



Российская Академия предпринимательства

Г. Т. Журавлев

**Использование ППП STATISTICA
для обработки и анализа первичной
социологической информации**

Учебно-практическое пособие (УПП)

**Москва
2008**

УДК 303.1
ББК 60.5я73
Ж 91

Рецензент

доктор философских наук, профессор,
зав. кафедрой социологии и связи с общественностью

Маршак А.Л.

Ж91 Журавлев, Г.Т.

Использование ППП STATISTICA для обработки и анализа первичной социологической информации: учебно-практическое пособие (УПП). – 2-е изд. доп. и перераб. – М.: Российская Академия предпринимательства; Агентство печати «Наука и образование», 2008. – 60 с.

УДК 303.1
ББК 60.5я73

Учебное пособие обсуждено и рекомендовано
к использованию в учебном процессе на заседании
Редакционно-издательского Совета.
Протокол № 8 от 28 августа 2008 года

Введение

ППП (пакет прикладных программ) STATISTICA представляет собой обширный набор методов решения различных исследовательских и управленческих задач. Изучение его в полном объеме дело трудоемкое. Обычно, в процессе изучения методов, студенты задают вопрос, а зачем это надо? Поэтому в предлагаемом УПП излагаются не только методы, но, прежде всего, задачи; суть которых, на наш взгляд, более понятна студентам, чем абстрактные методы. В 1968 г. мне довелось вместе с другими авторами опубликовать книгу, в которой излагались методы статистической обработки первичной социологической информации¹. Спустя 40 лет я не могу сказать, что эта книга оказала большое влияние на внедрение этих методов в практику социологических исследований и на квалификацию социологов в этой области. Разумеется, сдвиг в положительную сторону есть не малый. Но если говорить о массовом использовании статистических методов в социологии, то дело оставляет желать лучшего.

Многолетний педагогический опыт привел меня к мысли, что мы начинаем не с того. Надо начинать не с методов, а с ответа на вопрос: «зачем это надо?». Поэтому целесообразно начать с задач, которые встают перед социологом, организующим прикладное исследование, которые ставит перед ним заказчик, которые надо решать.

С проблемой исследовательских задач социолог сталкивается в самом начале подготовки исследования, на этапе разработки программы социологического исследования. Задачу иногда ставит заказчик исследования. Например, заказчик говорит: «Меня интересует, насколько фирма может увеличить производство и поставки нашего товара на рынок через пять лет?» Эта задача носит прогностический характер. Заказчика интересует прогноз выпуска товара на пять лет. Или, заказчик говорит: «Несмотря на постоянно растущую зарплату, внедрение новых машин, качество нашей продукции снижается, спрос падает. Какие причины влияют на это?» Здесь ставится задача поиска причинно-следственных связей, которые определяют качество производимой продукции. Но не в явном виде стоит и другая задача — оптимизации факторов, влияющих на качество продукции. Сколько надо платить работникам, какие должны быть взаимоотношения в трудовом коллективе, надо ли вовлекать работников в решение управленческих задач, повышать их квалификацию и т.д., ибо на качество продукции влияют не только производственно-технические условия, но и отношение работников к выполняемой работе, к фирме в

¹ См.: Методика и техника статистической обработки первичной социологической информации. М.: Наука, 1968.

целом. Подбор наиболее удачных параметров различных факторов есть проблема оптимизации.

В том случае, когда социолог впервые начинает изучать какой-либо объект (учебное заведение, промышленное предприятие, торговую фирму и пр.), перед ним возникает иная задача — описать объект, получить количественные характеристики различных показателей (средняя успеваемость, средняя заработная плата, средняя производительность труда и т.п.), чтобы иметь представление об этом объекте или сравнить его с другими сходными объектами. Здесь мы имеем дело с задачей описания изучаемого объекта.

Итак, мы выделили четыре задачи, которые наиболее часто встречаются в исследовательской и управленческой практике. Это — задача описания объекта (с нее начинается анализ эмпирических данных); задача, связанная с объяснением причинно-следственных связей; задача прогнозирования социальных процессов; задача оптимизации процессов.

