

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

кафедра социальной информатики

**ОСОБЕННОСТИ ОБРАБОТКИ ДАННЫХ
ГОСУДАРСТВЕННОЙ СТАТИСТИКИ С ПОМОЩЬЮ
ПРОГРАММЫ SPSS**

(автореферат бакалаврской работы)

студентки 5 курса 531 группы
направления 09.03.03 - Прикладная информатика
профиль Прикладная информатика в социологии
Социологического факультета
Видманкиной Ксении Александровны

Научный руководитель

кандидат социологических наук, доцент _____ С.В.Ситникова
подпись, дата

Зав. кафедрой

кандидат социологических наук, доцент _____ И.Г.Малинский
подпись, дата

Саратов 2018

Введение. Актуальность проблемы. Современная социологическая наука располагает значительным количеством эмпирической информации. Этому способствует не только неуклонный рост числа эмпирических социологических исследований, но и активное обращение социологов к другим источникам информации. В числе наиболее востребованных альтернативных источников информации находятся данные государственной статистики.

Привлекательность для исследователей статистических материалов, предоставляемых органами госстатистики, можно объяснить рядом моментов. Так, этот материал охватывает самые разные аспекты жизнедеятельности общества. Кроме того, статистический материал собирается и обобщается на протяжении длительного времени, что позволяет с его помощью отслеживать динамику социальных процессов. Наконец, он является доступным (по крайней мере, его часть) даже для рядовых исследователей. Вместе с тем, данные государственной статистики представляют собой значительный, но все же разрозненный объем информации. Несмотря на то, что интересующимся предлагаются не только простые (частоты и проценты), но и агрегированные показатели, обращение к статистическим материалам предполагает приложении серьезных аналитических усилий, глубокой статистической проработки информации. В связи с этим возрастает роль специальной аналитической работы.

Уже в 60-х годах для обозначения методов освоения результатов завершенных или ранее произведенных исследований в американской социологии был изобретен термин «вторичный анализ». Под ним понимают более углубленное изучение полученных данных с учетом новой исследовательской задачи или в связи с новыми данными.

Возможности, которые открывает вторичный анализ для эмпирических социологических исследований и фундаментальность методологических проблем социального познания, которые возникают с его появлением, свидетельствуют, что мы столкнулись не со случайным фактом в истории науки, а с фундаментальным явлением, которое ретроспективно

обнаруживается в прошлом и которое займет существенное место в будущем. Важно сегодня точно определить место и статус этого явления с тем, чтобы эффективно обозначить направления и границы его развития, в том числе применительно к анализу данных государственной статистики.

Степень научной разработанности. Характеризуя степень изученности данной проблемы, надо отметить, что термин «вторичный анализ», появившийся в обиходе западной социологической науки 30-40 лет назад, оказывается до сих пор мало распространенным в нашей стране и употребляется в отечественной литературе довольно редко и робко.

Всплеск интереса исследователей к данному проблемному полю пришелся на три хронологических периода. В 1975-1976 гг. к вопросу методологии вторичного анализа обращались такие исследователи, как В.Э. Шляпентох, О.М. Маслова, М.С. Мацковский, О.В. Ермакова и др. В 1984-1989 гг. разработкой данной проблемы занимались Г.И. Саганенко, М.М. Алхасов, А.О. Крыштановский и др. В 1995-1996 годы работы по вторичному анализу были написаны Г.Г. Татаровой, В.Б. Якубовичем, В.С. Сычевой, О.В. Ершовой, И.М. Клямкиным и др. В остальные годы данная тема практически не затрагивалась. И это наблюдается одновременно с широким распространением самих вторичных исследований.

В последнее время происходит медленное возрождение исследовательского интереса к методологическим аспектам вторичного анализа, и среди исследователей, занимающихся данной проблемой, можно назвать Сазонова Б.В., Коржеву Э.М., Жукову Т.И., Горяинова В.П., Погорецкого В.Г., Гиенко В.В. и др., что еще раз подтверждает актуальность выбранной темы.

Анализ статистических данных, предоставляемых органами государственной статистики, с точки зрения специфики их использования в рамках эмпирических социологических исследований реализуется, в основном, узкими специалистами, занимающимися количественными социологическими исследованиями и анализом их данных. Так, в рамках V Социологической

Грушинской конференции «Большая социология: Расширение пространства данных» на секции «Статистические данные в социологии» состоялось обсуждение теоретико-методологической проблемы использования статистических данных, предоставляемых государственной и ведомственной статистикой, в социологических исследовательских целях, в том числе были представлены практические результаты подобных попыток

Объектом исследования выступает метод вторичного анализа, **предметом** – применение метода вторичного анализа для анализа баз данных государственной статистики.

