

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Саратовский государственный университет имени Н.Г.Чернышевского»

Балашовский институт (филиал)

**Методические указания для проведения практических занятий
по теории вероятностей и математической статистике
для направления 080100 Экономика**

Балашов 2012

УДК 519.2
ББК 22.17

Автор-составитель
И.Б. Волосенцева

Методические указания для проведения практических занятий по дисциплине «Теория вероятностей и математическая статистика» составлены в соответствии с учебной программой и предназначены для студентов 2 курса факультета математики, экономики и информатики направления 080100 «Экономика». В них представлены примеры задач из предметной области с подробным их решением.

Рекомендуется к опубликованию в электронной библиотеке кафедрой математики Балашовского института (филиала) Саратовского государственного университета имени Н. Г. Чернышевского.

Работа представлена в авторской редакции.

© Волосенцева И.Б., 2012

Содержание

I. Изучение дисперсии распределения с помощью статистического эксперимента.....	4
II. Проверка гипотез о законе распределения.....	5
III. Дисперсионный анализ. Сравнение параметров более чем двух случайных величин.....	7
IV. Корреляционный и регрессионный анализ. Выявление связи между величинами.....	9
V. Элементы регрессионного анализа.....	12
VI. Корреляционная модель.....	15
Список литературы.....	18
Приложение	19

I. Изучение дисперсии распределения с помощью статистического эксперимента

Проводя практическое занятие по данной теме, преподаватель должен уделить внимание вычислению по выборке точечной оценки для генеральной дисперсии σ^2 и среднеквадратического отклонения и доверительных интервалов для них. Необходимо уделить внимание проверке гипотезы о численном значении дисперсии. Студенты должны выработать в сознании определенный алгоритм действия по данной проблеме, покажем это на конкретном примере.

Задача 1. Выполнение норм выработки характеризуется таблицей:

Процент выполнения	90–100	100–110	110–120	120–130	130–140	Σ
Середина интервала	95	105	115	125	135	
Число рабочих	10	160	100	60	20	350

Найти смещенную и несмещенную оценку для дисперсии выполнения норм выработки и 95 процентный доверительный интервал для генерального среднеквадратического отклонения. Проверить гипотезу о том, что среднеквадратическое отклонение выполнения плана равно 10 (уровень значимости 0,05).

Проведем решение задачи по определенному алгоритму:

1) Выполняем точечную оценку для среднего по формулам:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \text{ — для выборки, заданной вариационным рядом; и}$$

$$\bar{x} = \frac{1}{n} \sum_{j=1}^k x_j m_j = \sum_{j=1}^k x_j \tilde{p}_j \text{ — для выборки, заданной таблицей группировки.}$$

Следовательно,

$$\bar{x} = \frac{1}{350} (95 \cdot 10 + 105 \cdot 160 + 115 \cdot 100 + 125 \cdot 60 + 135 \cdot 20) = \frac{39450}{350} = 112,71.$$

2) Вторым шагом вычисляем смещенную точечную оценку для дисперсии по формулам:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad - \quad \text{для выборки, заданной вариационным рядом; и}$$

$$S^2 = \frac{1}{n} \sum_{j=1}^n x_j^2 m_j - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \frac{m_j}{n} \quad - \quad \text{для выборки, заданной таблицей. Тогда } S^2 = 12796,428 - 12703,544 = 92,884.$$

3) Затем вычисляем несмещенную точечную оценку для дисперсии по формуле:

$$s^2 = \frac{n}{n-1} S^2, \quad \sigma = \sqrt{s^2} \quad - \quad \text{оценка для среднеквадратичного отклонения.}$$

$$\text{Имеем } s^2 = \frac{350}{349} 92,884 = 93,15, \quad \sigma = \sqrt{93,15} = 9,65.$$

4) Далее вычисляем границы генеральной дисперсии (используя таблицу 1 Приложения). В таблице 1 для $n = 349$ и $\beta = 0,9$ с помощью линейной интерпретации находим коэффициенты для вычисления верхней и нижней границы доверительного интервала:

$$k_1 = 0,660 + 349/(1000 - 39) \cdot (0,918 - 0,660) = 0,75$$

$$k_2 = 1,618 + 349/(1000 - 39) \cdot (1,094 - 1,618) = 1,43$$

$$0,75 \cdot 93,15 < \sigma^2 < 1,43 \cdot 93,15$$

$$8,36 < \sigma^2 < 11,53$$

5) Наконец, проверяем гипотезу о числовом значении дисперсии, так как значение 10 покрывается построенным доверительным интервалом, гипотеза о том, что $\sigma = 10$ не противоречит данным опыта.