В ППП STATISTICA разработаны методы решения этих задач. Разумеется, круг задач, решаемых социологами и менеджерами, собственниками фирм, значительно шире, но следует начать с простого, а затем, овладев этими методами, студенты или молодые специалисты самостоятельно научатся решать и более сложные задачи.

Использование ППП STATISTICA, а также SPSS, позволяет создать уникальную среду, в которой статистическая обработка и анализ данных превращается из рутинного дела в увлекательное исследование. Вместо того, чтобы тратить время на утомительные расчеты, исследователь больше сил и времени использует для анализа, перебора многих вариантов причинно-следственных связей или оптимизации изучаемых процессов, что позволяет получать многовариантные решения с использованием компьютерных технологий и современных методов анализа. Важнейшим преимуществом такого подхода является огромное быстроедействие компьютеров; по сути, двумя-тремя нажатиями клавиш за доли секунды мы получаем результат, который вручную мы получили бы за месяц-другой напряженного рутинного труда.

Это быстроедействие позволяет в короткие промежутки времени получать большое число таблиц, графиков, коэффициентов корреляции, уравнений регрессии и др. и анализировать разные варианты, которые будут для нашей фирмы наилучшими. Именно поиск наилучших, или оптимальных, вариантов и является наиболее творческим и интересным процессом в исследовательской и аналитической работе. Обычно менеджер принимает интуитивные решения, половина из них являются ошибочными; а расчет разных вариантов позволяет оценить наилучшее решение из возможных и реализовать его. Доля ошибочных решений резко уменьшается.

1. Задачи описания объектов

Эти задачи возникают, когда исследователь впервые начинает изучать объект. Исследователя интересуют количественные характеристики тех или иных свойств обследуемого объекта. Например, дан заказ изучить эффективность работы ряда вузов. Вполне естественно, что необходимо иметь общее представление об этих учебных заведениях. Исследователя будут интересовать такие данные — общее количество студентов, в том числе по факультетам; средняя успеваемость; средняя стипендия; средняя плата за обучение; среднее количество студентов в группе; наиболее часто получаемые на экзаменах оценки; среднее число пропусков занятий; наиболее часто встречающееся число пропусков занятий и др. Очевидно, разные вузы будут иметь разные числовые характеристики. На этом основании судят о лучшем или худшем вузе.

Эти данные необходимы для двух целей: 1) получить общее представление об изучаемых объектах; 2) сравнить два (или более) объекта.

Для описания объектов используются следующие статистические показатели: средняя арифметическая величина (M); мода (Mo); медиана (Me); а также меры рассеяния признака: стандартное (среднеквадратическое) отклонение (σ); дисперсия (σ^2).

Средняя арифметическая величина выражает величину признака, отнесенную к единице совокупности и абстрагированную от индивидуальных особенностей отдельных единиц. Она подсчитывается по формуле:

$$M = (x_1 + x_2 + x_3 + \dots + x_n) / n; \quad (1)$$

где M — средняя арифметическая величина,
 x — индивидуальные значения величины признака,
 n — число единиц наблюдения,
/ — знак деления.

Например, рассчитаем среднюю зарплату пяти рабочих:

$$M = (3000 \text{ руб.} + 4000 \text{ руб.} + 4000 \text{ руб.} + 5000 \text{ руб.} + 6000 \text{ руб.}) / 5 = 4400 \text{ руб.}$$

Как видно, средняя заработная плата группы рабочих (обследованной совокупности) равна 4400 руб. В другой группе работников она может быть 7000 руб. Значит, в поисках рабочего места работник будет выбирать то предприятие, где выше средняя зарплата.

Здесь под *признаком* понимается любое свойство объекта (человека, фирмы, коллектива, города и др.), которое можно измерить. Измеренное значение признака будет его величиной. Например, возраст измеряется количеством прожитых лет; индивид А имеет 22 года, это есть величина признака этой единицы совокупности — индивида А.

Наряду со средней арифметической величиной рассчитываются также так называемые структурные средние – мода (*Mo*) и медиана (*Me*).

