, но не надежно
  - в) надежно, но не валидно
  - г) ни надежно, ни валидно
  - д) не репрезентативно
22. Взаимосвязь между значениями переменной выражается
- а) количеством измерений
  - б) надежностью измерений
  - в) валидностью измерений
  - г) дисперсией измерений
  - д) уровнем измерений
23. В прикладных психологических исследованиях большинство обсчитываемых переменных – это
- а) номинальные переменные
  - б) порядковые переменные
  - в) интервальные переменные
  - г) переменные отношений
  - д) в принципе любые переменные
24. Наиболее распространенной причиной шкалирования выступает необходимость
- а) понять, что ответил испытуемый
  - б) исправить ответ испытуемого
  - в) подсказать испытуемому правильный ответ
  - г) выразить ответ испытуемого в баллах
  - д) узнать, на какой вопрос ответил испытуемый
25. Значение, ниже которого оказываются 25% случаев, называется
- а) первый процентиль
  - б) первый квартиль
  - в) медиана
  - г) мода
  - д) стандартное отклонение
26. Цель шкалирования по Гуттману состоит в разработке
- а) параллельного измерения
  - б) валидного континуума
  - в) одномерного континуума
  - г) надежного континуума
  - д) многомерного конструкта
27. Приблизительная истинность в отношении выводов о причинно-следственных либо о причинных взаимосвязях называется
- а) валидность конструктов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность

28. Какая валидность указывает на то, хотели ли вы выполнить именно ту программу, которую выполнили, либо на то, хотели ли вы наблюдать именно то, что в итоге наблюдается?
- а) валидность конструкторов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность
29. Феномен регрессии к среднему возникает
- а) только с психологическими переменными
  - б) только с математическими переменными
  - в) только с независимыми переменными
  - г) только с зависимыми переменными
  - д) с любыми переменными
30. Когда имеется выборка, пусть и неслучайная, но представляющая собой хорошую подвыборку из генеральной совокупности, то регрессия к среднему
- а) не возникнет
  - б) не окажет сильного влияния
  - в) не окажет никакого влияния
  - г) окажет сильное влияние
  - д) окажет такое влияние, какое нам нужно
31. Зная формулу вычисления процента регрессии, можно вычислить
- а) линейную корреляцию
  - б) ранговую корреляцию
  - в) регрессию к среднему
  - г) альфу Кронбаха
  - д) расколотые корреляции
32. По сравнению с непосредственно представленными данными, выводы на основе статистических расчетов во многих случаях
- а) намного точнее
  - б) намного приблизительнее
  - в) намного шире
  - г) намного уже
  - д) одинаково
33. Значение степени вероятности, на основании которого решается, можно ли считать полученный результат достоверным, равно
- а) 3,14
  - б) 0,1
  - в) 0,01
  - г) 0,05
  - д) 1
34. Что произойдет с сигналом, если сделать взаимосвязь более заметной?
- а) сигнал усилится
  - б) сигнал ослабеет
  - в) шум усилится

- г) шум станет невыносимым  
д) ничего не изменится
35. Нулевая гипотеза названа так потому, что она обычно соотносится со случаем
- а) нет погрешности
  - б) нет искажения
  - в) нет надежности
  - г) нет эффекта
  - д) нет правил
36. Чем ниже  $\alpha$ , тем ниже
- а) валидность
  - б) надежность
  - в) погрешность
  - г) вероятность
  - д) мощность
37. Как влияет увеличение  $\alpha$  на строгость тестирования?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет
38. Способ хранения данных в базе для их последующего анализа образует её
- а) систему
  - б) структуру
  - в) функцию
  - г) разновидность
  - д) комплекс
39. Для представления количественных описаний в удобной форме, применяются
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
40. Если в ряду 500 значений, то медианой будет
- а) значение номер 0
  - б) значение номер 1
  - в) значение номер 250
  - г) значение номер 251
  - д) значение номер 500
41. Отношение ряда чисел к среднему по выборке называется
- а) стандартная ошибка
  - б) стандартное отклонение
  - в) стандартное значение
  - г) стандартный размер

- д) стандартная оценка
42. Число, которым описывается степень взаимосвязи между переменными, называется
- а) размах
  - б) дисперсия
  - в) медиана
  - г) мода
  - д) корреляция
43. Когда на основании имеющихся данных необходимо сделать обобщенные заключения, мы прибегаем
- а) к статистическим оценкам
  - б) к статистическим выводам
  - в) к статистическим тенденциям
  - г) к статистическим отклонениям
  - д) к статистическим погрешностям
44. Ковариационная модель анализа коррекции надежности необходима при анализе
- а) по факторному экспериментальному плану
  - б) по плану посттестового рандомизированного эксперимента
  - в) по экспериментальному плану ковариационного анализа
  - г) по плану точечного замещения регрессии
  - д) по плану неэквивалентных групп
45. Когда необходимо учитывать различия между средними значениями относительно распределения или вариабельности баллов, уместно применение
- а) корреляционной матрицы
  - б) медианы
  - в) моды
  - г) t-критерия
  - д) квотной выборки
46. Для проверки значимости t-критерия, нужно задать
- а) доверительный интервал
  - б) уровень значимости
  - в) уровень случайности
  - г) интервал вероятности
  - д) уровень риска
47. Зная альфа-уровень, количество степеней свободы и значение t-критерия, можно найти t-значение в стандартных таблицах значимости и определить – достаточно ли велико t-значение для того, чтобы считать его значимым. Если ответ «да», тогда делается вывод о том, что различие между средними двух групп
- а) велико
  - б) мало
  - в) положительно
  - г) отрицательно
  - д) существует

48. Количественная переменная, используемая в регрессионном анализе для представления выборочных подгрупп, называется

- а) независимая переменная
- б) зависимая переменная
- в) индикаторная (фиктивная) переменная
- г) шум
- д) константа

49. В основе t-критерия, дисперсионного анализа (ANOVA), ковариационного анализа (ANCOVA), регрессионного анализа и многих многомерных методов, включая факторный анализ, кластерный анализ, многомерное шкалирование, анализ дискриминант функций, канонические корреляции и т. п. заложена

- а) ковариационная модель
- б) теоретическая модель
- в) алгебраическая модель
- г) общая линейная модель
- д) модель мысленного эксперимента

50. Для того, чтобы описать линию, проходящую через двумерный график, необходим компонент, который называется

- а) интервал
- б) отклонение
- в) оценка
- г) ошибка
- д) наклон

51. Каким образом среднее значение отражает каждое значение в распределении

- а) отражает точно
- б) отражает неточно
- в) не отражает совсем
- г) в принципе неважно
- д) затрудняюсь ответить

52. Если состояние до эксперимента описывается несколькими переменными, то их включают в комплекс

- а) переменных  $x$
- б) переменных  $y$
- в) переменных  $z$
- г) индикаторной (фиктивной) переменной
- д) паттернов

53. Разность по горизонтали между проекциями средних значений в экспериментальной группе и в контрольной группе на горизонтальной оси, называется

- а) сходство между средними значениями
- б) различие между средними значениями
- в) близость между средними значениями
- г) тождественность средних значений

- д) ошибка в средних значениях
54. Большая вариабельность означает большой
- а) сигнал
  - б) шум
  - в) гам
  - г) кошмар
  - д) валидность
55. Наиболее общим способом определения терапевтического эффекта выступает
- а) дисперсионный анализ (ANOVA)
  - б) ковариационный анализ (ANCOVA)
  - в) регрессионный анализ
  - г) факторный анализ
  - д) кластерный анализ
56. В регрессионной модели простого факторного плана  $2 \times 2$  для каждого фактора применяются
- а) независимые переменные
  - б) зависимые переменные
  - в) индикаторные (фиктивные) переменные
  - г) паттерны
  - д) коэффициенты наклона
57. При конкретизации статистической модели рандомизированного анализа блоков используется множество
- а) независимых переменных
  - б) зависимых переменных
  - в) индикаторных (фиктивных) переменных
  - г) паттернов
  - д) коэффициентов наклона
58. При конкретизации статистической модели плана ковариационного анализа, применяются
- а) независимые переменные
  - б) зависимые переменные
  - в) индикаторные (фиктивные) переменные
  - г) паттерны
  - д) коэффициенты наклона
59. Средние различия между экспериментальной группой и контрольной группой отражают тот коэффициент, который связан с переменной
- а) переменной X
  - б) переменной Z
  - в) переменной Y
  - г) переменной e
  - д) переменной  $\beta$
60. Вертикальное расстояние от линии регрессии до каждой точки называется
- а) остатком от деления
  - б) остатком от вычитания

- в) сухим остатком
- г) врожденными задатками
- д) атмосферными осадками

61. При регрессионном анализе линии регрессии проводятся с учетом смещения точек

- а) влево
- б) вправо
- в) по вертикали
- г) по горизонтали
- д) по часовой стрелке

62. Когда мы прибавляем ошибку измерения, допущенную до эксперимента, то мы фактически удлиняем

- а) вертикальное измерение
- б) горизонтальное измерение
- в) четвертое измерение
- г) порядковое измерение
- д) интервальное измерение

63. Необходимым условием искажения при исследовании по плану неэквивалентных групп оказывается

- а) точность измерения
- б) надежность измерения
- в) валидность измерения
- г) тщательность измерения
- д) ошибка измерения

64. Отсекающий критерий должен соблюдаться

- а) по мере возможности
- б) по мере надобности
- в) по желанию
- г) от случая к случаю
- д) без исключений

65. Если вычесть значение отсекания из каждого доэкспериментального значения, тогда те доэкспериментальные значения, которые изначально были отсекающими значениями, окажутся равными

- а) минус 1
- б) ноль
- в) плюс 1
- г) 3,14
- д) 95%

66. По причине того, что в истинную модель оказались включены те члены, которые могут не быть необходимы, имеется вероятность того, что оценка окажется

- а) равной нулю
- б) отличной от нуля
- в) действительной
- г) недействительной



- д) удивительной
67. Оценку вертикального вытеснения по отношению к линии регрессии выражает коэффициент, ассоциированный с
- а) переменной  $x$
  - б) переменной  $y$
  - в) переменной  $z$
  - г) индикаторной (фиктивной) переменной
  - д) паттерном

## Вариант II

1. Сформулируйте ответы на следующие вопросы:

1. Что называется «выборочным распределением»?
2. Сформулируйте правило 65%, 95% и 99%.
3. По каким причинам стратифицированная выборка более предпочтительна, чем простая случайная выборка?
4. В чем заключается различие между вероятностной выборкой и невероятностной выборкой? Означает ли это различие, что невероятностные выборки не репрезентативны?
5. Что включают в модальную выборку?
6. Кто участвует в панельных исследованиях?
7. Что такое операционализация?
8. Что оценивает прогностическая валидность?
9. Перечислите требования, которые необходимо соблюдать для того, чтобы обеспечить валидность конструкторов.
10. Что необходимо показать для подтверждения дискриминантной валидности?
11. Из чего состоит номологическая сеть?
12. Благодаря чему можно улучшить способность выявлять конвергентную валидность и дискриминантную валидность?
13. Как называется собрание или упорядочение операционализаций, релевантных теоретическому паттерну?
14. Как получить корреляцию паттерна согласования?
15. Из каких компонентов складывается любое измерение?
16. Каким образом могут быть уменьшены случайные и систематические ошибки измерения?
17. Можно ли рассчитать дисперсию истинных значений?
18. Как определить интервал оценки надежности?
19. Каким образом можно обеспечить надежность внешних наблюдений? Сколько способов Вы знаете?
20. В чем заключается главное отличие подхода на основе параллельных форм от расколотой пополам надежности?

21. Что такое расколотая пополам надежность?
22. Почему уровень измерения важен?
23. Каким образом соподчиняются уровни измерений?
24. В чем отличие шкалы ответов от настоящего шкалирования?
25. Каковы преимущества применения многомерной модели?
26. Какие шкалы называются реверсивными?
27. Что такое интерквартильная широта?
28. Каким исследованиям релевантна внутренняя валидность?
29. Что такое ковариация причины и следствия?
30. В каких случаях возникает регрессия к среднему? Что это такое?
31. По какой формуле вычисляется процент регрессии к среднему?
32. Что такое «социальные» угрозы внутренней валидности? Перечислите их и охарактеризуйте.
33. В чём отличие валидности выводов от внутренней валидности?
34. На каких допущениях основан научный анализ?
35. Что такое размер эффекта?
36. Почему нулевая гипотеза так называется?
37. Что означает  $\alpha$  равная 0,01?
38. Назовите и охарактеризуйте наиболее распространенные преобразования данных.
39. Что называется распределением? Приведите пример.
40. Что называется в статистике модой?
41. По какой формуле рассчитывается корреляция?
42. Какая матрица называется симметричной?
43. В семью каких статистических моделей входит этот критерий?
44. В каких случаях уместно применять t-критерий?
45. Чем является дисперсия?
46. Что следует знать для того, чтобы найти t-значение в стандартных таблицах значимости?
47. Где применяют индикаторные (фиктивные) переменные?
48. Как определяется эффективность каждой группы?
49. Чему соответствует каждая точка на двумерном графике?
50. Какой анализ называется регрессионным?
51. Почему был выбран термин «ошибка»?
52. Что представляет собой общая линейная модель?
53. Что включается в комплекс переменных  $x$  уравнения общей линейной модели?
54. Числовые соответствия каких значений следует знать для проверки конкретных исследовательских гипотез относительно взаимосвязей между переменными либо относительно различий между группами?
55. В каком виде можно представить каждую группу испытуемых?
56. Чему соответствует большая вариабельность?
57. Как оценивают размер эффекта?
58. Что такое наклон  $\beta_1$ ?
59. Чему равно t-значение для коэффициента  $\beta_1$ ?