II. Проверка гипотез о законе распределения

По второму вопросу необходимо провести проверку гипотез о законе распределения. Приведем пример.

Задача 2. Расход сырья на единицу продукции по старой технологии составил:

Расход сырья x_i	304	307	308	Σ
Число изделий m_i	1	4	4	9

А по новой технологии:

Расход сырья x_i	303	304	306	308	Σ
Число изделий m_i	2	6	4	1	13

Для проверки гипотезы о совпадении генеральных средних для старой и новой технологии вычислим эмпирические значения

$$\bar{x} = 307,11; \bar{y} = 304,77; S_x^2 = 2,378; S_y^2 = 1,555.$$

При уровне значимости 0,1 выполнить проверку о совпадении генеральных дисперсий (о том, что различие выборочных оценок дисперсий незначимо) для того, чтобы проверить гипотезу о совпадении средних по критерию Стьюдента.

Решение:

1) Вычисляем несмещенные оценки для генеральных дисперсий обеих совокупностей S_1^2 и S_2^2 и статистику для проверки гипотезы – их отношение $F = S_1^2/S_2^2$.

$$S_1^2 = S_x^2 = 2,378$$

$$S_2^2 = S_y^2 = 1,555$$

$$F = S_1^2/S_2^2 = 2,378/1,555 = 1,53$$

$$n_1 = 9 - 1 = 8; n_2 = 13 - 1 = 12$$

2) Затем выполняем проверку гипотезы. Критерии проверки для гипотезы H_0 , состоящей в том, что $\sigma_1^2 = \sigma_2^2$, по статистике $F = S_1^2/S_2^2$ для уровня значимости α :

$$H_1: \sigma^2 = \sigma_1^2 > \sigma_2^2 - F > F_{\alpha, n_1-1, n_2-1},$$

$$H_2: \sigma^2 = \sigma_1^2 \neq \sigma_2^2 - F > F_{\alpha/2, n_1-1, n_2-1},$$

$$\text{либо } F < F_{1-\alpha/2, n_1-1, n_2-1} \quad (F < \frac{1}{F_{\alpha/2, n_2-1, n_1-1}}).$$

Граница критической области F_{α, n_1-1, n_2-1} берется из таблицы критических точек распределения F (таблица 2 Приложения). Для отыскания значений $F_{1-\alpha/2, n_1-1, n_2-1}$ надо воспользоваться равенством:

для $p > 0,5$ $F_p(k_1, k_2) = \frac{1}{F_{1-p}(k_2, k_1)}$ (очевидно, если выборки поменять местами). Значит, в нашем случае $H_1: \sigma^2 = \sigma_1^2 \neq \sigma_2^2$, следовательно, надо выполнить проверки: $F > F_{\alpha/2, n_1-1, n_2-1}$ либо $F < F_{1-\alpha/2, n_1-1, n_2-1}$ ($F < \frac{1}{F_{\alpha/2, n_2-1, n_1-1}}$).

$$F_{\alpha/2, n_1-1, n_2-1} = 2,85,$$

$$F_{\alpha/2, n_2-1, n_1-1} = 3,28,$$

$$F_{1-\alpha/2, n_1-1, n_2-1} = 1/F_{\alpha/2, n_2-1, n_1-1} = 1/3,28 = 0,3,$$

значит, критическая область такова: $(0; 0,3); (2,58; \infty)$.

Значение статистики 1,53 не попало в критическую область, следовательно, гипотезу о равенстве дисперсии расхода сырья по старой и новой технологии принимаем и проверку о равенстве математических ожиданий проводить можно.

III. Дисперсионный анализ. Сравнение параметров более чем двух случайных величин

В третьем пункте знакомим студентов с дисперсионным анализом. Проверим статистические гипотезы о законе распределения с помощью критерия согласия Пирсона или критерия Колмогорова. Проверка статистических гипотез об однородности двух выборок с помощью критерия Колмогорова-Смирнова и гипотезы об однородности m выборок с помощью критерия χ^2 .

Покажем на примере.

Задача 3. При 120 подбрасываниях игральной кости единица выпала 25 раз, двойка 19 раз, тройка 15 раз, четверка 22 раза, пятерка 15 раз, шестерка 21 раз. Согласуется ли это с тем, что кость правильной формы?

Решение:

1) Сначала вычисляем значение статистики $U = \chi^2 = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}$, где n – число испытаний (используем критерий согласия Пирсона χ^2).

