

ОСНОВЫ СТАТИСТИКИ

Вводная лекция





Здравствуйтесь!

Немного поговорим об авторе книги, которая явилась базовым учебником при разработке курса лекций. Введение статистических методов в экспериментальное исследование часто происходит уже после получения результатов. Однако это нежелательно, поскольку отсутствие составленного заранее статистического плана может послужить причиной ошибочных выводов. В учебнике Джозефа Хили рассматриваются инструменты для социальных исследований. Книга позволяет извлечь информацию из данных, узнать их структуру и на основе полученной информации провести дальнейший анализ. В этом учебнике рассмотрена базовая концепция статистики, которую следует знать. Понимание статистики в контексте исследования социологии дает представление об общей концепции планирования эксперимента и возможностях статистического анализа.

Статистические данные – часть обыденного языка социологии и других социальных наук, включая политологию, социальную работу, государственное управление, уголовное судопроизводство, городские исследования и геронтологию. Эти дисциплины основаны на исследовании и обычно используют статистику, чтобы выразить знание и обсудить теорию.

Знание статистики позволит вам понять данные исследования, провести количественное исследование самим, способствовать повышению знания социологии и достигнуть вашего полного потенциала как социологу. Основная цель этого учебника состоит в том, чтобы развивать статистическую грамотность. Статистически грамотный человек понимает и ценит роль статистики в процессе исследования, сможет компетентно выполнить основные вычисления, может прочитать и оценить профессиональную литературу в его области, а также любых отчетах о научно-исследовательской работе, с которыми он может столкнуться.

В качестве инструмента автор предлагает использовать статистический пакет «SPSS», но в курсе лекций нами упор все таки будет сделан на использование распространённого и доступного инструмента – MS Excel.

Кроме того, для формирования материала были привлечены и дополнительные источники, такие как курс лекций профессора Шалабанова А.К. и Роганова Д.А., официальные источники – www.statistica.ru, www.businessman.ru, www.datuapstrade.lv, statanaliz.info, а также работы Орлова А.И. в области статистики. Для более точного перевода, большего понимания и адаптации материала Джозефа Хили к нашей реальности, для применения полученных навыков слушателям предлагается работать с источниками статистических данных представляемых для обозрения Агентством по статистике Республики Казахстан.

Что такое статистика и зачем она нужна? Как сказали Илья Ильф и Евгений Петров, «Статистика знает все» (из произведения «12 стульев»).

Что вы делаете, когда видите большой объем цифр, графиков и слово «статистика»? Думаем, как правило, просто игнорируете их. Но только не сегодня.

Согласно статистическим данным, деревьев на Земле больше, чем звезд на Млечном Пути. Сейчас на планете примерно 3 триллиона деревьев, в то время как звезд – 400 миллиардов.

В среднем человек за день проходит 7500 шагов, при этом средняя продолжительность жизни составляет 80 лет. Имея только эти данные, исследователи высчитали, что за всю жизнь можно пройти целых 180 000 километров. Такое расстояние позволяет обойти Землю по экватору 4,5 раза.

Статистики даже выяснили, в какой день рождается больше всего детей. Так, лидирующее место принадлежит четвергу (17,5 %), на втором месте – понедельник (14,2 %), а вот самый непопулярный для рождения день – это суббота (12,5 %).

На самом деле статистика является частью повседневной жизни социологии и других социальных наук, нашей повседневной жизни. Статистические методы дают жизнь, как нам кажется, «мертвым» цифрам.



После статистической обработки данных мы получаем ответы на многие вопросы. Таким образом, цифры начинают говорить. Оценка удовлетворенности населения или работников, выявление предпочтений избирателей, создание и функционирование системы показателей работы учреждения и многие другие опросы позволяют получать точные и достоверные данные, так как это информация из первых рук – непосредственно от населения. В связи с этим особенно важно приобрести навыки в обобщении со статистическими данными и поиске скрытых закономерностей.

Дадим классическое определение статистики. Статистика – это общественная наука, изучающая явления и процессы общественной жизни, она раскрывает законы возникновения и развития этих явлений и их взаимосвязи. Нельзя не обратить на внимание на еще одно понятие – это статистические методы обработки информации.