Цель работы – обосновать эвристический потенциал вторичного анализа данных государственной статистики для социологического исследования с использованием программы SPSS. Для достижения цели работы были поставлены следующие **задачи**:

1. рассмотреть базы данных государственной статистики как источник социальной информации;
2. охарактеризовать вторичный анализ данных как особый метод проведения социологического исследования, выявить его специфику;
3. применить метод вторичного анализа статистических данных для разработки социальной карты Саратовской области.

В качестве **эмпирической базы** были использованы данные государственной статистики по Саратовскому региону¹. Эмпирическая база данных создана в программе статистической обработки информации SPSS.

Научная новизна заключается в обосновании анализа данных государственной статистики как формы вторичного анализа.

Структура бакалаврской работы. Работа состоит из введения, трех разделов, заключения, списка использованных источников и приложений.

Основное содержание работы. В первом разделе «Данные государственной статистики как объект социологической интерпретации» рассматриваются

¹Города и районы Саратовской области в 2016 году: Статистический сборник / Территориальный орган Федеральной службы государственной статистики по Саратовской области. Саратов, 2017. 217 с.

основные принципы анализа статистической информации с точки зрения социологического исследования, дается краткая история развития статистики в нашей стране, а также характеризуются организационные и структурные особенности сбора статистических данных.

В западной социологии давно укоренена традиция использования различных статистических данных как самостоятельных источников информации об обществе. В российской социологии это направление пока представлено слабо, хотя потенциал у него, безусловно, имеется. Спецификой отечественной государственной статистики является сбор огромного массива данных, к которым исследователи не всегда имеют доступ. Но даже когда доступ есть, часто они неправильно трактуют эти данные. Основная сложность для социологов связана с тем, что наименования статистических показателей часто совпадают со словами повседневной речи, но имеют иные смыслы, а потому неверно интерпретируются. Кроме того, статистические данные довольно редко собираются по методикам, разработанным совместно с социологами. Как правило, в государственной или корпоративной статистике сбор данных направлен на достижение внутрикорпоративных целей.

Для обеспечения более точного соотношения статистических и социологических данных К.Д. Титаев предлагает использовать особую исследовательскую стратегию: составить их каталог, проинтерпретировать их, т.е. придать им социологические смыслы, на основе этих смыслов осуществить концептуализацию каждого статистического показателя. И только после завершения данного этапа можно ставить задачи и задавать вопросы о том, как эти статистические показатели соотносятся между собой и что говорят о социальной реальности.

Становление государственной статистики как инструмента познания общества произошло в конце XVIII века и было связано с оживлением свободного предпринимательства. Среди основных направлений развития государственной статистики на данном этапе можно отметить организационно-методологическое совершенствование работы органов государственной

статистики и появление первых масштабных статистических обследований, результаты которых впоследствии легли в основу реформ 1861г.

Пореформенный период (1861-1917) вошел в историю как период земской статистики. Огромный статистический материал, собранный и разработанный земскими статистиками, стал надежной основой глубоких исследований экономики пореформенной России, в первую очередь, русской деревни. Русской статистической науке этого периода была присуща глубокая теоретическая аргументация, ей принадлежала

В начальный этап советской статистики (1917 - 1930 гг.) проводилось большое число специально организованных статистических переписей и обследований, плодотворно работали различные научные коллективы, был построен первый баланс народного хозяйства, совершенствовалась организационная структура государственной статистики. В целом этот период выделяют плюрализм мнений, концепций, идей, критическое использование достижений зарубежной статистики. Дальнейшее развитие советской статистики было осложнено созданием в 30-е годы административно-бюрократической системы, массовыми репрессиями, коснувшимися выдающихся экономистов и статистиков. В период Великой Отечественной войны статистика решала важнейшие задачи по оперативному учету материальных и трудовых ресурсов, переводу производительных сил страны в восточные территории. В послевоенное время статистика стала играть еще более видную роль: развернулись балансовые работы, углубилась теория индексного метода, получили распространение экономико-математические модели и методы.

Современная российская государственная статистика столкнулась с необходимостью адаптации к новым условиям функционирования, связанным с изменениями политической и социально-экономической обстановки в стране и на международной арене, в связи с чем занимается решением текущих задач обеспечения информационных потребностей общества и параллельно разрабатывает комплекс мероприятий, связанных с адаптацией к новым

условиям существования. В настоящее время все информационные ресурсы Росстата можно разделить на два типа: информационные системы Росстата и информационные ресурсы с размещенными наборами открытых данных. В свою очередь, первые подразделяются на Центральную базу статистических данных и Банк готовых документов.