В дискретном случае генеральные частоты p_i – это вероятности значений x_i , частоты n_i – количество значений x_i , встретившихся в выборке. Если генеральное распределение непрерывное, то выборка группируется. Для интервалов группировки выборки (x_i, x_{i+1}) подсчитываются частоты n_i (число попаданий данных опыта в интервал) и теоретические вероятности p_i попадания в интервал (x_i, x_{i+1}) . Далее, для вычисления вероятности p_i надо воспользоваться таблицами генерального распределения $F(x)$, следовательно, $p_i = F(x_{i+1}) - F(x_i)$. Тогда в нашей задаче генеральное распределение – дискретное, находим относительные частоты, они известны:

$$p_i = \frac{1}{6} \quad (i = \overline{1, 6}), \quad n_1=25, \quad n_2=19, \quad n_3=15, \quad n_4=22, \quad n_5=15, \quad n_6=21, \quad n=120, \\ np_i=120/6=20.$$

Вычисляем значение статистики U :

$$U = \chi^2 = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i} = \frac{1}{20} [(25 - 20)^2 + (19 - 20)^2 + (15 - 20)^2 + \\ + 22 - 20 + 15 - 20 + 21 - 20] = \frac{1}{20} [25 + 1 + 25 + 4 + 25 + 1] = \frac{80}{20} = 4.$$

2) Далее для выбранного уровня значимости α по таблице χ^2 распределения (таблица 2 Приложения) находим критическое значение $\chi_{\alpha, k}^2$ для числа степеней свободы $k = m - r - 1$ (m – число интервалов группировки или принимаемых значений, r – число параметров генерального распределения, замененных оценками). Находим критическое значение $\chi_{\alpha, k}^2$,

$$k = m - r - 1 = 6 - 0 - 1 = 5, \quad \chi_{0,05; 5}^2 = 11,1.$$

3) Затем студенты должны сравнить U и $\chi_{\alpha, k}^2$. Если $U > \chi_{\alpha, k}^2$, то гипотеза H_0 отвергается; если же $U < \chi_{\alpha, k}^2$, гипотеза H_0 не противоречит опытным данным. Значит, студенты должны выяснить, попало ли значение статистики

в критическую область: $4 < 11,1$, следовательно, в критическую область значение статистики не попало и гипотеза о симметричности игральной кости согласуется с данными опыта на уровне значимости 0,05.

IV. Корреляционный и регрессионный анализ. Выявление связи между величинами

По четвертому вопросу необходимо отработать вопрос о проверке гипотезы о равенстве m математических ожиданий (элементы дисперсионного анализа).

Покажем на примере.

Задача 4. Произведено по 4 испытания на каждом из трех уровней фактора F . Методом дисперсионного анализа при уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытаний даны в таблице:

Номер испытания	F_1	F_2	F_3
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34
$\Sigma/4$	35	25	27

Решение:

Для упрощения расчетов можно прийти к данным $y_{ij} = 10^k x_{ij} - C$. Вычтем из всех значений $29 = (35 + 25 + 27)/3$. И предлагаю составить новую расчетную таблицу, дополняя столбцами, содержащими квадраты исправленных выборочных значений и строками для хранения сумм по строке и столбцу:

$i \setminus j$	F_1		F_2		F_3		Σ
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	
1	9	81	-9	81	-8	64	
2	7	49	-5	25	-7	49	
3	6	36	-3	9	2	4	
4	2	4	1	1	5	25	
$R_j = \sum_{i=1}^n y_{ij}^2$		170		116		142	$R_j=428$
$T_j = \sum_{i=1}^n y_{ij}$	24		-16		-8		$\Sigma T_j=0$
T_j^2	576		256		64		$\Sigma T_i^2=896$

1) Первым шагом предлагается выполнять проверку средствами дисперсионного анализа. В дисперсионном анализе для каждого из j ($j=1, 2, \dots, m$) уровней фактора проводится n испытаний. Пусть x_{ij} – значение измеряемого признака X , полученное в i -м опыте при j -м уровне фактора. Необходимо перейти от x_{ij} к $y_{ij} = 10^k x_{ij} - C$. Результаты будем располагать в двумерной таблице размерности $n \times m$. Положим, что данные заданы так, что i (номер строки) – это номер наблюдения, а j (номер столбца) – это номер уровня фактора. Для каждого уровня фактора j (столбца) вычисляем суммированием по столбцу:

$$R_j = \sum_{i=1}^n x_{ij}^2, \quad T_j = \sum_{i=1}^n x_{ij}, \text{ или}$$

$$R_j = \sum_{i=1}^n y_{ij}^2, \quad T_j = \sum_{i=1}^n y_{ij}.$$

Значит, число уровней фактора $m=3$, число испытаний на каждом уровне $n=4$.