Мода представляет собой значение изучаемого признака, которое повторяется с наибольшей частотой. В вышеприведенном примере чаще других встречается заработная плата 4000 руб. Это и есть модальное значение изучаемого признака. $Mo = 4000$ руб.

Медианой называется значение признака, приходящееся на середину ранжированной (упорядоченной по убыванию или возрастанию) совокупности. Медиана – значение признака, которое разбивает выборку на две равные части. Половина наблюдений лежит ниже медианы, и половина наблюдений лежит выше медианы. Медианой будет значение признака у объекта, который находится в середине этого ряда. Если мы всех обследованных индивидов построим в один ряд по возрастанию или убыванию их зарплаты, то медианой будет значение признака у индивида (объекта), который находится в середине этого ряда. Можно не выстраивать людей, а составить список, в котором фамилии будут располагаться не по алфавиту, а по убыванию или возрастанию их зарплаты. В вышеприведенном примере медиана равна 4000 руб., поскольку в середине ряда находится значение 4000 руб. В нашем примере мода и медиана численно равны. Это обычно бывает в симметричных распределениях. Но чаще они несколько отличаются одна от другой.

Перейдем к расчетам этих величин с использованием пакета прикладных программ (ППП) STATISTICA. В настоящем учебно-методическом пособии используется STATISTICA 6. (STATISTICA 5 – русифицированный вариант). Различия между 5 и 6 вариантами незначительны. Последовательность команд такая:

Команды. Запуск ППП STATISTICA.

1. Пуск.
2. Программы.
3. STATISTICA.
4. Открывается окно статистических модулей.

Название модуля высвечивается в верхней строке окна.

Вторая сверху строка окна (строка меню) содержит ряд пунктов меню, общих для Windows приложений: File – Файл, Edit – Правка, View – Вид, Window – Окно, Help – Помощь. Есть пункты меню, которые присущи ППП STATISTICA: Insert, Format, Statistics, Graphs, Tools, Data, Help. По мере работы с ППП мы будем рассматривать, какие функции они выполняют.

Многие операции в ППП STATISTICA схожи с операциями в Word и Excel, поэтому студенты, работающие с этими программами, без труда освоят и пакет STATISTICA. Как показывает опыт, трудность зак-

лючается не в том, как пользоваться процедурами, заложенными в этом ППП, а в знании, понимании и правильном истолковании полученных статистических показателей. Поэтому целесообразно параллельно с освоением ППП STATISTICA освежить знания в области теории статистики или заново изучать эту дисциплину.

Для дальнейшей работы следует выполнить следующие действия:

- ввести данные или открыть файл данных, если он был создан раньше;
- выбрать переменные для анализа;
- выбрать метод анализа из меню в стартовой панели модуля;
- выбрать конкретную вычислительную процедуру и задать ее параметры;
- произвести запуск вычислительной процедуры;
- выбрать следующий шаг анализа.

Прежде всего, надо ввести данные наблюдений (например, данные анкетного опроса). Для этого (как и в Excel) открываем новую исходную матрицу.

Команды:

1. Курсор в левый верхний угол и щелчок мышью «Создать».

На экране появится чистая таблица 10 x 10. Файлы в папке STATISTICA организованы аналогично файлам в электронных таблицах (например, Excel). Такой файл можно рассматривать как таблицу (Spreadsheet), в которой столбцы являются переменными (Variables), например, вопросы анкеты, а в строках записываются значения переменных — наблюдения (Cases), анкеты. Для анкет это означает, что переменными выступают вопросы анкеты, а наблюдениями — сами анкеты. Ответы на вопросы анкеты вводятся в столбцы и строки.

После щелчка мышью «Создать» на экране появится вспомогательное окно (Create New Document) с командами. Нас интересует малое окно в центре большого окна, в котором есть две цифры 10. Это означает, что мы получили матрицу 10 x 10, в которую надо вводить данные опроса. Но чаще всего, нам нужны другие размерности матрицы, поэтому надо стрелками либо добавить, либо уменьшить число вариантов (вопросов) и наблюдений (анкет).

Это можно сделать иначе.