60. Сколько факторов представлено в регрессионной модели простого факторного плана  $2 \times 2$ ?
61. Каким символом в алгебраической записи статистической модели рандомизированного анализа блоков обозначена экспериментальная группа?
62. Чему в алгебраической записи статистической модели плана ковариационного анализа для одной ковариаты и одной контрольной и экспериментальной группы данной модели соответствует индикаторная (фиктивная) переменная  $Z_i$ ?
63. Что позволяет оценить статистическая модель анализа плана для неэквивалентных групп?
64. Что показывает индикаторная (фиктивная) переменная в этом уравнении?
65. Какие значения являются положительными (отрицательными)?
66. Как называется такой случай, когда ошибка измерения отсутствует?
67. Как отражается на наклоне линий регрессии ошибка измерения, допущенная до эксперимента?
68. Где уменьшается наклон линии?
69. Происходит ли сглаживание линий и искажение данных в рандомизированном случае? Почему?
70. С чем соотносится надежность измерения?
71. В чем заключается идея формулы для коррекции ошибки измерения?
72. Какой вид имеет корректировка надежности, равной 0,8?
73. Что должна включать в себя статистическая модель структуры плана перегиба линии регрессии в основном виде?
74. Какое распределение значений должно существовать для контрольной группы и экспериментальной группы?
75. Каким уравнением описывается точным образом конкретизированная модель?
76. Каким образом надлежит планировать стратегию моделирования?
77. Из скольких этапов состоит анализ перегиба линии регрессии? Назовите и охарактеризуйте эти этапы.
78. Что выражает коэффициент регрессии, связанный с членом  $Z$ ?
79. К какой ещё модели анализа предъявляются требования, аналогичные требованиям при статистическом анализе вытеснения точек регрессии?

## 2. Выполните следующие тестовые задания:

1. Причина, по которой мы обращаемся к выборке, заключается в том, что мы стремимся оценить
  - а) испытуемого
  - б) выборку
  - в) результат
  - г) генеральную совокупность
  - д) ответ
2. Среднее выборочного распределения в действительности эквивалентно

- а) 3,14  
б) нулю  
в) сумме квадратов катетов  
г) статистике  
д) параметру
3. Когда вместо выборки исследуется генеральная совокупность, то среднее значение является
- а) характеристикой  
б) ответом  
в) статистикой  
г) параметром  
д) габаритом
4. Таблица случайных чисел, компьютерный генератор случайных чисел или механическое устройство используются при составлении
- а) простой случайной выборки  
б) произвольной выборки  
в) кластерной выборки  
г) экспертной выборки  
д) выборки «снежный ком»
5. Исследователи предпочитают вероятностные или случайные выборочные методы перед невероятностными, считая их
- а) более современными  
б) более распространенными  
в) более точными и надежными  
г) более понятными  
д) более близкими
6. Валидность конструкторов имеет отношение
- а) к распределению  
б) к произведению  
в) к отклонению  
г) к параметру  
д) к обобщению
7. То, насколько хорошо операционализация отражает конструктор, определяется
- а) прогностической валидностью  
б) конвергентной валидностью  
в) дискриминантной валидностью  
г) одновременной валидностью  
д) валидностью транслирования
8. Отвергает идею о том, что в качестве основы для определения конструкторов мы можем полагаться на операциональные формулировки
- а) фрейдизм  
б) универсализм  
в) дефиниционализм  
г) реляционализм

- д) операционализм
9. Свидетельствует ли паттерн, подтверждающий и конвергентную валидность, и дискриминантную валидность, о том, что измерения данного конструкта действительно измеряют именно этот конструкт:
- а) разумеется, да
  - б) разумеется, нет
  - в) строгие правила отсутствуют
  - г) в принципе неважно
  - д) затрудняюсь ответить
10. Матрица или таблица корреляций, устроенная таким образом, чтобы облегчить интерпретацию оценок валидности конструктов, называется
- а) номологическая сеть
  - б) линейная матрица
  - в) симметричная матрица
  - г) выборочное распределение
  - д) матрица мультисвойств-мультиметодов
11. В матрице мультисвойств-мультиметодов коэффициенты на диагоналях валидности должны быть
- а) отличаться от нуля
  - б) равны нулю
  - в) равны между собой
  - г) неравны между собой
  - д) в принципе неважно
12. Идея об одновременном рассмотрении паттерна конвергенции и дискриминации аналогична по своей цели
- а) матрице корреляций
  - б) матрице мультисвойств-мультиметодов
  - в) доверительному интервалу
  - г) выборочному распределению
  - д) номологической сети
13. Данные, применяемые для исследования теоретической модели, образуют
- а) наглядный паттерн
  - б) фактический паттерн
  - в) теоретический паттерн
  - г) наблюдаемый паттерн
  - д) психологический паттерн
14. Теория истинного значения – это теория
- а) отношения
  - б) измерения
  - в) рассуждения
  - г) вычисления
  - д) применения
15. Если бы мы увидели все случайные ошибки в распределении, то их сумма оказалась бы
- а) равной 3,14

- б) равной нулю
  - в) четной
  - г) нечетной
  - д) надежной
16. В научном исследовании термин «надежность» означает допускающий возможность
- а) измерения
  - б) вычисления
  - в) сомнения
  - г) отрицания
  - д) повторения
17. Надежность можно себе представить в качестве пропорции
- а) вероятности в измерении
  - б) наглядности в измерении
  - в) истины в измерении
  - г) метода в измерении
  - д) принципа в измерении
18. Если измерение абсолютно ненадежно, то ошибка измерения
- а) равна 0
  - б) равна 1
  - в) равна 3,14
  - г) равна стандартному отклонению
  - д) равна стандартной ошибке
19. При оценке согласованности результатов по пунктам одного и того же тестового задания, применяется
- а) надежность внешних наблюдений
  - б) тест-ретестовая надежность
  - в) надежность параллельных форм
  - г) внутренне согласованная надежность
  - д) старый проверенный способ
20. Чем короче временной интервал между двумя измерениями измеряемого конструкта, тем
- а) меньше корреляция
  - б) больше корреляция
  - в) симметричнее корреляция
  - г) негативнее корреляция
  - д) позитивнее корреляция
21. При каком измерении, количественное обозначение оказывается всего лишь уникальным наименованием свойства
- а) номинальное измерение
  - б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
22. Уровни измерений соподчиняются

- а) параллельно
  - б) последовательно
  - в) синхронно
  - г) иерархически
  - д) взаимно
23. Какое количество размеров может быть у шкалы:
- а) один размер
  - б) два размера
  - в) много размеров
  - г) сколько угодно размеров
  - д) столько, сколько испытуемых
24. Какому процентилю соответствует третий квартиль
- а) первому процентилю
  - б) двадцать пятому процентилю
  - в) пятидесятому процентилю
  - г) семьдесят пятому процентилю
  - д) тысячному процентилю
25. Если по шкале Гутсмана испытуемый набрал восемь баллов, значит, с первыми восемью утверждениями испытуемый
- а) не знаком
  - б) не согласился
  - в) знаком слабо
  - г) согласился
  - д) ни о чём не говорит
26. В большинстве исследований, связанных с наблюдением и с описанием, внутренняя валидность
- а) возможна
  - б) реальна
  - в) допускается
  - г) релевантна
  - д) не релевантна
27. Какая валидность занимается решением проблемы «третьей переменной»?
- а) валидность конструкторов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность
28. Какое отношения имеет регрессия к среднему к общей тенденции к созреванию?
- а) никакого отношения
  - б) косвенное отношение
  - в) непосредственное отношение
  - г) случайное отношение
  - д) чисто деловое отношение

29. Если выборка оказывается экстремальной по отношению к генеральной совокупности, то регрессия к среднему
- а) не возникнет
  - б) не окажет сильного влияния
  - в) не окажет никакого влияния
  - г) окажет сильное влияние
  - д) окажет такое влияние, какое нам нужно
30. Насколько смещается выборочная группа, если корреляция  $r = 0,5$ ?
- а) на 100%
  - б) на 80%
  - в) на 50%
  - г) на 20%
  - д) на 0%
31. Какой характер имеют статистические выводы относительно данных?
- а) конкретный
  - б) общий
  - в) валидный
  - г) надежный
  - д) предварительный
32. Значение степени вероятности 0,05 означает, что из каждых 100 раз, подвергаемых статистическому анализу, полученный результат имеет вероятность происходить случайным образом, по крайней мере
- а) 100 раз
  - б) 95 раз
  - в) 50 раз
  - г) 5 раз
  - д) 0 раз
33. Что произойдет с шумом, если увеличить надежность?
- а) сигнал усилится
  - б) сигнал ослабеет
  - в) шум усилится
  - г) шум станет невыносимым
  - д) ничего не изменится
34. В психологических исследованиях значение  $\alpha$  как правило выбирается равным
- а) 3,14
  - б) 0,1
  - в) 0,01
  - г) 0,05
  - д) 1
35. Чем ниже  $\alpha$ , тем меньше
- а) валидность конструкта
  - б) надежность измерения
  - в) погрешность измерения
  - г) вероятность ошибки



- д) мощность критерия
36. Как влияет увеличение  $\alpha$  на мощность?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет
37. Описание данных и путь доступа к данным указывается
- а) в книге чисел
  - б) в книге перемен
  - в) в книге кодов
  - г) в книге жалоб
  - д) в красной книге
38. Помогают нам разумным образом упорядочить большие объемы собранных данных
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
39. Наиболее распространенное значение в ряду чисел, называется
- а) среднее
  - б) мода
  - в) медиана
  - г) дисперсия
  - д) ранг
40. Для того, чтобы рассчитать стандартное отклонение, сначала необходимо найти расстояние между каждым значением и
- а) модой
  - б) медианой
  - в) средним
  - г) дисперсией
  - д) отклонением
41. Выражение «позитивная взаимосвязь» означает, что более высоким значениям по одной переменной соответствуют по второй переменной
- а) более низкие значения
  - б) более высокие значения
  - в) равные значения
  - г) нулевые значения
  - д) число  $\pi$
42. При сравнении средней результативности в двух группах, полезно применять
- а) корреляционную матрицу
  - б) медиану
  - в) моду

- г) t-критерий  
д) квотную выборку
43. На традиционной модели ковариационного анализа (ANCOVA) непосредственно основан анализ
- а) по факторному экспериментальному плану
  - б) по плану посттестового рандомизированного эксперимента
  - в) по экспериментальному плану ковариационного анализа
  - г) по плану точечного замещения регрессии
  - д) по плану неэквивалентных групп
44. Формула для расчета t-критерия представляет собой
- а) интеграл
  - б) радикал
  - в) квадратный корень
  - г) дробь
  - д) матрицу
45. В большинстве психологических исследований эмпирически принято задавать уровень альфа равным
- а) 3,14
  - б) 1,0
  - в) 0,9
  - г) 0,05
  - д) 0
46. Уровень альфа, равный 0,05, означает, что, даже если статистически значимого различия не существует, мы обнаруживаем статистически значимое различие между средними значениями
- а) сто раз из ста
  - б) 99 раз из ста
  - в) 95 раз из ста
  - г) 65 раз из ста
  - д) пять раз из ста
47. В простейшем случае применяется индикаторная (фиктивная) переменная со значениями
- а) 1 и 2
  - б) 0 и 2
  - в) 0 и 1
  - г) 0 и 10
  - д) 2 и 1
48. Взаимосвязь, при которой испытуемые, балл которых до эксперимента был выше, получили более высокий балл после эксперимента, и наоборот, называется
- а) нулевая взаимосвязь
  - б) двусторонняя взаимосвязь
  - в) односторонняя взаимосвязь
  - г) позитивная взаимосвязь
  - д) негативная взаимосвязь

49. Величина, описывающая расстояние по вертикали от прямой линии до каждой точки, называется
- а) интервал
  - б) отклонение
  - в) оценка
  - г) ошибка
  - д) наклон
50. Каким образом линия регрессии отражает каждое значение в двумерном распределении
- а) отражает полно
  - б) отражает кратко
  - в) не отражает совсем
  - г) в принципе неважно
  - д) затрудняюсь ответить
51. Колоколообразную кривую можно вообразить в виде
- а) в виде матрицы корреляций
  - б) в виде уравнения
  - в) в виде гистограммы
  - г) в виде доверительного интервала
  - д) в виде континуума
52. Межгрупповые различия более очевидны, когда велик
- а) сигнал
  - б) шум
  - в) гам
  - г) кошмар
  - д) валидность
53. Формула  $y_i = \beta_0 + \beta_1 Z_i + \epsilon_i$  представляет собой
- а) уравнение параболы
  - б) уравнение гиперболы
  - в) уравнение прямой
  - г) уравнение окружности
  - д) уравнение химической реакции
54. В данной модели, основные эффекты выражаются через статистики, в которых значения бета выступают смежными
- а) с переменной  $x$
  - б) с переменной  $y$
  - в) с переменной  $z$
  - г) с индикаторной (фиктивной) переменной
  - д) с коэффициенты наклона
55. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $\epsilon_i$  означает
- а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания

- д) индикаторную (фиктивную) переменную
56. Экспериментальной группе соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $u_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
57. Для того, чтобы определить, есть ли статистически значимое различие между полученным значением и истинным значением, нужно задать
- а) ширину и высоту
  - б) угол наклона
  - в) стандартное отклонение
  - г) стандартную ошибку
  - д) доверительный интервал
58. Регрессионный анализ предполагает проведение линии через те данные, которые образуют наименьшую сумму квадратов
- а) остатков от деления
  - б) остатков от вычитания
  - в) сухих остатков
  - г) врожденных задатков
  - д) атмосферных осадков
59. Если ошибки рандомизированы, то мы получим столько остатков разностей, смещающих наклон линии одновременно и вверх и вниз, что в среднем наклон линии
- а) сильно увеличится
  - б) сильно уменьшится
  - в) несильно увеличится
  - г) несильно уменьшится
  - д) не изменится
60. Если ошибка в измерении до эксперимента отсутствует, тогда надежность измерения равна
- а) минус 1
  - б) ноль
  - в) плюс 1
  - г) 3,14
  - д) 95%
61. Искажения, имеющие отношение к отсекающему значению, часто называются
- а) статистическими
  - б) противными
  - в) вредными
  - г) белыми
  - д) пушистыми

62. Изменение вертикального уровня (основной эффект), изменение наклона (эффект взаимодействия), либо и то, и другое указывает на то, что экспериментальный эффект

- а) велик
- б) мал
- в) не велик и не мал
- г) существует
- д) затрудняюсь ответить

63. Направление эффекта интерпретируется с учетом знака

- а) стандартной ошибки
- б) стандартного отклонения
- в) теоретического паттерна
- г) коэффициента регрессии
- д) коэффициента корреляции

64. В модели ковариационного анализа вытеснения точек регрессии переменная  $e_i$  выражает

- а) значение  $i$ -того члена
- б) преэспериментальный коэффициент
- в) среднее различие в эксперименте
- г) индикаторную (фиктивную) переменную
- д) остаток для  $i$ -того члена после вычитания

### Вариант III

1. Сформулируйте ответы на следующие вопросы:

1. Почему выборочное распределение рассматривают в редких случаях?
2. Как называются интервалы 65%, 95% и 99%?
3. Каков алгоритм составления систематической случайной выборки?
4. Почему исследователи предпочитают вероятностные или случайные выборочные методы перед невероятностными?
5. Приведите примеры произвольных выборок.
6. Кто участвует в панельных исследованиях?
7. Какие типы валидности встречаются в исследовательских работах, когда идет речь о качестве измерения?
8. Что оценивает одновременная валидность?
9. Сформулируйте определения конвергентной и дискриминантной валидности.
10. Как преодолевается неадекватное дооперациональное объяснение конструкторов?
11. На каких принципах основывается номологическая сеть?
12. Каким образом выявляются различные части матрицы мультисвойств-мультиметодов?
13. Что называют словом «паттерн»?

14. О чем можно сделать вывод на основании того, в какой степени паттерны согласуются друг с другом?
15. Из чего образуется вариабельность измерения?
16. Что такое надежность в научном исследовании?
17. Можно ли вычислить надежность?
18. Чему равна ошибка измерения, если измерение абсолютно надежно?
19. Каким образом оценить надежность внешних наблюдений в тех случаях, когда измерение является непрерывным?
20. Как проводится оценка внутренне согласованной надежности?
21. Чему в математическом смысле эквивалентна альфа Кронбаха?
22. Сколько уровней измерения существует? Как они называются?
23. Какая область измерения называется шкалированием?
24. Что такое шкала ответов?
25. Каким может быть двумерный конструкт? Приведите пример.
26. Чем различаются одномерные методы шкалирования?
27. В чем заключается цель шкалирования по Гуттману?
28. В чем заключается основной вопрос применительно к внутренней валидности?
29. Означает ли демонстрация существования взаимосвязи то, что это каузальная взаимосвязь? Почему?
30. Возможно ли на основании явления регрессии к среднему предвидеть то, каким образом сместятся баллы индивидуального испытуемого?
31. По какой формуле вычисляется процент регрессии к среднему?
32. Каков процент регрессии, если корреляция равна 0,5?
33. Какие причины затрудняют валидность выводов? Перечислите их и охарактеризуйте. С чем соотносятся эти причины?
34. Что такое нарушение предположений о статистических тестах? Приведите пример.
35. Какие стратегии увеличения размера эффекта Вы знаете?
36. В чем отличие ошибки первого типа от ошибки второго типа?
37. Что означает 0,05 уровень значимости? Приведите пример.
38. Что такое обращение баллов? По какой формуле оно рассчитывается?
39. Что такое частота распределения?
40. Что такое дисперсия? Сколько показателей может быть у дисперсии? Какой из них является более точным?
41. Как проверить корреляцию?
42. Что такое корреляционная матрица?
43. Что выступает основой для понимания того, каким образом производится сравнение групп?
44. Какие распределения обычно изображаются при помощи гистограммы?
45. Что означает название «степени свободы»?
46. Чему математически эквивалентен t-критерий?
47. Чем объясняется полезность индикаторных (фиктивных) переменных? Что это означает?

48. Чему равно уравнение для контрольной группы? Как выразить его алгебраически?
49. Что такое позитивная взаимосвязь?
50. Какой вид имеет уравнение прямой?
51. Какой анализ называется регрессионным?
52. Когда мы применяем к нашим данным двумерную линейную модель, то какие данные по каждому испытуемому нам известны? Что они означают?
53. В каком алгебраическом виде записывается общая линейная модель?
54. Какое значение соответствует в общей линейной модели представлению взаимосвязи  $x$  и  $y$ ?
55. Как выглядят различия между средними значениями?
56. Когда различия между группами более очевидны?
57. Какая алгебраическая модель применяется при постэкспериментальном анализе?
58. Как называются межгрупповые различия?
59. Чему равен наклон линии регрессии?
60. Какие факторы представлены в регрессионной модели простого факторного плана  $2 \times 2$ ?
61. Какими переменными в алгебраической записи статистической модели рандомизированного анализа блоков обозначены второй, третий и четвертый блоки?
62. Чему в алгебраической записи статистической модели плана ковариационного анализа для одной ковариаты и одной контрольной и экспериментальной группы данной модели соответствует значение  $\beta_0$ ?
63. Какие различия позволяет устанавливать ковариационный анализ (ANCOVA)?
64. Чему равен 95%-ный доверительный интервал? Что это означает?
65. Что такое остаточные разности?
66. Что происходило бы с линиями, если бы ошибки измерения не было?
67. В результате чего возникает псевдоэффект?
68. Какими причинами вызывается искажение при применении модели ковариационного анализа (ANCOVA) к плану исследования неэквивалентных групп?
69. Если половина измерений истинна, а другая половина измерений ошибочна, тогда чему равна надежность?
70. На какую величину требуется корректировать значения по каждому испытуемому?
71. Что означает «сжатие» распределения на величину, пропорциональную ошибке измерения?
72. При помощи чего оценивается нижняя граница надежности?
73. Что необходимо допустить прежде, чем приступить к рассмотрению конкретной статистической модели?
74. Что означает, что эксперимент проводится единообразно в отношении всех испытуемых?
75. Что значит «избыточная конкретизация» общей линейной модели?

76. В силу чего точная конкретизация истинной модели бывает труднодостижима?

77. С какой целью из каждого доэкспериментального значения вычитают отсекающие значения?

78. В чем состоит опасность процедуры конкретизации модели, предполагающей отбрасывание отдельных членов на отдельных этапах анализа?

79. В чём заключается цель анализа вытеснения точек регрессии?

## 2. Выполните следующие тестовые задания:

1. Если мы измеряем всю генеральную совокупность и рассчитываем среднее значение, то оно называется

- а) характеристика
- б) ответ
- в) статистика
- г) параметр
- д) габарит

2. Среднее значение от средних значений бесконечного количества выборок называется

- а) среднее значение суммы
- б) среднее значение произведения
- в) среднее значение выборочного распределения
- г) среднее значение матрицы
- д) среднее значение многочлена

3. Если отклониться от среднего вверх и вниз (т. е. влево и вправо) на одно стандартное значение, тогда мы охватим

- а) 65% всех случаев данного распределения
- б) 25% всех случаев данного распределения
- в) 95% всех случаев данного распределения
- г) 99% всех случаев данного распределения
- д) 50% всех случаев данного распределения

4. Стратифицированную случайную выборку также называют

- а) кластерной выборкой
- б) пропорциональной или квотной случайной выборкой
- в) экспертной выборкой
- г) простой случайной выборкой
- д) выборкой «снежный ком»

5. Выборка, составленная из тех людей, о которых известно, что они обладают специальными познаниями, называется

- а) кластерной выборкой
- б) многоэтапной выборкой
- в) экспертной выборкой
- г) простой случайной выборкой
- д) выборкой «снежный ком»



6. Если удастся показать, что измерения, которые теоретически считаются сильно взаимосвязанными, на практике сильно взаимосвязаны, то такая валидность называется
- а) внутренняя валидность
  - б) валидность выводов
  - в) валидность конструкторов
  - г) конвергентная валидность
  - д) дискриминантная валидность
7. То, насколько хорошо операционализация отражает конструкт, определяется
- а) прогностической валидностью
  - б) конвергентной валидностью
  - в) дискриминантной валидностью
  - г) одновременной валидностью
  - д) валидностью транслирования
8. Вещи не являются или черными, или белыми – все понятия более или менее соотносятся между собой, с точки зрения
- а) фрейдизма
  - б) универсализма
  - в) дефиниционализма
  - г) реляционализм
  - д) операционализма
9. Для подтверждения дискриминантной валидности, необходимо показать, что те измерения, которые не должны быть взаимосвязаны
- а) реально надежны
  - б) реально не взаимосвязаны
  - в) реально взаимосвязаны
  - г) в принципе неважно
  - д) затрудняюсь ответить
10. В отношении независимой переменной или причины в эксперименте возможен
- а) монооперациональный уклон
  - б) мультиоперациональный уклон
  - в) монометодический уклон
  - г) мультиметодический уклон
  - д) сочетание уклонов
11. По своей сути, матрица мультисвойств-мультиметодов – это
- а) транспонированная матрица
  - б) обратная матрица
  - в) единичная матрица
  - г) корреляционная матрица
  - д) выборочное распределение
12. В матрице мультисвойств-мультиметодов коэффициент валидности должен быть
- а) равен значениям в одном с ним столбце

- б) меньше, чем значения в одном с ним столбце
  - в) больше, чем значения в одном с ним столбце
  - г) равен нулю
  - д) в принципе неважно
13. Идея об одновременном рассмотрении паттерна конвергенции и дискриминации аналогична по своей цели
- а) матрице корреляций
  - б) матрице мультисвойств-мультиметодов
  - в) доверительному интервалу
  - г) выборочному распределению
  - д) номологической сети
14. Гипотеза о том, чего следует ждать от полученных данных, называется
- а) наглядный паттерн
  - б) фактический паттерн
  - в) теоретический паттерн
  - г) наблюдаемый паттерн
  - д) психологический паттерн
15. В теории истинного значения утверждается, что каждое измерение представляет собой
- а) произведение двух слагаемых
  - б) сумму двух слагаемых
  - в) разность двух слагаемых
  - г) квадрат двух слагаемых
  - д) матрицу двух слагаемых
16. На среднем функционировании группы испытуемых случайная ошибка
- а) отражается
  - б) преломляется
  - в) ускоряется
  - г) замедляется
  - д) не отражается
17. Для того, чтобы увидеть – насколько наблюдение согласовано, его можно
- а) измерить правильно
  - б) измерить внимательно
  - в) измерить тщательно
  - г) измерить точно
  - д) измерить дважды
18. Мы не можем рассчитать надежность, но мы можем её
- а) угадать
  - б) распознать
  - в) отменить
  - г) оценить
  - д) озвучить
19. Надежность всегда находится в интервале
- а) от  $-1$  до  $+1$
  - б) от  $0$  до  $+1$

- в) от  $-1$  до  $0$
- г) от  $-\infty$  до  $+\infty$
- д) от  $\max$  до  $\min$

20. Допустим, что внешние наблюдатели одинаковым образом распределили по категориям 86 наблюдений из 100. В этом случае процент согласия внешних наблюдателей составляет

- а) 0%
- б) 14 %
- в) 86%
- г) 100%
- д) 200%

21. Проблема в связи с оценкой надежности параллельных форм состоит в необходимости генерирования

- а) большого количества экспериментов
- б) большого количества испытуемых
- в) большого количества опросников
- г) большого количества пунктов опросника
- д) большого количества оценок испытуемых

22. При каком измерении, свойства допускают ранжирование

- а) номинальное измерение
- б) порядковое измерение
- в) интервальное измерение
- г) измерение отношений
- д) в принципе любое измерение

23. Область измерения, связанная с конструированием измерительного инструментария, которая сопрягает качественные конструкты с количественными метрическими единицами, называется

- а) тестирование
- б) ранжирование
- в) шкалирование
- г) функционирование
- д) коррелирование

24. Метрику шкалы можно себе представить в виде количества

- а) испытуемых
- б) ответов
- в) мнений
- г) баллов
- д) линий

25. Интерквартильная широта – это

- а) разность третьего и первого квартилей
- б) разность восьмого и четвертого квартилей
- в) разность любых квартилей
- г) разность квартилей и процентилей
- д) косинус угла альфа