Вычисляем значения y_{ij}^2 и помещаем в таблицу. Вычисляем значения R_j , T_j и T_j^2 и помещаем их в три последние строки таблицы.

2) Вторым шагом выполняем суммирование трех последних строк по строке. Результаты суммирования помещаем в последнем столбце трех последних строк.

3) Затем предлагается вычислить Q , Q_1 и Q_2 , опираясь на результаты суммирования последних трех строк по строке:

$$Q = \sum_{j=1}^m R_j - \frac{1}{mn} (\sum_{j=1}^m T_j)^2,$$

$$Q_1 = \frac{1}{n} \sum_{j=1}^m T_j^2 - \frac{1}{mn} (\sum_{j=1}^m T_j)^2 ,$$

$$Q_{\text{ост}} = Q_{\text{общ}} - Q_{\text{факт}} \quad \text{или} \quad Q_2 = Q - Q_1 .$$

Тогда число уровней фактора $m=3$, число испытаний на каждом уровне $n=4$.

$$Q_{\text{общ}} = \sum R_j - \frac{1}{3} \cdot 4 [\sum T_j]^2 = 428 - 0 = 428 ,$$

$$Q_{\text{факт}} = Q_1 = [\sum T_j^2]/4 - \frac{1}{3} \cdot 4 [\sum T_j]^2 = 896/4 - 0 = 224 ,$$

$$Q_{\text{ост}} = Q_2 = Q_{\text{общ}} - Q_{\text{факт}} = 428 - 224 = 204 .$$

4) Затем необходимо найти несмещенные оценки факторной и остаточной дисперсий. Для этого нужно разделить Q_1 и Q_2 на соответствующее число степеней свободы, значит:

$$S_1^2 = \frac{Q_1}{m-1} ; \quad S_2^2 = \frac{Q_2}{mn-m} .$$

5) Следующим шагом вычисляем значение статистики $F = S_1^2/S_2^2$, затем в таблице 2 Приложения критическое значение F_{α, k_1, k_2} . Тогда

$$F = \frac{S_1^2}{S_2^2} = \frac{112}{22,67} = 4,94 , \quad F_{0,05; 2; 9} = 4,26 .$$

6) $4,94 > 4,26$. Значение статистики попало в критическую область, значит, нулевую гипотезу о равенстве средних для разных уровней фактора отвергаем, следовательно, они различаются значимо.

7) Затем необходимо вычислить оценки для генерального среднего. Проводим оценки параметров:

$$\bar{y} = 0, \quad \bar{x} = 29$$

$$\lambda_1 = 35 - 29 = 6$$

$$\lambda_1 = 25 - 29 = -4$$

$$\lambda_1 = 27 - 29 = -2$$

$$S_{\text{ост}}^2 = 22,67 - \text{оценка для остаточной дисперсии}$$

8) Последним шагом вычисляем коэффициент детерминации – отношение факторной вариации к общей:

$$\rho^2 = Q_{\text{факт}}/Q_{\text{общ}} ,$$

$\rho^2 = 224/428 = 0,523$ (т.е. примерно половина (0,523 часть) дисперсии признака X объясняется зависимостью от фактора).

V. Элементы регрессионного анализа

В этом пункте плана необходимо познакомить с вычислением коэффициента корреляции, графическим изображением функциональной зависимости между величинами с помощью прямой, построенной методом наименьших квадратов.

Приведем пример:

Задача 5. Стоимость акций по годам некоторой фирмы указана в таблице:

Год t_i	1992	1993	1994	1995	1996
Стоимость y_i	9	9	11	10	12

Для выявления тенденции стоимости на ближайший год найти методом наименьших квадратов по этим данным прямую линию, вычислить прогнозируемое значение стоимости акций на 1997 г. и построить для этого значения доверительный интервал.

Решение:

1) Первым шагом необходимо предложить студентам составить таблицу для расчета. В ней должны быть столбцы: Δx_i , $y_i \Delta x_i$, $(\Delta x_i)^2$, y_i^2 , $ax_i + b$, $\Delta i = y_i - ax_i - b$, $(\Delta i)^2$ и 2 последние строки – для сумм по столбцу и этих сумм, поделенных на n . За начало отсчета возьмем 1994 г. и введем новую переменную $x = t - 1994$.