Для изучения различных общественных и социально-экономических явлений, а также некоторых процессов, происходящих в природе, проводятся статистические исследования. Любое статистическое исследование начинается с целенаправленного сбора информации об изучаемом явлении или процессе. Как правило, порядок преобразований первоначально полученной информации таков: сначала данные измерения упорядочивают и группируют. Первое, что следует оценить – это рамки, в которых могут находиться данные измерения. Затем составляют таблицы распределения данных. Табличное представление информации – это внесение в таблицу ряда данных и измерений. Таблицы распределения переводят в графики распределения. При графическом представлении больших объемов информации применяют столбчатые диаграммы, получают своего рода паспорт данных измерения, в котором собрано наибольшее количество основных числовых характеристик полученной информации. Одна из основных задач статистики состоит в обработке информации. Конечно, у статистики есть много других задач: получение и хранение информации, выработка различных прогнозов, оценка их достоверности и т.д., но ни одна из них не достижима без обработки данных.

Деятельность людей во множестве случаев предполагает работу с данными, а она в свою очередь может подразумевать не только оперирование ими, но и их изучение, обработку и анализ. Например, когда нужно уплотнить информацию, найти какие-то взаимосвязи или определить структуры. И как раз для аналитики в этом случае очень удобно пользоваться не только разными техниками мышления, но и применять статистические методы.

Особенностью методов статистического анализа является их комплексность, обусловленная многообразием форм статистических закономерностей, а также сложностью процесса статистических исследований. Однако мы хотим поговорить именно о таких методах, которые может применять каждый, причем делать это эффективно и с удовольствием.

Статистическое исследование может проводиться посредством следующих методик:

- Статистическое наблюдение;
- Сводка и группировка материалов статистического наблюдения;
- Абсолютные и относительные статистические величины;
- Вариационные ряды;
- Выборка;
- Корреляционный и регрессионный анализ;
- Ряды динамики.

Далее мы рассмотрим каждый из них более подробно. Но отметим, что представим лишь основные характеристики без подробного описания алгоритмов действий. Впрочем, понять их не составит никакого труда.

Статистическое наблюдение

Статистическое наблюдение является планомерным, организованным и в большинстве случаев систематическим сбором информации, направленным, главным образом, на явления социальной жизни. Реализуется данный метод через регистрацию предварительно определенных наиболее ярких признаков, цель которой состоит в последующем получении характеристик изучаемых явлений.



Статистическое наблюдение должно выполняться с учетом некоторых важных требований:

1. Оно должно полностью охватывать изучаемые явления;
2. Получаемые данные должны быть точными и достоверными;
3. Получаемые данные должны быть однообразными и легко сопоставимыми.

Также статистическое наблюдение может иметь две формы.

Отчетность – это такая форма статистического наблюдения, где информация поступает в конкретные статистические подразделения организаций, учреждений или предприятий. В этом случае данные вносятся в специальные отчеты.

Специально организованное наблюдение – наблюдение, которое организуется с определенной целью, чтобы получить сведения, которых не имеется в отчетах, или же для уточнения и установления достоверности информации отчетов. К этой форме относятся опросы (например, опросы мнений людей), перепись населения и т.п.

Кроме того, статистическое наблюдение может быть категоризировано на основе двух признаков: либо на основе характера регистрации данных, либо на основе охвата единиц наблюдения. К первой категории относятся опросы, документирование и прямое наблюдение, а ко второй – наблюдение сплошное и несплошное, т.е. выборочное.

Для получения данных при помощи статистического наблюдения можно применять такие способы как анкетирование, корреспондентская деятельность, самоисчисление (когда наблюдаемые, например, сами заполняют соответствующие документы), экспедиции и составление отчетов.

Сводка и группировка материалов статистического наблюдения

Говоря о втором методе, в первую очередь следует сказать о сводке. Сводка представляет собой процесс обработки определенных единичных фактов, которые образуют общую совокупность данных, собранных при наблюдении.

Если сводка проводится грамотно, огромное количество единичных данных об отдельных объектах наблюдения может превратиться в целый комплекс статистических таблиц и результатов. Также такое исследование способствует определению общих черт и закономерностей исследуемых явлений.

С учетом показателей точности и глубины изучения можно выделить простую и сложную сводку, но любая из них должна основываться на конкретных этапах:

- Выбирается группировочный признак;
- Определяется порядок формирования групп;
- Разрабатывается система показателей, позволяющих охарактеризовать группу и объект или явление в целом;
- Разрабатываются макеты таблиц, где будут представлены результаты сводки.