Команды:

1. Курсор на Data (Данные). На экране появится подменю.
2. Курсор на Vars (Варианты). Появляется еще одно подменю, в котором есть команды Add (прибавить) Delete (стереть). Пользуясь ими, мы получим искомую матрицу с необходимым числом вариантов (вопросов анкеты). Если нам надо добавить 5 столбиков, поскольку у нас

не 10, а 15 вопросов, например, в анкете, мы вводим в верхнюю строку этого окна цифру 5, и ОК. Если надо убавить, убрать 5 столбцов, вводим цифру 5 и ОК (Delete).

3. Курсор на Cases. Появится точно такое же подменю. Точно так же мы увеличиваем или уменьшаем число наблюдений (анкет).

Второй способ хорош тогда, когда, например, мы собрали 1000 анкет и создали матрицу 40x1000; но в процессе ремонта выборки нам потребовалось опросить еще 200 человек. Здесь мы используем второй способ формирования исходной матрицы данных.

Для примера решим такую задачу: рассчитать средние величины – M , M_o и M_e некоторой выборочной совокупности студентов. В качестве признаков мы возьмем следующие: x_1 – возраст, x_2 – среднемесячный денежный доход семьи, x_3 – среднемесячные расходы студента на потребление культуры (посещение театра и кино, покупка книг, экскурсии и пр.), x_4 – свободное время студента, часов в среднем за неделю, x_5 – время, затрачиваемое на поездки в институт, часов в среднем за неделю. Создадим исходную матрицу данных 5x41, где 5 это варианты (признаки), а 41 это число наблюдений, или анкет. (Табл.1)

Последовательность операций и команд.

1. «Создать».
2. На экране появилось окно «Create New Document».
3. В окне «Number of variables» устанавливаем цифру 5.
4. В окне «Number of cases» устанавливаем цифру 41.
5. ОК.
6. На экране появилась электронная таблица 5 x 41, не заполненная.
7. Сохраняем файл под каким-либо именем в STATISTICA.

В эту таблицу будем вносить данные нашего опроса. Все таблицы должны иметь номера и названия, а также в таблице верхние и боковые заголовки.

Таблица 1

Распределение респондентов по пяти признакам

Возраст, лет	Доход, руб.	Расходы на культуру, руб.	Свободное время, час.	Время на дорогу, час.
17	10 500	1520	2,55	2,49
17	11 000	1530	3,98	2,46
18	12 000	1420	4,23	2,31
18	11 150	1540	4,12	2,25
18	15 220	1545	4,11	2,12
18	13 100	1510	3,85	3,21
19	11 320	1430	4,02	3,12
19	12 320	1490	4,01	2,84

Продолжение таблицы

19	15 200	1340	3,81	2,72
19	14 650	1600	5,56	2,61
19	15 110	1620	5,25	2,36
20	15 200	1330	4,01	2,45
20	14 000	1650	3,71	2,23
20	15 100	1680	3,62	2,12
20	13 540	1720	3,51	1,82
20	15 900	1730	2,21	1,73
20	16 100	1740	2,32	1,65
21	14 200	1810	2,41	1,44
21	17 120	1710	3,42	3,45
21	16 450	1820	3,32	3,14
21	14 500	1830	1,91	3,32
21	15 110	1230	2,23	2,81
21	16 050	1290	1,73	2,75
21	17 120	1920	2,55	2,41
22	13 000	1930	2,88	2,23
22	17 620	1950	2,76	2,53
22	17 700	1800	2,51	2,22
22	15 650	2020	3,21	1,84
22	18 000	2030	1,23	1,93
22	18 150	2210	1,01	1,45
23	18 250	2120	2,99	3,42
23	19 000	1810	0,99	3,11
23	16 220	1910	0,85	1,12
23	17 200	2210	0,64	1,22
24	14 220	1520	1,66	2,23
24	18 120	2230	1,22	1,71
24	16 000	2100	3,22	2,11
24	14 400	1630	0,44	2,22
25	18 910	2360	1,02	2,32
25	19 560	1650	0,33	2,12
25	20 550	2400	4,55	2,44