	Δx_i	y_i	$y_i \Delta x_i$	$(\Delta x_i)^2$	y_i^2	$ax_i + b$	Δi	$(\Delta i)^2$
	-2	9	-18	4	81	8,8	0,2	0,04
	-1	9	-9	1	81	9,5	-0,5	0,25
	0	11	0	0	121	10,2	0,8	0,64
	1	10	10	1	100	10,9	-0,9	0,81
	2	12	24	4	144	11,6	0,4	0,16
Σ	0	51	7	10	527			1,9
Σ/n	0	10,2	1,4	2	$105,4 - (10,1)^2 = 3,39$			

$$\Delta i = y_i - ax_i - b$$

2) Далее вычисляем коэффициенты прямой. Уравнение прямой ищем в виде $y = a(x - \bar{x}) + \bar{y}$, где

$$a = \frac{\frac{1}{n} \sum x_i y_i - \frac{1}{n} \sum x_i \cdot \frac{1}{n} \sum y_i}{\frac{1}{n} \sum x_i^2 - \left(\frac{1}{n} \sum x_i\right)^2} = \frac{\sum y_i \Delta x_i}{\sum (\Delta x_i)^2}, \quad \Delta x_i = x_i - \bar{x},$$

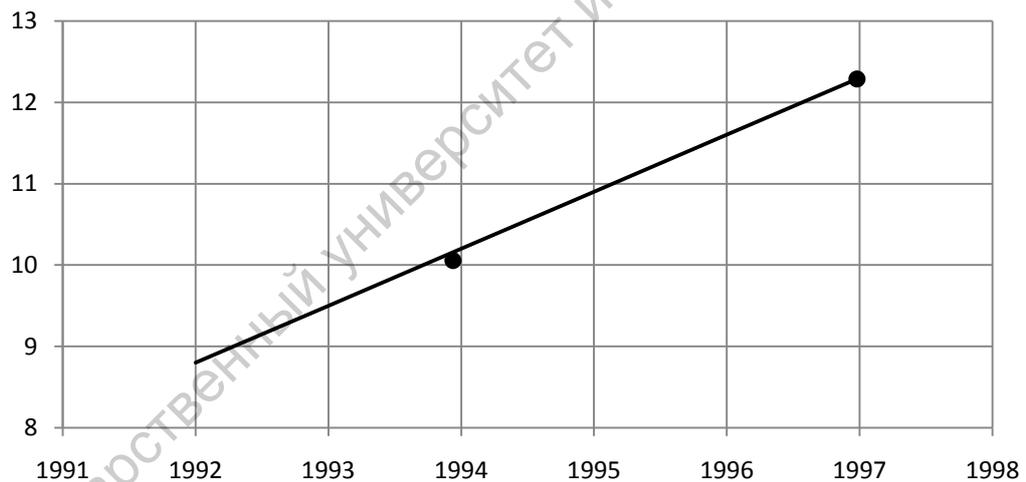
$$b = \bar{y} - a\bar{x}, \quad \bar{x} = \frac{1}{n} \sum x_i, \quad \bar{y} = \frac{1}{n} \sum y_i.$$

Значит $\bar{x} = 0$, $\bar{y} = 51/5 = 10,2$, $a = \sum y_i \Delta x_i / \sum (\Delta x_i)^2 = 7/10 = 0,7$.

Уравнение МНК-прямой: $y = 0,7x + 10,2$, где x измеряется в годах и отсчитывается от 1994 года.

3) Третьим шагом строим график прямой (прямую будем проводить через две точки $(\bar{x}; \bar{y})$ и $(x_i; ax + b)$ с любым x)

1997-у году соответствует $x = 3$. Проводим прямую через 2 точки: $(\bar{x}; \bar{y}) = (1994; 10,2)$ и $(1997; 12,3)$. Прогнозируем стоимость акций на 1997-й год $y = 0,7 \cdot 3 + 10,2 = 12,3$.