26. Когда судьи оценивают утверждения для формируемой шкалы, то согласие судей с оцениваемыми утверждениями
- а) желательно
  - б) обязательно
  - в) единодушное
  - г) ни о чём не говорит
  - д) не важно
27. В тех исследованиях, в которых делаются попытки установить каузальные взаимосвязи, внутренняя валидность
- а) возможна
  - б) реальна
  - в) допускается
  - г) релевантна
  - д) не релевантна
28. Для того, чтобы иметь основания утверждать, что вами продемонстрировано существование внутренней валидности, необходимо
- а) подобрать альтернативные достоверные объяснения
  - б) разобрать альтернативные достоверные объяснения
  - в) включить альтернативные достоверные объяснения
  - г) исключить альтернативные достоверные объяснения
  - д) обосновать альтернативные достоверные объяснения
29. Возможно ли на основании явления регрессии к среднему предвидеть то, каким образом сместятся баллы индивидуального испытуемого?
- а) всегда возможно
  - б) в ряде случаев возможно
  - в) возможно, если приобрести необходимый навык
  - г) возможно в редких случаях
  - д) никогда не возможно
30. Если ваша выборка состоит из тех, чьи баллы ниже среднего балла в генеральной совокупности, то регрессия к среднему заставит их
- а) опуститься
  - б) подняться
  - в) замереть
  - г) отжаться
  - д) исчезнуть
31. Если две переменные коррелируют абсолютно, то регрессия к среднему
- а) не возникнет
  - б) не окажет сильного влияния
  - в) не окажет никакого влияния
  - г) окажет сильное влияние
  - д) окажет такое влияние, какое нам нужно
32. Что позволяет делать с собранными данными дескриптивная статистика, в отличие от статистических выводов?
- а) рассчитывать
  - б) прогнозировать

- в) анализировать
  - г) описывать
  - д) наблюдать
33. В психологических исследованиях эмпирически установлено, что значение статистической мощности должно быть
- а) не менее 1
  - б) не менее 0,8
  - в) не менее 0,2
  - г) не более 1
  - д) равно 0
34. С идеей о том, что шум создает «ошибку», затрудняющую возможность увидеть взаимосвязь, связана
- а) валидность
  - б) надежность
  - в) погрешность
  - г) вероятность
  - д) мощность
35. Вероятность того, что эффект или различие оказались случайными, называется
- а) роковая ошибка
  - б) простительная ошибка
  - в) ошибка первого типа
  - г) ошибка второго типа
  - д) медвежья услуга
36.  $\alpha$  равная 0,01 означает, что вероятность утверждать, что различия отсутствуют, когда различия действительно реально отсутствуют, равна
- а) 0%
  - б) 1%
  - в) 50%
  - г) 99%
  - д) 100%
37. Как влияет увеличение  $\alpha$  на вероятность ошибки первого типа (т. е. утверждения, что различия существуют, когда на самом деле различий не существует)?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет
38. Основу любого количественного анализа данных образуют
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики

39. Обобщение частоты отдельных значений или рангов значений переменной, называется
- а) размышление
  - б) наблюдение
  - в) распределение
  - г) валидность
  - д) надежность
40. Если распределение является нормальным (т. е. колоколообразным), то среднее, медиана и мода
- а) больше нуля
  - б) в сумме равны нулю
  - в) возводятся в квадрат
  - г) равны друг другу
  - д) не существуют
41. Если среднее равно 20,8750, а стандартное отклонение составляет 7,0799, то в пределах между 6,7152 и 35,0348 окажутся примерно
- а) 100% значений
  - б) 99% значений
  - в) 95% значений
  - г) 50% значений
  - д) 0% значений
42. Выражение «позитивная взаимосвязь» означает, что более низким значениям по одной переменной соответствуют по второй переменной
- а) более низкие значения
  - б) более высокие значения
  - в) равные значения
  - г) нулевые значения
  - д) число  $\pi$
43. В семью статистических моделей, известной под именем общей линейной модели, входит
- а) корреляционная матрица
  - б) медиана
  - в) мода
  - г) t-критерий
  - д) квотная выборка
44. Когда необходимо сравнить между собой средние значения по двум группам, а в особенности – при проведении анализа по посттестовому двухгрупповому рандомизированному экспериментальному плану, уместно применение
- а) корреляционной матрицы
  - б) медианы
  - в) моды
  - г) t-критерия
  - д) квотной выборки
45. Формула для расчета t-критерия представляет собой

- а) отношение медианы к моде
- б) произведение валидности и надежности
- в) разность сомножителей
- г) отношение сигнала к шуму
- д) сумму отклонений от среднего

46. Уровень альфа, равный 0,05, означает, что, даже если статистически значимого различия не существует, мы обнаруживаем статистически значимое различие между средними значениями

- а) сто раз из ста
- б) 99 раз из ста
- в) 95 раз из ста
- г) 65 раз из ста
- д) пять раз из ста

47. Зная альфа-уровень, количество степеней свободы и значение t-критерия, можно найти t-значение в стандартных таблицах значимости и определить – достаточно ли велико t-значение для того, чтобы считать его значимым. Если ответ «да», тогда делается вывод о том, что различие между средними двух групп

- а) велико
- б) мало
- в) положительно
- г) отрицательно
- д) существует

48. Полезность индикаторных (фиктивных) переменных объясняется тем, что при их помощи мы можем представить несколько групп испытуемых в виде

- а) гистограммы
- б) выборки
- в) колоколообразной кривой
- г) паттерна
- д) уравнения

49. Когда мы проводим среди данных линию – мы создаем

- а) ковариационную модель
- б) теоретическую модель
- в) алгебраическую модель
- г) общую линейную модель
- д) модель мысленного эксперимента

50. На степень ошибочности описания каждой точки на графике с помощью уравнения для прямой, указывает величина, которая называется

- а) интервал
- б) отклонение
- в) оценка
- г) ошибка
- д) наклон

51. Каким образом среднее значение отражает каждое значение в распределении

- а) отражает точно
  - б) отражает кратко
  - в) не отражает совсем
  - г) в принципе неважно
  - д) затрудняюсь ответить
52. В каком случае легче всего сделать вывод о том, что средние значения двух групп не одинаковы?
- а) при низкой вариабельности
  - б) при высокой вариабельности
  - в) при равной вариабельности
  - г) при нулевой вариабельности
  - д) при искаженной вариабельности
53. Межгрупповые различия более очевидны при слабом
- а) сигнале
  - б) шуме
  - в) гаме
  - г) кошмаре
  - д) валидности
54. Если уравнение  $y = mx + b$  переформулировать в виде  $y = b + mx$ , это свойство называется
- а) свойство ассоциативности
  - б) свойство валентности
  - в) свойство кратности
  - г) свойство делимости
  - д) свойство коммутативности
55. Эффект взаимодействия оказывается смежным с произведением
- а) независимых переменных
  - б) зависимых переменных
  - в) индикаторных (фиктивных) переменных
  - г) паттернов
  - д) коэффициентов наклона
56. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $\beta_0$  означает
- а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания
  - д) индикаторную (фиктивную) переменную
57. Ковариации соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $y_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
58. 95%-ный доверительный интервал равен коэффициенту плюс/минус



- а) одно значение стандартной ошибки
  - б) два значения стандартной ошибки
  - в) три значения стандартной ошибки
  - г) девять значений стандартной ошибки
  - д) среднее значение стандартной ошибки
59. Там, где ошибка измерения отсутствует, наклон линий регрессии влияет
- а) наблюдаемый паттерн
  - б) теоретический паттерн
  - в) психологический паттерн
  - г) ничто не влияет
  - д) затрудняюсь ответить
60. Когда ошибка измерения допущена до эксперимента, то ошибки добавляются
- а) влево
  - б) вправо
  - в) по вертикали
  - г) по горизонтали
  - д) по часовой стрелке
61. Чем больше неэквивалентность между группами, тем искажение окажется
- а) невыносимее
  - б) непоправимее
  - в) слабее
  - г) сильнее
  - д) не изменяется
62. Распределение до и после эксперимента может быть описано при помощи
- а) гистограммы
  - б) паттерна
  - в) колоколообразной кривой
  - г) многочлена
  - д) выборки
63. Основной экспериментальный эффект оценивается при помощи
- а) стандартной ошибки
  - б) стандартного отклонения
  - в) теоретического паттерна
  - г) коэффициента регрессии
  - д) коэффициента корреляции
64. Если был рассчитан многочлен четвертого порядка, и если было установлено, что коэффициенты  $X^4$  и  $X^4Z$  оказались незначимыми, то эти члены
- а) можно перемножить
  - б) можно поделить
  - в) можно суммировать
  - г) можно вычесть
  - д) можно сократить

65. В модели ковариационного анализа вытеснения точек регрессии переменная  $\beta_1$  выражает

- а) значение  $i$ -того члена
- б) преэкспериментальный коэффициент
- в) среднее различие в эксперименте
- г) индикаторную (фиктивную) переменную
- д) остаток для  $i$ -того члена после вычитания

## Вариант IV

1. Сформулируйте ответы на следующие вопросы:

1. Объясните различие между стандартной ошибкой и стандартным отклонением.
2. Что такое вероятностный выборочный метод? Для чего он применяется?
3. Как называется выборка, составленная на основе популяции, распределенной по широкому географическому региону?
4. Приведите примеры произвольных выборок.
5. Что включают в модальную выборку?
6. В чем состоят преимущества экспертной выборки?
7. Чему служит лицевая валидность?
8. Что называют валидностью конструкторов?
9. При помощи чего обычно оценивается та степень, в которой любые два измерения соотносятся друг с другом?
10. В чем состоит монооперациональный уклон и как его преодолеть?
11. Каким образом выявляются различные части матрицы мультисвойств-мультиметодов?
12. Что служит основой для генерирования паттернов?
13. Когда применяются принципы согласования паттернов?
14. В чем заключаются преимущества и недостатки согласования паттернов?
15. Почему важна теория истинного значения?
16. Каким образом можно убедиться в том, что наблюдение согласовано (или допускает возможность повторения)?
17. Что служит показателем совместной дисперсии двух измерений?
18. В каком интервале всегда находится надежность?
19. Какой тип надежности можно назвать «калибровкой» внешних наблюдателей? Приведите пример.
20. Сколько существует способов оценки внутренней согласованности измерения?
21. Каковы преимущества и недостатки каждого из оценочных показателей надежности?
22. Что такое номинальное измерение?
23. Что такое измерение отношений?
24. Какие методы одномерного шкалирования Вы знаете?

25. Какая размерность может быть у шкалы?
26. Как называется шкалирование по Гуттману?
27. Каковы преимущества применения многомерной модели?
28. Почему внутренняя валидность характеризуется «нулевой обобщаемостью»? Что это значит?
29. Что такое проблема «третьей переменной»?
30. Регрессия к среднему происходит перспективно или ретроактивно?
31. Что необходимо знать для того, чтобы вычислить регрессию к среднему? Приведите пример.
32. По какой формуле вычисляется процент регрессии к среднему?
33. Что называется угрозой валидности выводов? Какие угрозы валидности выводов Вы знаете?
34. Меньше какого значения не может быть статистическая мощность? Что это означает?
35. Какими способами можно улучшить надежность?
36. Какими способами можно улучшить надежность?
37. При увеличении мощности, возрастает ли вероятность обнаружения эффекта?
38. Для чего применяются дескриптивные статистики?
39. В каком виде можно представить распределение частоты?
40. Как рассчитать дисперсию?
41. Как найти степень свободы корреляции?
42. Что такое корреляционная матрица?
43. На какие два основных типа подразделяются планы исследований?
44. Что означает выражение «статистическое различие между группами по средним значениям»?
45. При помощи чего проверяют вычисленное значение t-критерия?
46. Какое значение альфа-уровня принято задавать в большинстве психологических исследований? Что оно означает?
47. С чем можно сравнить индикаторные (фиктивные) переменные?
48. Как рассчитывается различие между группами? Чему оно равно?
49. Как создать линейную модель?
50. Что показывает наклон линии регрессии?
51. Чему соответствует каждая точка на двумерном графике?
52. Что описывает линия двумерной регрессии?
53. С чем соотносится переменная, соответствующая программе либо терапии, при экспериментальном или при квазиэкспериментальном исследовании?
54. В каком алгебраическом виде записывается общая линейная модель?
55. Почему в случае низкой вариабельности легче всего сделать вывод о том, что различия между группами существуют?
56. Каким образом может наклон оказаться различием между средними?
57. Как определяется наклон?
58. Почему необходимо принимать во внимание вариабельность значений выборки?

59. Что представляет собой уравнение алгебраической модели постэкспериментального анализа?
60. Что учитывает регрессионная модель простого факторного плана  $2 \times 2$ ?
61. Что означают в алгебраической записи статистической модели рандомизированного анализа блоков значения бета ( $\beta$ )?
62. Из чего состоит применяемая в ковариационном анализе матрица данных?
63. Какая проблема возникает в связи с моделью ковариационного анализа (ANCOVA) при обращении к плану для неэквивалентных групп?
64. По каким причинам возникают искажения при применении ковариационного анализа (ANCOVA) к плану неэквивалентных групп?
65. Что влияет на наклон линий регрессии, когда ошибка измерения отсутствует?
66. Почему псевдоэффект не возникает при рандомизированном плане ковариационного анализа?
67. Верно ли, что чем больше неэквивалентность между группами, тем сильнее окажется искажение? Объясните, почему.
68. Можем ли мы увидеть ошибку измерения непосредственно из наших данных? Почему?
69. Что представляет собой надежность?
70. К чему должны приблизиться баллы каждого испытуемого?
71. Какой вид имеет после корректировки формула ковариационного анализа (ANCOVA)?
72. Когда можно быть уверенным в том, что действительно удалось обнаружить значительный эффект на основании данных, которые не содержат ошибки измерения?
73. Какие допущения следует принять для того, чтобы аналитическая модель считалась адекватной? Назовите и охарактеризуйте эти допущения.
74. В чем заключается основная проблема, связанная с анализом данных по плану перегиба линии регрессии?
75. Каким уравнением описывается избыточным образом конкретизированная модель?
76. На что следует обращать внимание при рассмотрении графика доэкспериментальной взаимосвязи?
77. Какой алгебраический вид имеет конкретизация полной квадратичной модели? Охарактеризуйте каждый компонент этого выражения.
78. До каких пор следует продолжать процедуру усовершенствования модели?
79. Какой алгебраический вид принимает модель ковариационного анализа (ANCOVA)? Охарактеризуйте каждый компонент этого уравнения.