Важно заметить, что есть и разные формы сводки. Централизованная сводка, требующая передачи полученного первичного материала в вышестоящий центр для последующей обработки.

Децентрализованная сводка, где изучение данных происходит на нескольких ступенях по восходящей.

Выполняться же сводка может при помощи специализированного оборудования, например, с использованием компьютерного ПО или вручную.

Что же касается группировки, то этот процесс отличается разделением исследуемых данных на группы по признакам. Особенности поставленных статистическим анализом задач влияют на то, какой именно будет группировка: типологической, структурной или аналитической. Именно поэтому для сводки и группировки либо прибегают к услугам узкопрофильных специалистов, либо применяют конкретные техники мышления.



Абсолютные и относительные статистические величины

Абсолютные величины считаются самой первой формой представления статистических данных. С ее помощью удастся придать явлениям размерные характеристики, например, по времени, по протяженности, по объему, по площади, по массе и т.д.

Если требуется узнать об индивидуальных абсолютных статистических величинах, можно прибегнуть к замерам, оценке, подсчету или взвешиванию. А если нужно получить итоговые объемные показатели, следует использовать сводку и группировку. Нужно иметь в виду, что абсолютные статистические величины отличаются наличием единиц измерения. К таким единицам относят стоимостные, трудовые и натуральные.

А относительные величины выражают количественные соотношения, касающиеся явлений социальной жизни. Чтобы их получить, одни величины всегда делятся на другие. Показатель, с которым сравнивают (это знаменатель), называют основанием сравнения, а показатель, которой сравнивают (это числитель), называют отчетной величиной.

Относительные величины могут быть разными, что зависит от их содержательной части. Например, существуют величины сравнения, величины уровня развития, величины интенсивности конкретного процесса, величины координации, структуры, динамики и т.д.

Чтобы изучить какую-то совокупность по дифференцирующимся признакам, в статистическом анализе применяются средние величины – обобщающие качественные характеристики совокупности однородных явлений по какому-либо дифференцирующему признаку.

Крайне важным свойством средних величин является то, что они говорят о значениях конкретных признаков во всем их комплексе единым числом. Невзирая на то, что у отдельных единиц может наблюдаться количественная разница, средние величины выражают общие значения, свойственные всем единицам исследуемого комплекса. Получается, что при помощи характеристики чего-то одного можно получить характеристику целого.

Следует иметь в виду, что одним из самых важных условий применения средних величин, если проводится статистический анализ социальных явлений, считается однородность их комплекса, для которого и нужно узнать среднюю величину. А от такого, как именно будут представлены начальные данные для исчисления средней величины, будет зависеть и формула ее определения.

Вариационные ряды

В некоторых случаях данных о средних показателях тех или иных изучаемых величин может быть недостаточно, чтобы провести обработку, оценку и глубокий анализ какого-то явления или процесса. Тогда во внимание следует брать вариацию или разброс показателей отдельных единиц, который тоже представляет собой важную характеристику исследуемой совокупности.

На индивидуальные значения величин могут воздействовать многие факторы, а сами изучаемые явления или процессы могут быть очень многообразны, т.е. обладать вариацией (это многообразие и есть вариационные ряды), причины которой следует искать в сущности того, что изучается.

Вышеназванные абсолютные величины находятся в непосредственной зависимости от единиц измерения признаков, а значит, делают процесс изучения, оценки и сравнения двух и более вариационных рядов более сложным. А относительные показатели нужно вычислять в качестве соотношения абсолютных и средних показателей.

Выборка

Смысл выборочного метода (или проще – выборки) состоит в том, что по свойствам одной части определяются численные характеристики целого (это называется генеральной совокупностью). Основным выборочного метода является внутренняя связь, объединяющая части и целое, единичное и общее.

Метод выборки отличается рядом существенных преимуществ перед остальными, т.к. благодаря уменьшению количества наблюдений позволяет сократить объемы работы, затрачиваемые средства и



усилия, а также успешно получать данные о таких процессах и явлениях, где-либо нецелесообразно, либо просто невозможно исследовать их полностью.

Соответствие характеристик выборки характеристикам изучаемого явления или процесса будет зависеть от комплекса условий, и в первую очередь от того, как вообще будет реализовываться выборочный метод на практике. Это может быть как планомерный отбор, идущий по подготовленной схеме, так и непланомерный, когда выборка производится из генеральной совокупности.