4) Далее вычисляем коэффициент корреляции r . Для системы точек, заданной таблицей из двух строк, вычислим по формулам:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i y_i) - \bar{x} \bar{y}}{\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2} \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2}} =$$

$$= \frac{\frac{1}{n} \sum_{i=1}^n y_i \Delta x_i}{s_x s_y}$$

Имеем $S_x^2 = \frac{10}{5} = 2$; $S_y^2 = 3,39$; $s_x = \sqrt{2} = 1,41$; $s_y = \sqrt{3,39} = 1,84$.

$$r = \frac{\frac{1}{n} \sum_{i=1}^n y_i \Delta x_i}{s_x s_y} = \frac{1,4}{1,41 \cdot 1,84} = 0,54$$

5) Далее по плану вычисляем прогнозируемое значение и доверительный интервал для него. Доверительный интервал для прогнозируемого значения вычисляем по формуле:

$$ax + b - t_{1-\alpha; n-2} \cdot S_{y(x)} < Y(x) < ax + b + t_{1-\alpha; n-2} \cdot S_{y(x)},$$

$$\text{где } S_{y(x)}^2 = S^2 \left[1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^n (x-\bar{x})^2} \right],$$

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (ax_i + b - y_i)^2 - \text{оценка дисперсии } \sigma^2.$$

Тогда прогнозируемое значение для 1997 г.:

$$Y(x) = 0,7 \cdot 3 + 10,2 = 12,3.$$

Оценка для дисперсии σ^2 :

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (\Delta_i)^2 = \frac{1}{3} \sum (y_i - ax_i - b)^2 = \frac{1}{3} \cdot 1,9 = 0,63,$$

$$S_{y(x)}^2 = 0,63 \left[1 + \frac{1}{5} + \frac{(3-3)}{10} \right] = 1,323,$$

$$S_{y(x)} = 1,15; \quad t_{1-\alpha; n-2} = 3,18 \quad (\alpha = 0,05; n = 5; n - 2 = 3)$$

$$0,7 \cdot 3 + 10,2 - 3,18 \cdot 1,15 = 12,3 - 3,657 = 8,64,$$

$$0,7 \cdot 3 + 10,2 + 3,18 \cdot 1,15 = 12,3 + 3,657 = 15,96.$$

Значит, 95-и процентный доверительный интервал для стоимости акций в 1997-м году:

$$8,64 < Y(1997) < 15,96.$$

б) Затем проверяем значимость уравнения регрессии.

$Q_R = \sum_{i=1}^n (ax_i + b - \bar{y})^2$ – сумма квадратов отклонений, обусловленных регрессией,

$Q_{\text{ост}} = \sum_{i=1}^n [y_i - (ax_i + b)]^2$ – остаточная сумма квадратов, характеризующая отклонение от среднего в точке x_i , т.е. влияние неучтенных факторов.

Уравнение регрессии значимо на уровне α , если на уровне α фактически наблюдаемое значение статистики $F = \frac{Q_R(n-m)}{Q_{\text{ост}}(m-1)} = \frac{Q_R(n-2)}{Q_{\text{ост}}(2-1)} > F_{\alpha; 1; n-2}$,

$F_{\alpha; 1; n-2}$ – табличное значение F -критерия Фишера-Снедекора, определенное на уровне значимости α , n – число наблюдений. Число m оцениваемых по выборке параметров в нашем случае равно двум (a и b). Значение F показывает, в какой мере регрессия лучше оценивает значение зависимой переменной по сравнению со средней.

Вычисляем Q_R :

$$Q_R = 1,4 \cdot 1,4 + 0,7 \cdot 0,7 + 0 + 0,7 \cdot 0,7 + 1,4 \cdot 1,4 = 4,9,$$

$$Q_{\text{ост}} = 1,9; \quad F = 4,9 \cdot 3 / 1,9 = 7,74,$$

$$F_{\alpha; 1; n-2} = 6,61.$$

Т.к. $F > F_{\alpha; 1; n-2}$ ($7,74 > 6,61$), уравнение регрессии значимо на уровне $\alpha = 0,05$.

VI. Корреляционная модель

Этот пункт плана показывает вычисление коэффициента корреляции, построение двух прямых регрессии.

Задача 6. Найти коэффициент линейной корреляции между признаками X и Y и написать уравнения прямых регрессии для корреляционной таблицы признаков:

$x y$	11	16	21	26	31	36	n_{j*}
25	2	4					6
35		6	3				9
45			6	45	4		55
55			2	8	6		16
65				4	7	3	14
n_{*j}	2	10	11	57	17	3	100

Решение:

1) Первым шагом рекомендуется вычислить (если они не заданы) в корреляционной таблице для каждого x_i значения $n_{i*} = \sum_{j=1}^m n_{ij}$ (сколько раз такое значение x_i встретилось в паре с любыми y_j) и для каждого

y_j значения $n_{*j} = \sum_{i=1}^n n_{ij}$. В нашей корреляционной таблице эти значения заданы.