Второй раздел «Анализ данных государственной статистики как разновидность вторичного анализа данных» посвящен характеристике метода вторичного анализа данных, в том числе в случае работы со статистическими данными.

В России первые осознанные упоминания вторичного анализа как методологического термина относятся ко второй половине 70-х и началу 80-х годов. Этому способствовал целый ряд серьезных причин: само существование эмпирических данных; появление их архивов; рост числа организаций, регулярно проводящих масштабные первичные исследования и получающих огромные массивы данных; дороговизна проведения эмпирических исследований; появление крупных институциональных заказчиков; в образовательных организациях обращение к самым разным массивам данных открывает принципиально новые возможности для подготовки профессионалов в области общественных наук. И конечно, появление новых технических средств хранения, обработки, передачи и накопления информации - ЭВМ и Интернет.

Вторичный анализ в социологии – это стратегия и метод социологического исследования, когда анализу подвергаются существующие социологические данные для получения знаний в соответствии с новыми целями и задачами, не стоявшими при сборе этих данных. В отличие от базового исследования вторичный анализ не имеет этапов разработки вопросников, пилотажа, полевого опроса и сборов документов опроса.

Ключевым критерием, позволяющим рассматривать анализ статистических данных в качестве вторичного анализа, является глубокая математическая обработка информации. Три оставшиеся характеристики

вторичного анализа – отсутствие специально организованного социологического исследования, постановка новых исследовательских целей и задач и смена авторства – присущи анализу материалов госстатистики имманентно.

В третьем разделе «Социальное картографирование Саратовской области с опорой на статистические данные» описывается пример вторичного анализа данных государственной статистики, целью данного исследования было создание социальной карты Саратовской области.

Заключение. Статистические данные, предоставляемые государственными органами, часто используются учеными в исследовательских целях. Тем более, что спецификой отечественной системы государственного и ведомственного статистического наблюдения является сбор просто огромного объема данных, о которых исследователи часто не подозревают. А если даже и знают, то не имеют к ним доступа, т.к. для этого надо уметь писать правильные бюрократические запросы. В случае же когда доступ к информации организован, исследователи часто неправильно трактуют эти данные, поскольку статистические данные создаются и организуются по особым принципам, отличающимся от социологических.

Для обеспечения более точного соотношения статистических и социологических данных необходимо использовать особую исследовательскую стратегию: надо идти от возможностей самих статистических данных, а именно составить их каталог, проинтерпретировать, осуществить концептуализацию каждого статистического показателя, и лишь затем ставить исследовательские задачи. Другими словами, анализ статистических данных в рамках социологических исследований является очень перспективным направлением, требующим выработку кардинально отличной стратегии работы с информацией, по сравнению с собственно социологической. Эффективное использование статданных также требует серьезного обращения к математическому аналитическому аппарату, в том числе с помощью специализированных программ статистической обработки данных типа SPSS.

Это позволяет рассматривать анализ статданных как разновидность вторичного анализа.

Под *вторичным анализом* понимается совокупность методов и приемов получения нового знания, характеризующихся следующими признаками:

1. это методы, которые используются в случае, когда исследователи отказываются от проведения специально организованного эмпирического исследования, сбора нового эмпирического материала и довольствуются старой информацией из ранее проведенных исследований;

2. в отличие от первичного анализа, вторичный аналитик ставит перед собой новые исследовательские цели и задачи, не стоявшие при сборе первичных данных;

3. основным этапом, отличающим первичный анализ от вторичного, является этап формирования «информационного поля» на базе первичной информации. Именно на этом этапе проявляется специфика проводимого вторичного анализа, дающая основание отнести его к тому или иному типу;

4. и наконец, как правило, постановка новых исследовательских задач осуществляется другими авторами.

Анализ статистических данных можно рассматривать как разновидность вторичного анализа при выполнении определенного условия –обязательной глубокой математической обработки информации. Иногда постановка той или иной социологической проблемы может потребовать провести преобразование имеющейся статистической информации, другими словами, работать с ней так, как если бы она была получена в результате проведения социологического исследования, в том числе с помощью специализированных программ статистической обработки данных. В этом случае исследователь имеет дело со вторичным анализом. Дополнительные характеристики вторичного анализа – отсутствие специально организованного социологического исследования, постановка новых исследовательских целей и задач и смена авторства – и так присущи анализу материалов госстатистики.