Но во всех случаях выборочный метод должен быть типичным и соответствовать критериям объективности. Данные требования нужно выполнять всегда, т.к. именно от них будет зависеть соответствие характеристик метода и характеристик того, что подвергается статистическому анализу.

Таким образом, перед обработкой выборочного материала необходимо провести его тщательную проверку, избавившись тем самым от всего ненужного и второстепенного. Одновременно с этим, составляя выборку, в обязательном порядке нужно обходить стороной любую самодеятельность. Это означает, что ни в коем случае не следует делать выборку только из вариантов, кажущихся типичными, а все другие – отбрасывать.

Эффективная и качественная выборка должна составляться объективно, т.е. производить ее нужно так, чтобы были исключены любые субъективные влияния и предвзятые побуждения. И чтобы это условие было соблюдено должным образом, требуется прибегнуть к принципу рандомизации или, проще говоря, к принципу случайного отбора вариантов из всей их генеральной совокупности.

Представленный принцип служит основой теории выборочного метода, и следовать ему нужно всегда, когда требуется создать эффективную выборочную совокупность, причем случаи планомерного отбора исключением здесь не являются.

Корреляционный и регрессионный анализ

Корреляционный анализ и регрессионный анализ – это два высокоэффективных метода, позволяющие проводить анализ больших объемов данных для изучения возможной взаимосвязи двух или большего количества показателей.

В случае с корреляционным анализом задачами являются:

Измерить тесноту имеющейся связи дифференцирующихся признаков;

Определить неизвестные причинные связи;

Оценить факторы, в наибольшей степени воздействующие на окончательный признак.

А в случае с регрессионным анализом задачи следующие:

- Определить форму связи;
- Установить степень воздействия независимых показателей на зависимый;
- Определить расчетные значения зависимого показателя.

Чтобы решить все вышеназванные задачи, практически всегда нужно применять и корреляционный и регрессионный анализ в комплексе.

Ряды динамики

Посредством этого метода статистического анализа очень удобно определять интенсивность или скорость, с которой развиваются явления, находить тенденцию их развития, выделять колебания, сравнивать динамику развития, находить взаимосвязь развивающихся во времени явлений.

Ряд динамики – это такой ряд, в котором во времени последовательно расположены статистические показатели, изменения которых характеризуют процесс развития исследуемого объекта или явления.

Ряд динамики включает в себя два компонента:

Период или момент времени, связанный с имеющимися данными;



Уровень или статистический показатель.

В совокупности эти компоненты представляют собой два члена ряда динамики, где первый член (временной период) обозначается буквой «t», а второй (уровень) – буквой «y».

Исходя из длительности временных промежутков, с которыми взаимосвязаны уровни, ряды динамики могут быть моментными и интервальными. Интервальные ряды позволяют складывать уровни для получения общей величины периодов, следующих один за другим, а в моментных такой возможности нет, но этого там и не требуется.

Ряды динамики также существуют с равными и разными интервалами. Суть же интервалов в моментных и интервальных рядах всегда разная. В первом случае интервалом является временной промежуток между датами, к которым привязаны данные для анализа (удобно использовать такой ряд, например, для определения количества действий за месяц, год и т.д.). А во втором случае – временной промежуток, к которому привязана совокупность обобщенных данных (такой ряд можно использовать для определения качества тех же самых действий за месяц, год и т.п.). Интервалы могут быть равными и разными, независимо от типа ряда.

Естественно, чтобы научиться грамотно применять каждый из методов статистического анализа, недостаточно просто знать о них, ведь, по сути, статистика – это целая наука, требующая еще и определенных навыков и умений. Но чтобы она давалась проще, можно и нужно тренировать свое мышление и улучшать когнитивные способности.

В остальном же исследование, оценка, обработка и анализ информации – очень интересные процессы. И даже в тех случаях, когда это не приводит к какому-то конкретному результату, за время исследования можно узнать множество интересных вещей. Статистический анализ нашел свое применение в огромном количестве сфер деятельности человека, а вы можете использовать его в учебе, работе, бизнесе и других областях, включая развитие детей и самообразование.