2) Далее вычисляем:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_{i*}, \quad S_x^2 = \frac{\sum_{i=1}^k x_i^2 n_{i*}}{n} - (\bar{x})^2,$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^m y_j n_{*j}, \quad S_y^2 = \frac{\sum_{j=1}^m y_j^2 n_{*j}}{n} - (\bar{y})^2,$$

$$\mu = \overline{xy} - \bar{x}\bar{y} = \frac{\sum_{i=1}^k \sum_{j=1}^m x_i y_j n_{ij}}{n} - \bar{x}\bar{y},$$

$$r_{xy} = \frac{\mu}{S_x S_y};$$

значимость коэффициента корреляции проверяется с помощью статистики:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

$$\text{Тогда, } \bar{x} = \frac{1}{100} (25 \cdot 6 + 35 \cdot 9 + 45 \cdot 55 + 55 \cdot 16 + 65 \cdot 14) = 47,30;$$

$$S_x^2 = \frac{1}{100} (25 \cdot 25 \cdot 6 + 35 \cdot 35 \cdot 9 + 45 \cdot 45 \cdot 55 + 55 \cdot 55 \cdot 16 + 65 \cdot 65 \cdot 14 - 47,3 \cdot 47,3 \cdot 100) = 99,71; \quad S_x = 9,985;$$

$$\bar{y} = \frac{1}{100} (11 \cdot 2 + 16 \cdot 10 + 21 \cdot 11 + 26 \cdot 57 + 31 \cdot 17 + 36 \cdot 3) = 25,30;$$

$$S_y^2 = \frac{1}{100} (11 \cdot 11 \cdot 2 + 16 \cdot 16 \cdot 10 + 21 \cdot 21 \cdot 11 + 26 \cdot 26 \cdot 57 + 31 \cdot 31 \cdot 17 + 36 \cdot 36 \cdot 3 - 25,3 \cdot 25,3 \cdot 100) = 24,01; \quad S_y = 4,9;$$

$$\mu = \frac{1}{100} (25 \cdot 11 \cdot 2 + 25 \cdot 16 \cdot 4 + 35 \cdot 16 \cdot 6 + 35 \cdot 21 \cdot 3 + 45 \cdot 21 \cdot 6 + 45 \cdot 26 \cdot 45 + 45 \cdot 31 \cdot 4 + 55 \cdot 21 \cdot 2 + 55 \cdot 26 \cdot 8 + 55 \cdot 31 \cdot 6 + 65 \cdot 26 \cdot 4 + 65 \cdot 31 \cdot 7 + 65 \cdot 36 \cdot 3) - 47,3 \cdot 25,3 = 1234,80 - 1196,69 = 38,11;$$

$$r = \frac{38,11}{9,985 \cdot 4,9} = 0,7789.$$

3) Выписываем уравнения регрессии через коэффициенты a_{yx} и a_{xy} прямых регрессии, выписываем тангенсы угла наклона прямых регрессии по формулам:

$$y_x - \bar{y} = a_{yx}(x - \bar{x}) = \frac{\mu}{S_x^2}(x - \bar{x});$$

$$x_y - \bar{x} = a_{xy}(y - \bar{y}) = \frac{\mu}{S_y^2}(y - \bar{y});$$

$$k_1 = a_{yx}; \quad k_2 = \frac{1}{a_{xy}}.$$

Подставляем значения:

$$a_{yx} = \frac{38,11}{99,71} = 0,382;$$

$$a_{xy} = \frac{38,11}{24,01} = 1,587.$$

$$k_1 = 0,392, \quad k_2 = \frac{1}{1,63} = 0,613.$$

Прямые регрессии Y на X и X на Y :