2. Выполните следующие тестовые задания:

1. Если сформировать несколько выборок из одной и той же популяции, то во всех выборках статистическая оценка окажется
  - а) различной
  - б) сходной
  - в) близкой к нулю
  - г) равной нулю
  - д) не поддающейся расчетам
2. Распределение средних значений вокруг среднего от средних в выборочном распределении, называется
  - а) выборочным средним
  - б) средним арифметическим
  - в) стандартным средним
  - г) стандартной ошибкой
  - д) стандартным отклонением
3. Если отклониться от среднего вверх и вниз (т. е. влево и вправо) на три стандартных значения, тогда мы охватим
  - а) 65% всех случаев данного распределения
  - б) 25% всех случаев данного распределения
  - в) 95% всех случаев данного распределения
  - г) 99% всех случаев данного распределения
  - д) 50% всех случаев данного распределения
4. Комбинирование выборочных методов называется
  - а) кластерной выборкой
  - б) многоэтапной выборкой
  - в) экспертной выборкой
  - г) простой случайной выборкой
  - д) выборкой «снежный ком»
5. Различие между вероятностной выборкой и невероятностной выборкой заключается в том, что для вероятностной выборки необходим
  - а) случайный отбор
  - б) естественный отбор
  - в) мысленный эксперимент
  - г) нормальное распределение
  - д) алгебраическая модель
6. Теория относительно того, каким образом методы измерения в эксперименте соотносятся друг с другом и с другими теоретическими компонентами, называется
  - а) теоретическая валидность
  - б) наблюдаемая валидность
  - в) теоретический паттерн
  - г) наблюдаемый паттерн
  - д) выборочное распределение
7. То, насколько хорошо удалось транслирование конструкта в операционализацию, определяется
  - а) лицевой валидностью

- б) конвергентной валидностью
  - в) дискриминантной валидностью
  - г) прогностической валидностью
  - д) валидностью транслирования
8. Отвергает идею о том, что в качестве основы для определения конструкторов мы можем полагаться на операциональные формулировки
- а) фрейдизм
  - б) универсализм
  - в) дефиниционализм
  - г) реляционализм
  - д) операционализм
9. Если дискриминантная валидность присутствует, то взаимосвязь между измерениями двух различных конструкторов должна быть
- а) кратной 3,14
  - б) кратной целому числу
  - в) надежной
  - г) очень большой
  - д) очень малой
10. Относится к измерениям и к наблюдениям, но не относится к причинам
- а) монооперациональный уклон
  - б) мультиоперациональный уклон
  - в) монометодический уклон
  - г) мультиметодический уклон
  - д) сочетание обоих уклонов
11. В матрице мультисвойств-мультиметодов все корреляции, для которых использованы одинаковые методы измерения, образуют
- а) диагонали валидности
  - б) треугольники гетеросвойств-монометодов
  - в) треугольники гетеросвойств-гетерометодов
  - г) блоки монометодов
  - д) блоки гетерометодов
12. Матрица мультисвойств-мультиметодов требует, чтобы каждое свойство измерялось
- а) любым методом
  - б) каждым методом
  - в) новым методом
  - г) проверенным методом
  - д) передовым методом
13. Любое упорядочение объектов или единиц называется
- а) параметром
  - б) статистикой
  - в) выборкой
  - г) генеральной совокупностью
  - д) паттерном
14. Данные, применяемые для исследования теоретической модели, образуют

- а) наглядный паттерн
  - б) фактический паттерн
  - в) теоретический паттерн
  - г) наблюдаемый паттерн
  - д) психологический паттерн
15. Теория истинного значения выступает фундаментом
- а) теории эволюции
  - б) теории относительности
  - в) теории вероятности
  - г) теории надежности
  - д) теории личности
16. Случайная ошибка считается
- а) сигналом
  - б) кошмаром
  - в) шумом
  - г) гамом
  - д) равной нулю
17. Когда мы наблюдаем количественное значение того, что мы измеряем, мы обычно рассматриваем это значение состоящим
- а) из одного компонента
  - б) из двух компонентов
  - в) из многих компонентов
  - г) из сходных компонентов
  - д) из теоретических компонентов
18. В числителе формулы для оценки надежности находится
- а) оценка регрессии истинного значения
  - б) оценка дисперсии истинного значения
  - в) оценка корреляции истинного значения
  - г) оценка суммы истинного значения
  - д) оценка разности истинного значения
19. Надежность, равная 0,8, означает, что вариабельность характеризуется истинностью
- а) на 20%
  - б) на 40%
  - в) на 80%
  - г) на 100%
  - д) на 200%
20. Когда измерение является непрерывным, то для оценок внешних наблюдателей требуется рассчитывать
- а) ошибки
  - б) погрешности
  - в) отклонения
  - г) искажения
  - д) корреляции

21. Для оценки тест-ретестовой надежности следует предложить одному и тому же внешнему наблюдателю выставить кодированные оценки
- а) дважды
  - б) трижды
  - в) сто раз
  - г) ежедневно
  - д) не выставлять кодированные оценки
22. При каком измерении всегда существует абсолютный нуль
- а) номинальное измерение
  - б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
23. Суммирующим шкалированием называется
- а) шкалирование Тёрстоуна
  - б) шкалирование Лайкерта
  - в) шкалирование Гутмана
  - г) шкалирование с равными интервалами
  - д) кумулятивное шкалирование
24. Сколько размеров имеет конструкт под названием семантический дифференциал
- а) один размер
  - б) два размера
  - в) три размера
  - г) много размеров
  - д) столько, сколько испытуемых
25. Шкалу Лайкерта иногда называют «суммирующей» шкалой, потому что итоговый балл испытуемого по шкале представляет собой
- а) корреляцию оценок по всем пунктам шкалы
  - б) дисперсию оценок по всем пунктам шкалы
  - в) стандартную ошибку оценок по всем пунктам шкалы
  - г) стандартное отклонение оценок по всем пунктам шкалы
  - д) сумму оценок по всем пунктам шкалы
26. Когда деления на шкале направлены в обратную сторону от общего направления шкалы, то такая разновидность называется
- а) кумулятивной шкалой
  - б) реверсивной шкалой
  - в) интервальной шкалой
  - г) порядковой шкалой
  - д) номинальной шкалой
27. Чем характеризуется внутренняя валидность? характеризуется «нулевой обобщаемостью»
- а) 65% доверительным интервалом
  - б) 95% доверительным интервалом
  - в) 99% доверительным интервалом



- г) стопроцентной обобщаемостью
- д) нулевой обобщаемостью

28. Допустим, мы отобрали 10% из генеральной совокупности, исходя из тех баллов по тестированию, которые были получены до эксперимента. Какова вероятность того, что при тестировании после эксперимента нижние десять процентов снова окажутся состоящими из той же самой группы?

- а) 100%
- б) 99%
- в) 95%
- г) 65%
- д) 0%

29. Если выборка состоит из испытуемых с высокими баллами, то регрессия к среднему заставит их

- а) опуститься
- б) подняться
- в) замереть
- г) отжаться
- д) исчезнуть

30. Если по первому измерению выборка не сильно отличается от генеральной совокупности, то регрессия к среднему будет

- а) большой
- б) небольшой
- в) средней
- г) достаточной
- д) не возникнет

31. Насколько смещается выборочная группа, если корреляция  $r = 0,5$ ?

- а) на 100%
- б) на 80%
- в) на 50%
- г) на 20%
- д) на 0%

32. Степень обоснованности, достоверности либо вероятности тех выводов, к которым мы пришли относительно взаимосвязи в наших данных, называется

- а) валидность конструктов
- б) валидность выводов
- в) внешняя валидность
- г) внутренняя валидность
- д) конвергентная валидность

33. В психологических исследованиях эмпирически установлено, что значение статистической мощности должно быть

- а) не менее 1
- б) не менее 0,8
- в) не менее 0,2
- г) не более 1
- д) равно 0

34. Ошибка, которую мы допускаем, когда отвергаем нулевую гипотезу, поскольку статистика принимает значение, принадлежащее критической области, в то время, как нулевая гипотеза верна, называется
- а) роковая ошибка
  - б) простительная ошибка
  - в) ошибка первого типа
  - г) ошибка второго типа
  - д) медвежья услуга
35. Хорошая реализации исследования позволяет улучшить
- а) валидность
  - б) надежность
  - в) погрешность
  - г) вероятность
  - д) мощность
36. Вероятность отрицания эффекта, даже если он всё-таки есть, называется
- а) роковая ошибка
  - б) простительная ошибка
  - в) ошибка первого типа
  - г) ошибка второго типа
  - д) медвежья услуга
37.  $\alpha$  равная 0,01 означает, что вероятность утверждать, что различия отсутствуют, когда различия действительно реально отсутствуют, равна
- а) 0%
  - б) 1%
  - в) 50%
  - г) 99%
  - д) 100%
38. Судить о том, закономерны или случайны наблюдаемые различия между группами, нам позволяют
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
39. Оценка «центра» распределения значений, называется
- а) выраженная тенденция
  - б) сглаженная тенденция
  - в) оптимальная тенденция
  - г) центральная тенденция
  - д) современная тенденция
40. Распределение значений вокруг центральной тенденции называется
- а) дислексия
  - б) депрессия
  - в) дисперсия
  - г) регрессия

- д) прогрессия
41. Если среднее равно 20,8750, а стандартное отклонение составляет 7,0799, то в пределах между 6,7152 и 35,0348 окажутся примерно
- а) 100% значений
  - б) 99% значений
  - в) 95% значений
  - г) 50% значений
  - д) 0% значений
42. Рассчитав корреляцию, следует рассмотреть возможность того, что наблюдаемая корреляция может оказаться
- а) положительной
  - б) отрицательной
  - в) нулевой
  - г) достоверной
  - д) случайной
43. Индикаторной (фиктивной) переменной называется такая переменная, которая применяет дискретные значения
- а) 1 и 2
  - б) 0 и 2
  - в) 0 и 1
  - г) 0 и 10
  - д) 2 и 1
44. Выражение «статистическое различие между группами по средним значениям» означает, что две колоколообразные кривые
- а) накладываются друг на друга
  - б) не накладываются друг на друга
  - в) параллельны друг другу
  - г) пересекают ось  $x$
  - д) приближаются к оси  $y$
45. Когда необходимо учитывать различия между средними значениями относительно распределения или вариабельности баллов, уместно применение
- а) корреляционной матрицы
  - б) медианы
  - в) моды
  - г)  $t$ -критерия
  - д) квотной выборки
46. Если первое среднее будет больше, чем второе среднее, тогда  $t$ -критерий будет
- а) нулевым
  - б) рациональным
  - в) отрицательным
  - г) положительным
  - д) случайным
47. Название «степени свободы» означает число оцениваемых по данной выборке

- а) характеристик
  - б) ответов
  - в) статистик
  - г) параметров
  - д) габаритов
48. Несмотря на то, что индикаторные (фиктивные) переменные являются переменными номинального уровня (шкала наименований), но статистически с ними можно обращаться как с переменными
- а) нулевого уровня
  - б) порядкового уровня
  - в) интервального уровня
  - г) уровня отношений
  - д) в принципе любого уровня
49. Изменение показателей после эксперимента по отношению к тому, что было до эксперимента, выражается
- а) сдвигом линии
  - б) разворотом линии
  - в) наклоном линии
  - г) искривлением линии
  - д) колоколообразной кривой
50. Линию двумерной регрессии можно представить себе так же, как и любой другой
- а) коэффициент корреляции
  - б) коэффициент отклонения
  - в) статистический коэффициент
  - г) статистический вывод
  - д) дескриптивную статистику
51. Общая линейная модель представляет собой
- а) гистограмму
  - б) выборку
  - в) колоколообразную кривую
  - г) паттерн
  - д) уравнение
52. Исследователь отвечает за приведение общего уравнения к такому конкретному виду, который обеспечивает наилучшее
- а) упрощение
  - б) искажение
  - в) уточнение
  - г) обобщение
  - д) вычисление
53. При низкой вариабельности взаимное наложение двух колоколообразных кривых оказывается
- а) наибольшим
  - б) наименьшим
  - в) равным

- г) нулевым  
д) смещенным
54. Стандартная ошибка несет в себе информацию о
- а) выборочном среднем
  - б) среднем арифметическом
  - в) стандартном отклонении
  - г) выборочной совокупности
  - д) генеральной совокупности
55. Постэкспериментальное различие между средними для двух групп выражается
- а) значением для  $i$ -того члена
  - б) остатком для  $i$ -того члена после вычитания
  - в)  $i$ -тым членом экспериментальной группы
  - г)  $i$ -тым членом контрольной группы
  - д) коэффициентом наклона
56. Поскольку в общей модели имеются две индикаторные (фиктивные) переменные, у каждой из которых существуют два значения, то мы можем на основе данной общей модели составить
- а) 1 отдельное уравнение
  - б) 2 отдельных уравнения
  - в) 4 отдельных уравнения
  - г) много отдельных уравнения
  - д) ни одного отдельного уравнения
57. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $Z$  означает
- а) среднее различие в эксперименте
  - б) полученное значение для  $i$ -того члена
  - в) коэффициент для отрезка, отсекаемого на координатной оси
  - г) остаток для  $i$ -того члена после вычитания
  - д) индикаторную (фиктивную) переменную
58. Среднему различию в эксперименте соответствует
- а) переменная  $X_i$
  - б) переменная  $Z_i$
  - в) переменная  $u_i$
  - г) переменная  $e_i$
  - д) переменная  $\beta_2$
59. Если бы не было ошибки измерения вообще, то баллы нашего испытуемого попали бы прямо
- а) на линию регрессии
  - б) на стену
  - в) на потолок
  - г) в книгу рекордов Гиннеса
  - д) под кривую нормального распределения
60. О каком-либо эффекте можно говорить лишь в том случае, если между двумя линиями наблюдается