Очень важный момент – озаглавить столбцы, то есть ввести названия вопросов и указать шкалы ответов. Для этого подводим курсор к var1 (в нашей исходной матрице 5 x 41 столбцы поименованы – var1, var2, var3, var4, var5). Вместо этих обозначений мы должны написать названия вопросов: *двойной щелчок* мышкой. На экране появится картинка. Во втором сверху слева окне (Name) написано «Var1». Вместо этого пишем «Возраст» (без кавычек). Справа есть две стрелки (под словом Cancel): стрелка, которая показывает направо, листает варианты дальше по возрастианию. Другая, левая стрелка позволяет листать варианты по уменьшению.

Подводим курсор на правую стрелку, щелчок мышью и на экране появляется Var2. Пишем возраст, и т.д. *Названия не должны быть длинными; вопросы надо излагать сокращенно.*

В нижнем окне мы можем указать шкалы. Если, например, возраст был бы шкалирован (1 – до 17 лет, 2 – 18-20 лет, 3 – 21-24 года и т.д.), то в нижнем окне мы напечатали бы эти обозначения, эту шкалу. В результате того, что столбцы озаглавлены, последующие таблицы, графики и коэффициенты будут названы.

Итак, первичная социологическая информация введена в память компьютера. Начинаем обработку этих данных. Сначала займемся подсчетом *средних величин*.

Команды:

1. Курсор на Statistics. Щелчок мышью. На экране появилось подменю.
2. Курсор на Basic Statistics/Tables. Щелчок мышью. Новое подменю.
3. Курсор на Descriptive statistics.
4. ОК. На экране появилось окно, в котором перечислены статистические показатели, которые можно получить. Нас пока интересуют средняя арифметическая, мода и медиана. Если на экране нет этого перечня, то курсор на Advanced. Щелчок мышью и этот перечень появится.
5. Выделяем галочкой Mean (средняя арифметическая), Moda (мода), Median (медиана).
6. Курсор на Summari. Щелчок мышью. На экране появилось окно, в котором написаны наши варианты (возраст, доход, расходы на культуру, свободное время, время на дорогу).
7. Либо курсором активизируем (затемняем) все признаки, поскольку нас интересуют средние всех показателей; либо курсор на Select All.
8. ОК.

На экране появилась таблица (табл. 2), в которой указаны интересующие нас средние величины.

Multiple означает, что распределения по доходам, расходам денег на потребление культуры и времени на дорогу не имеют наиболее часто встречающейся величины признака. Если мы посмотрим на таблицу 1, то увидим, что, например, признак «Время на дорогу» имеет два значения признака, которые встречаются по три раза – 2,12 и 2,23 часа.

Как видно из таблицы 2, средний возраст обследованных 21,05 года, модальное значение 21 год. Студент, который находится в середине ранжированного ряда по увеличению возраста, имеет также возраст 21 год. Здесь мода и медиана совпадают.

Значения средних величин

	Средняя арифметическая	Медиана	Мода
1. Возраст, лет	21,05	21,00	21,00
2. Доход, руб.	15475,85	15220,00	<i>Multiple</i>
3. Расходы на культуру, руб.	1753,29	1720,00	Multiple
4. Свободное время, час.	2,78	2,88	2,55
5. Время на дорогу, час.	2,34	2,31	Multiple

Средний доход обследованных студентов (среднемесячный) составляет 15 475,85 руб. Медиана равна 15 220,00 руб. Средняя арифметическая несколько больше медианы; это говорит о том, что в выборке лиц с высоким доходом несколько больше, чем лиц с низким доходом. Они превышают среднюю арифметическую величину.

Денежные расходы на потребление культуры (приобретение книг, посещение кинотеатров и пр.) в среднем составляют 1753,29 руб. Медиана и здесь меньше средней арифметической – 1720,00 руб. Это значит, что число лиц, которые тратят на культуру больше средней арифметической, превышает число лиц, которые тратят на культуру меньше средней арифметической.