Нельзя не обратить внимания на развитие статистики в Республике Казахстан, раскрыв некоторые исторические факты. Статистика Казахстана своими корнями уходит в далекое прошлое. Имеются исторические подтверждения статистическим сведениям о первом Казахском государстве – Казахском Ханстве. В начале его образования (1459 год) в долинах рек Шу и Талас (на территории нынешней Жамбылской области) численность населения составляла 200 тыс. человек, а к концу XV – века достигла 1 миллиона. Однако зарождение на территории современного Казахстана регулярной и централизованной статистической деятельности относится ко второй половине XVIII века, т.е. к периоду вхождения Казахстана в состав Российской империи. Первая всеобщая перепись населения на его территории, как и во всей царской России, была проведена по состоянию на 9 февраля (28 января) 1897 года. Первым официальным государственным статистическим органом, образованным на территории Казахстана, является Туркестанский губернский статкомитет (дата образования 22 января 1868 года) и подведомственные ему статбюро в Сырдарьинской и Семиреченской областях. В середине 70-х годов XIX века организовался Уральский областной статкомитет, в 1877 году – Семипалатинский и Акмолинский (в г. Омске) и в 1895 году – Тургайский областные статкомитеты. Однако до 1920 года в Казахстане не было единого статистического органа, объединяющего указанные и другие местные статистические службы. С образованием (26 августа 1920 года) Казахской Автономной Социалистической Республики в составе РСФСР, Правительство Казахской АССР своим постановлением от 8 ноября 1920 года утвердило «Положение о государственной статистике в Казахской АССР» и образовало Статистическое управление АССР. Таким образом, датой образования единых централизованных статистических органов Казахстана принято считать 8 ноября 1920 года.

За последние пятнадцать лет Агентство Республики Казахстан по статистике условно прошло следующие фазы развития:

1) 1992–1996 годы – формирование Агентства как национального органа, создающего основной методологический ноу-хау, внедрение стандарта системы национальных счетов (СНС 93).



2) 1996–1998 годы – освоение методологии составления интегрированных счетов и таблиц СНС; запуск систематического использования международно-согласованных статистических классификторов; начало создания статистических регистров; введение статистических методов производства информации по малым предприятиям; введение новых технологий информации и связи.

3) 1995–2005 годы – фактическое внедрение международных классификаций во всех областях статистического производства; успешное проведение первой казахстанской переписи населения 1999 года и развитие демографической социальной статистики; внедрение прогрессивных методов массовой обработки данных, получение технической помощи в рамках международного сотрудничества.

4) 2006–2008 годы – реализация Программы совершенствования государственной статистики, включая пересмотр методологий и классификаций, адаптации развивающихся международных стандартов, начало внедрения системы метаданных и интегрированных классификаций;

5) С 2008 года – модернизация процессов обработки данных, пересмотр системы сбора данных, внедрение современных технологий в области обработки данных (хранилище данных, электронная статистическая отчетность и др.), активные инвестиции в развитие персонала и расширение международного сотрудничества.

Вернемся непосредственно к теме нашего разговора. Если говорим «статистика», мы подразумеваем понятие расчеты, следовательно, математика. Какие же знания нам необходимы для успешного освоения курса «Основы статистики», прежде всего конечно же всевозможные арифметические действия как положительными числами, так и с отрицательными, умение работать с обыкновенными и десятичными дробями))). Основными математическими операциями являются сложение (+), вычитание (-), умножение (*), деление (/).

Например, что получим, если сложить два числа с разными знаками

$$5+(-2)=5-2=3, \text{ или } 5-(-3)=5+3=8,$$

$$-4*(-3)=12, (-4)*3=-12 \text{ и наоборот } 3*(-4)=-12$$

Если же разделить числа с разными знаками:

$$-4:2=-2; \quad 4:-2=-2; \quad -4:-2=2$$

Рассмотрим следующие выражения:

$$5(4)=20; \quad 8+2 \times 222=8+2 \times 42=8+82=8+4=12;$$

$$8+2-4(5-1)=10-44=10-9=9;$$

$$15-96/2=63/2=2$$

Еще не менее важные операции, это нахождение

$$\sum X_i = 2+4+5+7=18$$

суммы ряда

$$\text{возведение числа в степень, например квадрат } X^2=(X)(X)=(4)(4)=16$$

$$\text{извлечение из под коренного выражения } X=16=4$$

Подведем итоги.

Что мы сегодня узнали, что такое статистика и какова ее роль в современности, что представляют собой статистические методы, и как с ними найти общий язык. Немного поговорили об истории развития статистики в Республике Казахстан и об авторе учебник, а на основе которого построен данный курс.