- а) горизонтальное смещение
  - б) вертикальное смещение
  - в) прямой угол
  - г) выравнивание
  - д) параллакс
61. Регрессионный анализ предполагает проведение линий в соответствии
- а) с вертикальным смещением
  - б) с горизонтальным смещением
  - в) с смещением влево
  - г) с смещением вправо
  - д) с смещением по часовой стрелке
62. Если ошибка в измерении до эксперимента отсутствует, тогда надежность измерения равна
- а) минус 1
  - б) ноль
  - в) плюс 1
  - г) 3,14
  - д) 95%
63. Отсутствие прыжка или перегиба кривой линии говорит нам о том, что экспериментальный эффект
- а) велик
  - б) мал
  - в) не велик и не мал
  - г) отсутствует
  - д) затрудняюсь ответить
64. Основной экспериментальный эффект оценивается при помощи
- а) стандартной ошибки
  - б) стандартного отклонения
  - в) теоретического паттерна
  - г) коэффициента регрессии
  - д) коэффициента корреляции
65. Если в точке отрезания имеется вертикальный перегиб, то он будет выражаться
- а) стандартной ошибкой
  - б) стандартным отклонением
  - в) теоретическим паттерном
  - г) коэффициентом регрессии
  - д) коэффициентом корреляции
66. В модели ковариационного анализа вытеснения точек регрессии переменная  $y_i$  выражает
- а) значение  $i$ -того члена
  - б) преэкспериментальный коэффициент
  - в) среднее различие в эксперименте
  - г) индикаторную (фиктивную) переменную
  - д) остаток для  $i$ -того члена после вычитания

## Вариант V

1. Сформулируйте ответы на следующие вопросы:

1. Почему чем больше объем выборки, тем меньше стандартная ошибка?
2. Перечислите характеристики простой случайной выборки.
3. Какова последовательность составления кластерной выборки?
4. Как называется комбинирование выборочных методов?
5. В чем состоит особенность целенаправленных выборок?
6. Какие типы квотных выборок вы знаете? В чем состоит различие между ними?
7. На что направлена содержательная валидность?
8. Что позволяет проверить критериальная валидность?
9. В чем состоит «дефиниционалистское» направление представления понятия конструктивной валидности?
10. В чем состоит главная проблема в связи с идеей конвергентности/дискриминантности?
11. В чем состоит монометодический уклон и как его преодолеть?
12. Каковы основные принципы интерпретации мультисвойств-мультиметодов?
13. Что всегда предполагает соответствие паттернов?
14. Что может включать в себя паттерн наблюдаемого результата?
15. Что такое случайная ошибка?
16. В каком виде можно себе представить надежность?
17. Как определить интервал оценки надежности?
18. Каким образом оценить надежность внешних наблюдений в тех случаях, когда измерение является непрерывным?
19. Что служит показателем совместной дисперсии двух измерений?
20. Сколько существует способов оценки внутренней согласованности измерения?
21. Обособлены ли друг от друга понятия надежности и валидности? В виде чего их можно рассматривать?
22. Что такое порядковое измерение?
23. На какие две широкие категории обычно подразделяются шкалы?
24. Сформулируйте определение шкалирования по С. Стивенсу. Что оно означает? Приведите пример.
25. Каковы цели шкалирования?
26. Приведите примеры одномерных конструктов.
27. С чего начинается разработка методов одномерного шкалирования?
28. Может ли быть так, что исследование обладает внутренней валидностью, но не обладает валидностью конструктов? Приведите пример.
29. Для чего необходимо «исключить» альтернативные достоверные объяснения? Как это сделать?

30. Регрессия направлена вверх или вниз?
31. Какая угроза внутренней валидности существует, когда план исследования включает в себя несколько групп?
32. Что такое проблема «третьей переменной»?
33. Какие факторы влияют на статистическую мощность?
34. Что означает 0,05 уровень значимости? Приведите пример.
35. Что такое альфа-уровень?
36. Что означает более низкий альфа-уровень?
37. Вероятность ошибки какого типа увеличивается при увеличении  $\alpha$ ?
38. В чем отличие дескриптивных статистик от статистических выводов?
39. Что такое центральная тенденция распределения?
40. Какие выводы можно сделать из предположения о нормальном распределении баллов?
41. Какие компоненты информации необходимы для проверки корреляции?
42. Как с помощью матрицы определить корреляцию для любой пары переменных?
43. В чем состоит отличие квазиэкспериментальных планов от экспериментальных планов?
44. Какой вид имеет формула для расчета t-критерия?
45. Как рассчитать стандартную ошибку?
46. Что нужно задать для проверки значимости t-критерия?
47. В чем состоит первый шаг представления большого объема информации с помощью индикаторных (фиктивных) переменных?
48. Какую последовательность действий следует выполнять для того, чтобы с помощью индикаторных (фиктивных) переменных отобразить уравнения для множества подгрупп?
49. К чему относится термин «линейная»?
50. Что описывает ошибка?
51. Что показывает линия двумерной регрессии?
52. Каким символом обозначаются в уравнении общей линейной модели индикаторные (фиктивные) переменные, которые кодируются символом  $x$ ?
53. Числовые соответствия каких значений следует знать для проверки конкретных исследовательских гипотез относительно взаимосвязей между переменными либо относительно различий между группами?
54. В чем отличие линии регрессии от полного распределения значений?
55. Назовите каждый способ определения терапевтического эффекта существует при анализе по постэкспериментальному рандомизированному плану. Что между ними общего?
56. Какой из трех способов является наиболее общим?
57. Сколько значений имеет переменная  $Z$ ? Какие это значения? Как называется такая переменная?
58. Как проще получить значение наклона  $\beta_1$ ?
59. О чем говорит положительное (отрицательное) значение наклона  $\beta_1$ ?
60. При помощи чего в регрессионной модели простого факторного плана  $2 \times 2$  выражаются основные эффекты?



61. Из чего в алгебраической записи статистической модели рандомизированного анализа блоков образуется матрица данных?
62. Что предполагает статистическая модель плана ковариационного анализа для одной ковариаты и одной контрольной и экспериментальной группы?
63. Какие переменные включает в себя уравнение модели ковариационного анализа (ANCOVA)?
64. Какой процедурой является регрессионный анализ? Поясните.
65. О чем говорит вертикальное смещение двух линий?
66. С каким измерением соотносится регрессионный анализ?
67. Почему возникает сглаживание линий?
68. К чему приводит сглаживание линий?
69. Каков возможный вариант решения проблемы, возникающей при плане исследования неэквивалентных групп?
70. Если ошибка измерения до эксперимента отсутствует, то чему равна надежность?
71. В чем её отличие от оригинальной формулы ковариационного анализа (ANCOVA)?
72. При помощи чего оценивается верхняя граница надежности?
73. Какие планы называются «пушистыми»?
74. Что происходит в случае, если данные криволинейны, а мы применяем к таким данным прямолинейную модель?
75. Что значит «недостаточная конкретизация» общей линейной модели?
76. Какой вид конкретизации более предпочтителен, если мы вынуждены конкретизировать модель неверным образом? Почему?
77. После того, как сделан выбор относительно варианта конкретизации модели, с чего начинается общая стратегия анализа?
78. В чём состоит консервативный способ определения того, каким образом следует усовершенствовать модель?
79. Что выражает коэффициент, ассоциированный в алгебраической модели ковариационного анализа (ANCOVA) с дихотомической экспериментальной переменной  $Z$ ?

## 2. Выполните следующие тестовые задания:

1. Если среднее значение для каждой выборки изобразить их в виде гистограммы, то окажется, что в основной массе
  - а) они равны нулю
  - б) они взаимно удалены
  - в) они приближаются к одному и тому же центральному значению
  - г) они кратны 3,14
  - д) они целые числа
2. Распределение бесконечного количества выборок одного и того же объема называют
  - а) матрица

- б) частное
  - в) многочлен
  - г) константа
  - д) выборочное распределение
3. Действительные значения параметров генеральной совокупности
- а) нам известны
  - б) нам неизвестны
  - в) приведены в справочнике
  - г) имеются в Интернете
  - д) затрудняюсь ответить
4. Если мы знаем средние значения выборочного распределения, и нам известна оценка стандартной ошибки, то мы можем определить
- а) ширину и высоту
  - б) угол наклона
  - в) стандартное отклонение
  - г) стандартную ошибку
  - д) доверительный интервал
5. Большая статистическая точность выборки означает, что
- а) дисперсия велика
  - б) дисперсия мала
  - в) дисперсия равна десятичному логарифму
  - г) дисперсия стремится к целому числу
  - д) дисперсия стремится к бесконечности
6. Разработать конкретные и осуществимые методы, позволяющие практически исследователям оценивать паттерны соответствия, т. е. оценивать конструктивную валидность своих исследований
- а) в принципе невозможно
  - б) в принципе возможно
  - в) в принципе согласен
  - г) в принципе не важно
  - д) затрудняюсь ответить
7. На проверку операционализации по отношению к релевантной содержательной области конструкта направлена
- а) содержательная валидность
  - б) валидность выводов
  - в) валидность конструктов
  - г) конвергентная валидность
  - д) дискриминантная валидность
8. Для того, чтобы оценить степень, в которой любые два измерения соотносятся друг с другом, обычно применяется
- а) стандартная ошибка
  - б) стандартное отклонение
  - в) теоретический паттерн
  - г) коэффициент дисперсии
  - д) коэффициент корреляции

9. Свидетельствует ли паттерн, подтверждающий и конвергентную валидность, и дискриминантную валидность, о том, что измерения данного конструкта действительно измеряют именно этот конструкт:
- а) разумеется, да
  - б) разумеется, нет
  - в) строгие правила отсутствуют
  - г) в принципе неважно
  - д) затрудняюсь ответить
10. Ограниченную обобщаемость конструктов можно назвать
- а) преднамеренными последствиями
  - б) непреднамеренными последствиями
  - в) неизбежными последствиями
  - г) необходимыми причинами
  - д) недостаточными причинами
11. Интерпретация матрицы мультисвойств-мультиметодов требует от исследователя способности
- а) способности фантазировать
  - б) способности измерять
  - в) способности рассуждать
  - г) способности вычислять
  - д) способности взаимодействовать
12. Для определения конвергентной валидности и дискриминантной валидности, фактор методов
- а) требуется
  - б) не требуется
  - в) отсутствует
  - г) допускается
  - д) в принципе неважно
13. Основой для генерирования паттернов служит
- а) практика
  - б) наблюдение
  - в) эксперимент
  - г) теория
  - д) измерение
14. Согласование паттернов позволяет рассматривать конвергенцию и дискриминацию
- а) в виде матрицы корреляций
  - б) в виде колоколообразной кривой
  - в) в виде гистограммы
  - г) в виде доверительного интервала
  - д) в виде континуума
15. Влияние любых факторов (например, настроения испытуемого) на измерение переменной в выборке, следует считать
- а) случайной ошибкой
  - б) простительной ошибкой

- в) систематической ошибкой  
г) уважительной причиной  
д) медвежьей услугой
16. В отличие от случайной ошибки, систематические ошибки имеют
- а) выраженную форму  
б) выраженную погрешность  
в) выраженную склонность  
г) выраженную тенденцию  
д) выраженный контур
17. Если наше измерение  $X$  надежно, то, если мы измерим или пронаблюдаем его дважды на одних и тех же испытуемых, то баллы в обоих случаях
- а) существенно искажутся  
б) существенно совпадут  
в) существенно отклонятся  
г) несущественно совпадут  
д) несущественно искажутся
18. В знаменателе формулы для оценки надежности находится
- а) регрессия измерения  
б) дисперсия измерения  
в) корреляция измерения  
г) сумма измерения  
д) разность измерения
19. При оценке степени, в которой согласуются между собой оценки одного и того же явления различными наблюдателями, применяется
- а) надежность внешних наблюдений  
б) тест-ретестовая надежность  
в) надежность параллельных форм  
г) внутренне согласованная надежность  
д) старый проверенный способ
20. Оценка тест-ретестовой надежности производится тогда, когда одна и та же тестовая методика применяется к одной и той же либо к аналогичной выборке
- а) не менее одного раза  
б) не менее двух раз  
в) не более одного раза  
г) не более двух раз  
д) тестовая методика вообще не применяется
21. Если тестовых заданий действительно много, то чаще всего для оценки внутренней согласованности используется
- а) тест-ретестовый оценочный показатель  
б) расколотые корреляции  
в) средняя корреляция по общему баллу  
г) средняя корреляция между пунктами методики  
д) альфа Кронбаха
22. При каком измерении играет роль дистанция между измерениями

- а) номинальное измерение
  - б) порядковое измерение
  - в) интервальное измерение
  - г) измерение отношений
  - д) в принципе любое измерение
23. Согласно определению С. Стивенса, шкалирование есть приписывание чисел объектам в соответствии
- а) со всеми правилами
  - б) с любыми правилами
  - в) с некоторыми правилами
  - г) с одним правилом
  - д) правил шкалирования не существует
24. Самая главная причина, по которой применяется одномерное шкалирование, заключается в том, что исследователь считает тот конструкт, который он измеряет
- а) надежным
  - б) валидным
  - в) основным
  - г) вспомогательным
  - д) одномерным
25. Когда деления на шкале направлены в обратную сторону от общего направления шкалы, то такая разновидность называется
- а) кумулятивной шкалой
  - б) реверсивной шкалой
  - в) интервальной шкалой
  - г) порядковой шкалой
  - д) номинальной шкалой
26. Когда судьи оценивают утверждения для формируемой шкалы, то согласие судей с оцениваемыми утверждениями
- а) желательно
  - б) обязательно
  - в) единодушное
  - г) ни о чём не говорит
  - д) не важно
27. Какая валидность подразумевает подтверждения того, что то, что было сделано в данном исследовании (т. е. программа) вызывает то, что наблюдается (т. е. результат)?
- а) валидность конструктов
  - б) валидность выводов
  - в) внешняя валидность
  - г) внутренняя валидность
  - д) конвергентная валидность
28. Какая валидность указывает на то, хотели ли вы выполнить именно ту программу, которую выполнили, либо на то, хотели ли вы наблюдать именно то, что в итоге наблюдается?