$$y = 25,3 + 0,382(x - 47,30) = 0,382x + 7,23,$$

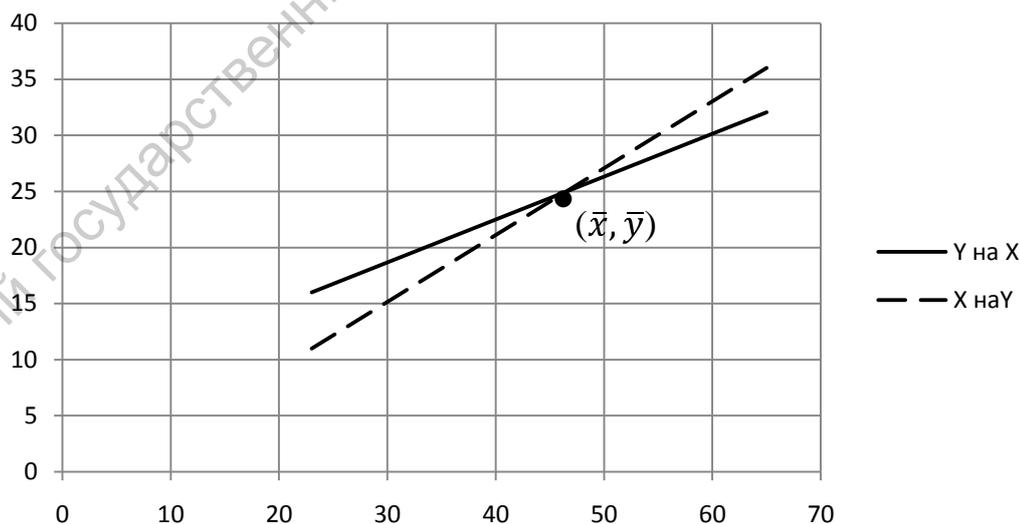
$$y(65) = 0,382 \cdot 65 + 7,23 = 32,06;$$

$$x = 47,30 + 1,587(y - 25,3) = 1,587y + 7,15,$$

$$x(21) = 1,587 \cdot 21 + 7,15 = 40,5,$$

$$x(16) = 1,587 \cdot 16 + 7,15 = 32,54.$$

4) Наконец, строим графики прямых регрессии (проводим прямые через точку (\bar{x}, \bar{y}) и еще одну точку, лежащую на прямой регрессии).



Список литературы

1. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] : учеб. пособие для вузов / В. Е. Гмурман. – 9-е изд. – М. : Высш. шк., 2003. – 479 с.
2. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике [Текст] : учеб. пособие для студентов вузов / В. Е. Гмурман. – 8-е изд., стер. – М. : Высш. шк., 2003. – 405 с.
3. Справочник по математике для экономистов [Электронный ресурс] : учеб. пособие / ред. В. И. Ермаков. – 3-е изд., перераб. и доп. – Электрон. дан. – М. : ИНФРА-М, 2009. – 464 с. – (Высшее образование)
4. Болотюк, В.А. Практикум и индивидуальные задания по курсу теории вероятностей (типовые расчеты) [Текст] : учебное пособие / В.А. Болотюк, Л.А. Болотюк, А.Г. Гринь, И.П. Гринь и др. – 1-е изд. – М.: Лань, 2010. – 288с.

Приложение

Таблица 1.

Границы доверительных интервалов для дисперсии в зависимости от числа степеней свободы и надежности.

Число степеней свободы	Надежность					
	0,90		0,95		0,99	
	Нижн. граница	Верх. граница	Нижн. граница	Верх. граница	Нижн. граница	Верх. граница
5	0,402	3,691	0,348	5,054	0,270	9,927
6	0,432	3,201	0,377	4,211	0,297	7,569
7	0,457	2,876	0,402	3,679	0,319	6,239
8	0,479	2,651	0,424	3,314	0,341	5,304
9	0,498	2,481	0,443	3,048	0,360	4,680
10	0,514	2,348	0,460	2,844	0,376	4,243
11	0,529	2,242	0,475	2,683	0,391	3,897
12	0,543	2,153	0,489	2,553	0,405	3,635
13	0,555	2,080	0,502	2,440	0,417	3,424
14	0,566	2,018	0,513	2,354	0,429	3,243
15	0,576	1,965	0,525	2,272	0,440	3,091
16	0,586	1,918	0,534	2,208	0,450	2,961
17	0,594	1,876	0,544	2,147	0,461	2,833
18	0,602	1,839	0,552	2,095	0,470	2,739
19	0,610	1,805	0,560	2,050	0,477	2,664
20	0,617	1,776	0,568	2,006	0,487	2,573
30	0,671	1,584	0,626	1,743	0,548	2,121
39	0,703	1,490	0,660	1,618	0,586	1,914
1000	0,931		0,918	1,094		

Таблица 2.

Критические значения для F -распределения для $\alpha=0,05$ (по вертикали k_2 – число степеней свободы знаменателя, по горизонтали k_1 – число степеней свободы числителя)

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,0	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Учебно-методическое издание

Автор-составитель
Волосенцева Ирина Борисовна

**Методические указания для проведения практических
занятий по теории вероятностей и математической статистике
для направления 080100 Экономика**