В качестве иллюстрации возможностей вторичного анализа статистических данных было проведено небольшое авторское исследование, посвященное *созданию социальной карты Саратовской области*. Исследование включало 3 этапа:

1. Сокращение переменных с помощью факторного анализа: число переменных было уменьшено с 19 до 4, проинтерпретированных как «Население», «Убыль населения», «Здравоохранение» и «Воспроизводство населения»;

2. Объединение отдельных районов Саратовской области в кластеры в зависимости от выраженности тех или иных статистических показателей с помощью иерархического кластерного анализа случаев: всего было выделено 6 кластеров;

3. Характеристика кластеров с помощью создания таблиц сопряженности с кластерной переменной, расчета коэффициентов корреляции и сравнения средних. Проведенный анализ позволил выявить следующие характеристики групп районов:

1 кластер, включающий Энгельсский и Балаковский районы, отличается наличием самых крупных (помимо Саратова) городов области. Районы выделяются высокой плотностью проживающего на их территории населения, средним уровнем воспроизводства населения при низком коэффициенте смертности, высокими значениями коэффициента миграции. Актуальны проблемы аварийного жилья, переполненности детских садов и высокий уровень преступности. Тем не менее, районы относительно благополучны в экономическом плане. В инфраструктурном плане районы также выглядят развитыми.

2 кластер, состоящий из Питерского, Хвалынского, Аркадакского, Турковского, Духовницкого, Дергачевского, Базарно-Карабулакского, Балтайского и Ивантеевского районов, характеризуется средней плотностью заселения территорий. Отмечаются высокие значения коэффициентов смертности и младенческой смертности при низком коэффициенте

миграционного прироста населения. Естественный прирост населения на среднем уровне. Низкая доля детей среди населения. С экономической точки зрения районы данного кластера находятся в упадке. Инфраструктура также не развита. Среди позитивных сторон можно отметить незначительную долю семей, проживающих в аварийном жилье, низкий уровень преступности и средний уровень медицинской обеспеченности.

3 кластер - Красноармейский, Краснопартизанский, Романовский, Федоровский, Лысогорский и Самойловский районы, выделяется самыми низкими показателями по заселенности территории и низким уровнем воспроизводства и естественного прироста населения. Средний коэффициент смертности и низкий коэффициент младенческой смертности, что можно связать с развитостью здравоохранения. В экономике районов отмечаются негативные тенденции. Инфраструктура не развита. Из положительных тенденций надо отметить низкий уровень преступности и незначительное число семей, нуждающихся в улучшении жилищных условий при (неожиданно!) очень высоком темпе получения нового жилья в отчетном году.

4 кластер - Саратовский, Татищевский, Новобурасский, Ровенский, Советский и Воскресенский районы, характеризуется средней плотностью заселения территорий, средним же уровнем воспроизводства населения, низким уровнем смертности, средним уровнем естественного прироста населения при высоких темпах иммиграции.. Однако в экономическом плане негативные тенденции. С точки зрения инфраструктурной оснащенности можно отметить небольшое число библиотек и учреждений культурно-досугового типа, которые дополняются крайне тяжелой ситуацией в сфере здравоохранения. Среди положительных моментов можно перечислить низкую долю нуждающихся в переселении в новое жилье и низкий уровень преступности.

5 кластер –это Марксовский, Пугачевский, Аткарский, Петровский, Ртищевский, Балашовский и Вольский районы, центрами которых выступают малые города. Территория районов плотно занята селами и поселками, что формирует среднюю плотность заселения. Воспроизводство населения на очень

низком уровне, смертность, естественный и миграционный прирост – на среднем. Большое число семей проживает в аварийном жилье, при этом темпы расселения являются низкими. Экономически районы не благополучны. Среди положительных моментов развитость инфраструктуры и относительно низкий уровень преступности.

6 кластер –Александрово-Гайский, Ершовский, Новоузенский, Екатериновский, Калининский, Краснокутский и Перелюбский районы, не имеют городов и отличаются низкой плотностью заселения. Высокий уровень воспроизводства населения, главным образом, за счет высокой рождаемости, и высокий уровень естественного прироста населения. При этом уровень смертности средний, а уровень миграционного прироста – низкий. С экономической точки зрения данный кластер является самым неблагополучным. Число семей, нуждающихся в улучшении жилищных условий, невелико, но и темпы расселения очень низкие. Среди положительных моментов можно отметить относительно развитую инфраструктуру, сфера здравоохранения развита на среднем уровне. Преступность низкая.

Таким образом, вторичных анализ статистических данных с применением ряда математических методов обработки данных позволил проиллюстрировать эвристические возможности вторичного анализа статистических данных и создать социальную карту Саратовской области, позволяющую провести глубокий и всесторонний социологический анализ нашего региона.