- а) валидность конструкторов
- б) валидность выводов
- в) внешняя валидность
- г) внутренняя валидность
- д) конвергентная валидность

29. Возможно ли на основании явления регрессии к среднему предвидеть то, каким образом сместятся баллы индивидуального испытуемого?

- а) всегда возможно
- б) в ряде случаев возможно
- в) возможно, если приобрести необходимый навык
- г) возможно в редких случаях
- д) никогда не возможно

30. Если по первому измерению выборка не сильно отличается от генеральной совокупности, то регрессия к среднему будет

- а) большой
- б) небольшой
- в) средней
- г) достаточной
- д) не возникнет

31. Насколько смещается выборочная группа, если корреляция  $r = 0,2$ ?

- а) на 100%
- б) на 80%
- в) на 50%
- г) на 20%
- д) на 0%

32. Все факторы, которые затрудняют поиск взаимосвязи, которую вы пытаетесь увидеть, называются

- а) сигналом
- б) кошмаром
- в) шумом
- г) гамом
- д) валидностью

33. Отношение сигнала о взаимосвязи к контекстуальному шуму называется

- а) размер сигнала
- б) размер шума
- в) размер взаимосвязи
- г) размер эффекта
- д) размер отношения

34. Что произойдет с сигналом, если сделать взаимосвязь более заметной?

- а) сигнал усилится
- б) сигнал ослабеет
- в) шум усилится
- г) шум станет невыносимым
- д) ничего не изменится

35. Вероятность того, что вы наблюдаете эффект тогда, когда он происходит, называется
- а) валидность
  - б) надежность
  - в) погрешность
  - г) вероятность
  - д) мощность
36. Более низкий  $\alpha$ -уровень означает, что тестирование проводится
- а) более строгим образом
  - б) менее строгим образом
  - в) более рациональным
  - г) менее рациональным
  - д) ничего не означает
37. Как влияет увеличение  $\alpha$  на вероятность ошибки второго типа (т. е. утверждения, что различия отсутствуют, когда в действительности различия существуют)?
- а) нормализует
  - б) увеличивает
  - в) уменьшает
  - г) усредняет
  - д) никак не влияет
38. Описывают то, что есть и то, что показывают данные
- а) коэффициенты корреляции
  - б) коэффициенты отклонения
  - в) статистические коэффициенты
  - г) статистические выводы
  - д) дескриптивные статистики
39. Балл, который находится точно посередине ряда чисел, называется
- а) среднее
  - б) мода
  - в) медиана
  - г) дисперсия
  - д) ранг
40. Максимальное значение минус минимальное значение называется
- а) промах
  - б) размах
  - в) замах
  - г) тах
  - д) min
41. Для того, чтобы рассчитать стандартное отклонение, сначала необходимо найти расстояние между каждым значением и
- а) модой
  - б) медианой
  - в) средним
  - г) дисперсией

- д) отклонением
42. Уровень значимости  $\alpha = 0,05$  означает, что мы проводим проверку того, что вероятность случайности полученной корреляции не превышает 5 из 100
- а) 99 из 100
  - б) 95 из 100
  - в) 50 из 100
  - г) 5 из 100
  - д) 1 из 100
43. Перемешивание, в результате которого множество объектов располагаются в случайном порядке, называется
- а) операционализация
  - б) структуризация
  - в) систематизация
  - г) реструктуризация
  - д) рандомизация
44. Выражение «статистическое различие между группами по средним значениям» означает
- а) высокую вариабельность
  - б) низкую вариабельность
  - в) приемлемую вариабельность
  - г) среднюю вариабельность
  - д) нулевую вариабельность
45. Когда необходимо учитывать различия между средними значениями относительно распределения или вариабельности баллов, уместно применение
- а) корреляционной матрицы
  - б) медианы
  - в) моды
  - г) t-критерия
  - д) квотной выборки
46. Если первое среднее меньше, чем второе среднее, тогда t-критерий будет
- а) нулевым
  - б) рациональным
  - в) отрицательным
  - г) положительным
  - д) случайным
47. Количество степеней свободы для t-критерия равно сумме человек в обеих группах
- а) плюс один
  - б) минус один
  - в) плюс два
  - г) минус два
  - д) плюс-минус один-два
48. Ключевым элементом простой регрессионной модели двухгруппового посттестового рандомизированного эксперимента  $y_i = \beta_0 + \beta_1 Z_i + e_i$  оказывается



- а) полученное значение для  $i$ -того члена
  - б) коэффициент для отрезка, отсекаемого на координатной оси
  - в)  $i$ -тый член экспериментальной группы
  - г)  $i$ -тый член контрольной группы
  - д) величина различия между экспериментальной и контрольной группами
49. Если бы уравнение описывало облако из точек совершенным образом, тогда бы каждая точка
- а) попала мимо линии
  - б) попала на линию
  - в) попала на окружность
  - г) попала на периметр
  - д) попала на колоколообразную кривую
50. Каким образом линия регрессии отражает каждое значение в двумерном распределении
- а) отражает точно
  - б) отражает неточно
  - в) не отражает совсем
  - г) в принципе неважно
  - д) затрудняюсь ответить
51. При экспериментальном или при квазиэкспериментальном исследовании, переменная, соответствующая программе либо терапии, соответствует
- а) независимой переменной
  - б) зависимой переменной
  - в) индикаторной (фиктивной) переменной
  - г) шуму
  - д) константе
52. Исследователь отвечает за приведение общего уравнения к такому конкретному виду, который обеспечивает наилучшее
- а) упрощение
  - б) искажение
  - в) уточнение
  - г) обобщение
  - д) вычисление
53. Небольшое различие между средними трудно обнаружить, если вариабельность
- а) велика
  - б) мала
  - в) положительна
  - г) отрицательна
  - д) существует
54. Описывает различия между группами относительно вариабельности всех значений в группах
- а) корреляционная матрица
  - б) медиана

- в) мода  
г) t-критерий  
д) квотная выборка
55. Если уравнение  $y = mx + b$  переформулировать в виде  $y = b + mx$ , это свойство называется
- а) свойство ассоциативности  
б) свойство валентности  
в) свойство кратности  
г) свойство делимости  
д) свойство коммутативности
56. Для того, чтобы увидеть различия между уровнями одного фактора, необходимо эти уравнения
- а) суммировать  
б) перемножить  
в) поделить друг на друга  
г) вычесть одно из другого  
д) приравнять к нулю
57. В алгебраической записи статистической модели рандомизированного анализа блоков для регрессионного анализа, символ  $u$  означает
- а) среднее различие в эксперименте  
б) полученное значение для  $i$ -того члена  
в) коэффициент для отрезка, отсекаемого на координатной оси  
г) остаток для  $i$ -того члена после вычитания  
д) индикаторную (фиктивную) переменную
58. Остатку для  $i$ -того члена после вычитания соответствует
- а) переменная  $X_i$   
б) переменная  $Z_i$   
в) переменная  $u_i$   
г) переменная  $e_i$   
д) переменная  $\beta_2$
59. Регрессионный анализ – это процедура
- а) с наименьшим количеством треугольников  
б) с наименьшим количеством квадратов  
в) с наименьшим количеством кружков  
г) с наименьшим количеством крестиков  
д) с наименьшим количеством галочек
60. Повлияет ли на наклон линий регрессии ошибка измерения, которая допущена после эксперимента?
- а) разумеется, да  
б) разумеется, нет  
в) строгие правила отсутствуют  
г) в принципе неважно  
д) затрудняюсь ответить
61. Если бы ошибки измерения не было, то линии регрессии
- а) были бы перпендикулярны друг другу

- б) были бы параллельны друг другу  
 в) вращались бы относительно друг друга  
 г) накладывались бы друг на друга  
 д) исчезали бы и появлялись друг за другом
62. Если в измерениях до эксперимента содержатся только исключительно ошибки, тогда надежность измерения равна  
 а) минус 1  
 б) ноль  
 в) плюс 1  
 г) 3,14  
 д) 95%
63. Включение в анализ ненужных членов приравнивается к усилению  
 а) сигнала  
 б) шума  
 в) гама  
 г) кошмара  
 д) наблюдаемого паттерна
64. Если избыточная оценка функции многочлена была выполнена корректно, в соответствии с моделью распределения данных, тогда оценка экспериментального эффекта окажется  
 а) существующей  
 б) несуществующей  
 в) искаженной  
 г) неискаженной  
 д) равной нулю
65. Если был рассчитан многочлен четвертого порядка, и если было установлено, что коэффициенты  $X^4$  и  $X^4Z$  оказались незначимыми, то эти члены  
 а) можно перемножить  
 б) можно поделить  
 в) можно суммировать  
 г) можно вычесть  
 д) можно сократить
66. В модели ковариационного анализа вытеснения точек регрессии переменная  $Z_i$  выражает  
 а) значение  $i$ -того члена  
 б) преэкспериментальный коэффициент  
 в) среднее различие в эксперименте  
 г) индикаторную (фиктивную) переменную  
 д) остаток для  $i$ -того члена после вычитания

### Ключи к тестовым заданиям:

Тема 1.

1.	Б	9.	Г	17.	Д	25.	В
----	---	----	---	-----	---	-----	---

2.	Г	10.	Г	18.	А	26.	В
3.	Г	11.	Г	19.	А		
4.	Б	12.	Д	20.	Б		
5.	В	13.	Г	21.	Б		
6.	Д	14.	А	22.	Б		
7.	В	15.	Г	23.	А		
8.	Д	16.	Б	24.	Г		

Тема 2.

1.	Д	13.	Г	25.	Б	37.	Г
2.	Д	14.	Г	26.	Д	38.	Д
3.	Г	15.	Д	27.	В	39.	Д
4.	В	16.	Б	28.	Г	40.	Г
5.	Б	17.	В	29.	Г	41.	В
6.	Д	18.	Б	30.	В	42.	Д
7.	Д	19.	Д	31.	Б	43.	В
8.	Д	20.	Б	32.	А	44.	Г
9.	А	21.	В	33.	В	45.	Д
10.	А	22.	В	34.	Б	46.	Д
11.	Д	23.	А	35.	Б		
12.	Г	24.	В	36.	Г		

Тема 3.

1.	Б	10.	Г	19.	Б	28.	В
2.	Б	11.	Б	20.	Б	29.	Д
3.	Б	12.	Д	21.	А	30.	Б
4.	Г	13.	Д	22.	Б	31.	А
5.	А	14.	Б	23.	Б	32.	Б
6.	Д	15.	Б	24.	В	33.	Г
7.	Б	16.	А	25.	А	34.	А
8.	Д	17.	В	26.	Б	35.	Д
9.	В	18.	Г	27.	Г	36.	В

Тема 4.

1.	Д	7.	Г	13.	Д	19.	Д
2.	А	8.	В	14.	В	20.	Б
3.	Б	9.	Б	15.	Д	21.	В
4.	Г	10.	В	16.	Б	22.	Г
5.	В	11.	Г	17.	Г	23.	Д
6.	Г	12.	Г	18.	А		

Тема 5.

1.	Г	7.	Г	13.	Б	19.	В
2.	Д	8.	Г	14.	А	20.	Б
3.	Г	9.	Д	15.	Б	21.	В
4.	Д	10.	Г	16.	Б		
5.	Г	11.	Г	17.	Г		
6.	А	12.	А	18.	А		

Тема 6.

1.	В	8.	Б	15.	Д	22.	Г
2.	Б	9.	В	16.	Г	23.	Г
3.	Г	10.	Г	17.	Г	24.	Б
4.	Б	11.	А	18.	В	25.	В
5.	В	12.	А	19.	Г	26.	В
6.	Г	13.	Б	20.	А	27.	Б
7.	Г	14.	А	21.	Д		

Тема 7.

1.	Б	6.	Д	11.	В	16.	Б
2.	В	7.	Д	12.	Б	17.	В
3.	Д	8.	В	13.	Г	18.	В
4.	Г	9.	Г	14.	В		
5.	Д	10.	В	15.	Б		

Тема 8.

1.	Д
2.	Б
3.	А
4.	Д
5.	Г
6.	В
7.	Г

Тема 9.

1.	Б	7.	Г	13.	Г	19.	Г
2.	Г	8.	Г	14.	Г	20.	Г
3.	Г	9.	Б	15.	В	21.	Д
4.	В	10.	А	16.	Д		
5.	Д	11.	Г	17.	Г		

6.	Д	12.	Г	18.	Д		
----	---	-----	---	-----	---	--	--

Тема 10.

1.	В
2.	В
3.	Д
4.	В
5.	Д
6.	Д

Тема 11.

1.	Г	6.	Г	11.	Б	16.	А
2.	Г	7.	Г	12.	Б	17.	Г
3.	Г	8.	Г	13.	Б	18.	Г
4.	В	9.	Д	14.	Д	19.	Г
5.	Б	10.	Б	15.	В		

Тема 12.

1.	Б	5.	А	9.	В	13.	Д
2.	В	6.	Б	10.	Г	14.	Д
3.	А	7.	А	11.	В		
4.	Б	8.	Б	12.	Г		

Тема 13.

1.	В
2.	В
3.	В
4.	В
5.	Г

Тема 14.

1.	В
2.	Г
3.	В
4.	Д
5.	Б

Тема 15.

1.	В
2.	Б
3.	А
4.	Д
5.	Г
6.	В

Тема 16.

1.	Б	7.	Б	13.	Г	19.	В
2.	Д	8.	Г	14.	А	20.	Б
3.	Б	9.	Б	15.	Г	21.	Д
4.	А	10.	Б	16.	Б	22.	Г
5.	Б	11.	В	17.	Г	23.	Д
6.	Б	12.	Д	18.	Г	24.	Д

Тема 17.

1.	Д	5.	Б	9.	Г	13.	Д
2.	Д	6.	Б	10.	Г		
3.	Г	7.	Г	11.	Г		
4.	Г	8.	Г	12.	Г		

Тема 18.

1.	В
2.	Д
3.	Б
4.	А
5.	Г

### Модуль 5.

Вариант 1

1.	Б	18.	А	35.	Г	52.	А
2.	Д	19.	Б	36.	Д	53.	Б
3.	Д	20.	А	37.	В	54.	Б
4.	А	21.	В	38.	Б	55.	В
5.	Г	22.	Д	39.	Д	56.	В
6.	Д	23.	Г	40.	В	57.	В

7.	Д	24.	Г	41.	Б	58.	В
8.	Г	25.	Б	42.	Д	59.	Б
9.	Б	26.	В	43.	Б	60.	Б
10.	Д	27.	Г	44.	Д	61.	В
11.	А	28.	А	45.	Г	62.	Б
12.	Д	29.	Г	46.	Д	63.	Д
13.	Г	30.	Б	47.	Д	64.	Д
14.	Б	31.	В	48.	В	65.	Б
15.	Д	32.	В	49.	Г	66.	Г
16.	Б	33.	Г	50.	Г	67.	В
17.	А	34.	А	51.	Б		

Вариант 2

1.	Г	17.	В	33.	А	49.	Г
2.	Д	18.	Б	34.	Г	50.	Б
3.	Г	19.	Г	35.	Г	51.	В
4.	А	20.	Б	36.	Б	52.	А
5.	В	21.	А	37.	В	53.	Г
6.	Д	22.	Г	38.	Д	54.	В
7.	Д	23.	Г	39.	Б	55.	Г
8.	Г	24.	Г	40.	В	56.	Б
9.	Б	25.	Г	41.	Б	57.	Д
10.	Д	26.	Д	42.	Г	58.	Б
11.	А	27.	Г	43.	Г	59.	Д
12.	Д	28.	А	44.	Г	60.	В
13.	Г	29.	Г	45.	Г	61.	Д
14.	Б	30.	В	46.	Д	62.	Г
15.	Б	31.	Б	47.	В	63.	Г
16.	Д	32.	Г	48.	Г	64.	Д

Вариант 3

1.	Г	18.	Г	35.	В	52.	А
2.	В	19.	Б	36.	Г	53.	Б
3.	А	20.	В	37.	Б	54.	Д
4.	Б	21.	Г	38.	Д	55.	В
5.	В	22.	Б	39.	В	56.	В
6.	Г	23.	В	40.	Г	57.	А
7.	Д	24.	Д	41.	В	58.	Б
8.	Г	25.	А	42.	А	59.	Г
9.	Б	26.	Д	43.	Г	60.	Г
10.	А	27.	Г	44.	Г	61.	Г



11.	Г	28.	Г	45.	Г	62.	Г
12.	В	29.	Г	46.	Д	63.	Г
13.	Д	30.	Б	47.	Д	64.	Д
14.	В	31.	А	48.	Д	65.	Б
15.	Б	32.	Г	49.	Г		
16.	Д	33.	Б	50.	Г		
17.	Д	34.	Б	51.	Б		

Вариант 4

1.	Б	18.	Б	35.	А	52.	Г
2.	Г	19.	В	36.	Г	53.	Б
3.	Г	20.	Д	37.	Г	54.	В
4.	Б	21.	А	38.	Г	55.	Д
5.	А	22.	Г	39.	Г	56.	В
6.	В	23.	Б	40.	В	57.	Д
7.	А	24.	В	41.	В	58.	Д
8.	Г	25.	Д	42.	Д	59.	А
9.	Д	26.	Б	43.	В	60.	Б
10.	В	27.	Д	44.	Б	61.	А
11.	Г	28.	Д	45.	Г	62.	В
12.	Б	29.	А	46.	Г	63.	Г
13.	Д	30.	Б	47.	Г	64.	Г
14.	Г	31.	В	48.	В	65.	Г
15.	Г	32.	Б	49.	В	66.	А
16.	В	33.	Б	50.	Д		
17.	Б	34.	В	51.	Д		

Вариант 5

1.	В	18.	Б	35.	Д	52.	Г
2.	Д	19.	А	36.	А	53.	А
3.	Б	20.	Б	37.	В	54.	Г
4.	Д	21.	Д	38.	Д	55.	Д
5.	Б	22.	В	39.	В	56.	Г
6.	Б	23.	В	40.	Б	57.	Б
7.	А	24.	Д	41.	В	58.	Г
8.	Д	25.	Б	42.	Г	59.	Б
9.	Б	26.	Д	43.	Д	60.	Б
10.	Б	27.	Г	44.	А	61.	Г
11.	В	28.	А	45.	Г	62.	Б
12.	Б	29.	Г	46.	В	63.	Б
13.	Г	30.	Б	47.	Г	64.	Г

14.	Д	31.	В	48.	Д	65.	Д
15.	А	32.	Г	49.	Б	66.	Г
16.	Г	33.	Г	50.	Б		
17.	Б	34.	А	51.	В		

Вопросы к экзамену по дисциплине  
«Математические методы в психологии»

1. Выборочное распределение
2. Выборочная ошибка
3. Правило 65%, 95% и 99%
4. Простая случайная выборка
5. Систематическая случайная выборка
- 6.** Кластерная случайная выборка
7. Многоэтапная выборка
8. Произвольные выборки
9. Преднамеренные (целенаправленные) выборки
10. Модальная выборка
11. Экспертная выборка
12. Квотная выборка
13. Гетерогенная выборка
14. Выборка «снежный ком»
15. Измерение различных типов валидности
16. Лицевая валидность
17. Содержательная валидность
18. Критериальная валидность
19. Прогностическая валидность
20. Одновременная валидность
21. Валидность конструкта
22. Конвергентная валидность
23. Дискриминантная валидность
24. Неадекватное дооперациональное объяснение конструктов
25. Монооперациональный уклон
26. Монометодический уклон
27. Взаимодействие различных экспериментальных воздействий
28. Взаимодействие тестирования и воздействия
29. Ограниченная обобщаемость конструктов
30. Неразделенность конструктов и уровней конструктов
31. «Социальные» угрозы валидности конструктов (формулирование гипотез, предвзятое оценивание, ожидания экспериментатора)
32. Номологическая сеть
33. Общая характеристика матрицы мультисвойств-мультиметодов
34. Составные части матрицы мультисвойств-мультиметодов

35. Основные принципы интерпретации матрицы мультисвойств-мультиметодов
36. Преимущества и недостатки матрицы мультисвойств-мультиметодов
37. Модифицированная матрица мультисвойств-мультиметодов
38. Согласование паттернов конструктивной валидности
39. Согласованность паттернов и валидность конструкторов
40. Преимущества и недостатки согласования паттернов
41. Теория истинного значения
42. Ошибка измерения
43. Уменьшение ошибки измерения
44. Теория надежности
45. Надежность внешних наблюдений
46. Тест-ретестовая надежность
47. Надежность параллельных форм
48. Средняя корреляция между пунктами методики
49. Средняя корреляция по общему баллу
50. Расколотая надежность
51. Альфа Кронбаха
52. Сравнение оценочных показателей надежности
53. Надежность и валидность
54. Уровни измерения
55. Определение шкалирования
56. Цели шкалирования
57. Метрика шкалирования
58. Шкалирование Тёрстоуна
59. Шкалирование Лайкерта
60. Шкалирование Гутмана
61. Внутренняя валидность
62. Установление причины и следствия
63. Нарушения внутренней валидности, связанные с единственной группой
64. Регрессия к среднему
65. Анализ данных психологического исследования
66. Валидность выводов
67. Угрозы валидности выводов (вывод о том, что взаимосвязь отсутствует, хотя в действительности она существует)
68. Угрозы валидности выводов (вывод о том, что взаимосвязь существует, хотя в действительности она отсутствует)
69. Улучшение валидности выводов
70. Статистическая мощность
71. Матрица принятия статистического решения
72. Подготовка данных исследования для анализа
73. Deskриптивные статистики
74. Одномерный анализ

75. Стандартное отклонение
76. Определение и расчет корреляции
77. Проверка значимости корреляции
78. Корреляционная матрица
79. Экспериментальный и квазиэкспериментальный анализы
80. t-критерий
81. Статистический анализ с применением t-критерия
82. Индикаторная (фиктивная) переменная
83. Общая линейная модель статистического анализа
84. Двумерная линейная модель статистического анализа
85. Постэкспериментальный анализ
86. Анализ по факторному плану
87. Рандомизированный анализ блоков
88. Анализ ковариаций
89. Анализ неэквивалентных групп (проблема)
90. Анализ неэквивалентных групп (решение проблемы)
91. Анализ перегиба линии регрессии (допущения при анализе)
92. Анализ перегиба линии регрессии (проблема криволинейности)
93. Анализ перегиба линии регрессии (конкретизация модели)
94. Анализ перегиба линии регрессии (анализ стратегии)
95. Анализ перегиба линии регрессии (этапы анализа)
96. Анализ вытеснения точек регрессии

Темы для контрольных работ по дисциплине  
«Математические методы в психологии»

1. Выборочное распределение
2. Простая случайная выборка
3. Невероятностная выборка
4. Валидность транслирования
5. Критериальная валидность
6. Конвергентная и дискриминантная валидность
7. Номологическая сеть
8. Матрица мультисвойств-мультиметодов
9. Согласование паттернов конструктивной валидности
10. Теория истинного значения
11. Ошибка измерения
12. Теория надежности
13. Типы надежности
14. Надежность и валидность
15. Уровни измерения
16. Общие вопросы шкалирования
17. Шкалирование Тёрстоуна
18. Шкалирование Лайкерта
19. Шкалирование Гуттмана

- 20.Регрессия к среднему
- 21.Улучшение валидности выводов
- 22.Статистическая мощность
- 23.Дескриптивные статистики
- 24.Расчет и проверка значимости корреляции
- 25.Статистический анализ с применением t-критерия
- 26.Индикаторные (фиктивные) переменные
- 27.Общая линейная модель
- 28.Анализ по факторному плану
- 29.Рандомизированный анализ блоков
- 30.Анализ ковариаций
- 31.Анализ неэквивалентных групп
- 32.Ковариационный анализ (ANCOVA) со скорректированной надежностью
- 33.Анализ перегиба линии регрессии
- 34.Проблема криволинейности
- 35.Конкретизация общей линейной модели
- 36.Анализ вытеснения точек регрессии

#### Список литературы:

1. Аренс Х., Лейтер Ю. Многомерный дисперсионный анализ. М., 1985.
2. Беллман Р. Введение в теорию матриц. М., 1969.
3. Битинас Б. Многомерный анализ в педагогике и педагогической психологии. Вильнюс, 1971.
4. Браунли К. А. Статистическая теория и методология в науке и технике. М., 1977.
5. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., 1976.
6. Грабарь М. Н., Краснянская К. А. Применение математической статистики в педагогических исследованиях. М., 1977.
7. Дейвисон М. Многомерное шкалирование. Методы наглядного представления данных. М., 1988.
8. Ермолаев О. Ю. Математическая статистика для психологов. М., 2003.
9. Иберла К. Факторный анализ. М., 1980.
10. Карасев А. И. Теория вероятностей и математическая статистика. М., 1977.
11. Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Р. Справочник по теории вероятностей и математической статистике. М., 1985.
12. Кричевец А. Д., Шикин Е. В., Дьячков А. Г. Математика для психологов. М., 2003.

13. Крылов В. Ю. Геометрическое представление данных в психологических исследованиях. М., 1990.
14. Лакин Г. Ф. Биометрия. М., 1990.
15. Логвиненко А. Д. Измерения в психологии. Математические основы. М., 1993.
16. Наследов А. Д. Математические методы психологического исследования: Анализ и интерпретация данных. Спб., 2004.
17. Образцов П. И. Методы и методология психолого-педагогического исследования. М., СПб., 2004.
18. Павлов Ю. В. Статистическая обработка результатов педагогического эксперимента. М., 1972.
19. Паповян С. С. Математические методы в социальной психологии. М., 1983.
20. Психология и математика. М., 1976.
21. Рубцова Н. Е., Ленков С. Л. Статистические методы в психологии. М., 2005.
22. Сочивко Д. В., Якунин В. А. Математические модели в психолого-педагогических исследованиях. Л., 1988.
23. Суходольский Г. В. Математическая психология. СПб., 1997.
24. Суходольский Г. В. Основы математической статистики для психологов. Л., 1972.
25. Тугушев Р. Х. Элементы математической статистики с критериями. Саратов, 2007.
26. Тугушев Р. Х. Системная персонология. Количественный и качественный анализ. Саратов, 1998.

Саратовский государственный университет имени Н.Г. Чернышевского