В том и другом случае, «средний доход» и «расходы на культуру», мы имеем не симметричное распределение, скошенное вправо, имеем правостороннюю асимметрию в распределении этих признаков. Это означает, что средняя арифметическая завышена; больше половины обследованных имеют показатели ниже средней арифметической.

Средняя арифметическая свободного времени составляет 2,78 часа. Медиана – 2,88 часа. Мода – 2,55 часа. Здесь в отличие от двух предыдущих распределений мы имеем левоасимметричное распределение. Число лиц, с меньшим в сравнении со средней арифметической величиной объемом свободного времени, превышает число лиц, имеющих свободного времени больше средней величины. Мода – 2,55 часа – меньше средней арифметической и медианы. Тем самым, чаще всего встречается значение признака «свободное время», которое ниже средней арифметической и медианы. Средняя арифметическая завышена за счет небольшого числа лиц, имеющих значительное больший объем свободного времени.

Время, затрачиваемое на передвижение по городу, в среднем составляет 2,34 часа. Медиана равна 2,31 часа. Здесь мы имеем правостороннюю асимметрию распределения этого признака. Значит число лиц, которые тратят на дорогу больше средней арифметической, превышает число лиц, затрагивающих на дорогу меньше средней арифметической.

Все эти данные ценны не только сами по себе, поскольку дают представление о выборочной совокупности студентов некоторого вуза, но они совершенно необходимы при сопоставлении двух или большего числа вузов или любых других объектов. Они позволяют судить, где лучше картина, а где — хуже.

Внимание! *Если мы перенесем таблицу из STATISTIC в WORD, она будет иметь вид, непригодный для чтения. Необходимо ее привести в соответствующий вид в WORD. Иначе говоря, таблицы надо редактировать. Надо заменить латинские буквы на кириллицу, вместо английских слов напечатать русские слова. Убрать лишние колонки и строки. Об этом ниже.*

Построение графиков

Распределение признаков можно представить графически, в более наглядном виде. Поэтому в рассматриваемом пакете есть команды для построения графиков.

Команды:

1. Курсор на Graphs. Щелчок мышью. Появляется подменю.
2. Histograms. Щелчок мышью. Появляется окно.
3. Variables. Щелчок мышью. Появляется окно.
4. Курсором активизируем все признаки, или курсор на Select All. Щелчок мышью. ОК.
5. Курсор на Advanced. Появляется окно.
6. Курсором помечаем галочкой Show percentages (показать проценты). ОК.
7. На экране появляются пять графиков.

Все графики должны иметь номера и названия, чтобы читатель понимал, что на графике изображено. Необходимо также пояснить текстом смысл графика, как мы поясняли содержание таблиц.

Здесь под столбиками обозначены значения признака, а над столбиками показано, сколько процентов в выборке составляют лица с такими значениями. Так что, 17 летних в выборке 5%, 18 летних — 10% и т.д.

В курсовой или дипломной работе целесообразнее графики получать в Excel. Дело в том, что, если в компьютере нет ППП STATISTICA, графики деформируются, их трудно читать. А поскольку в каждом компьютере есть Excel, эти графики читаются нормально. Поэтому проценты надо перенести в Excel и там построить графики.

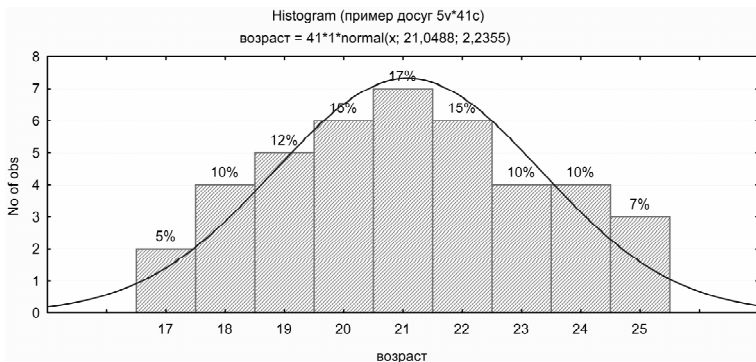


Рис. 1. Распределение респондентов по возрасту, %.

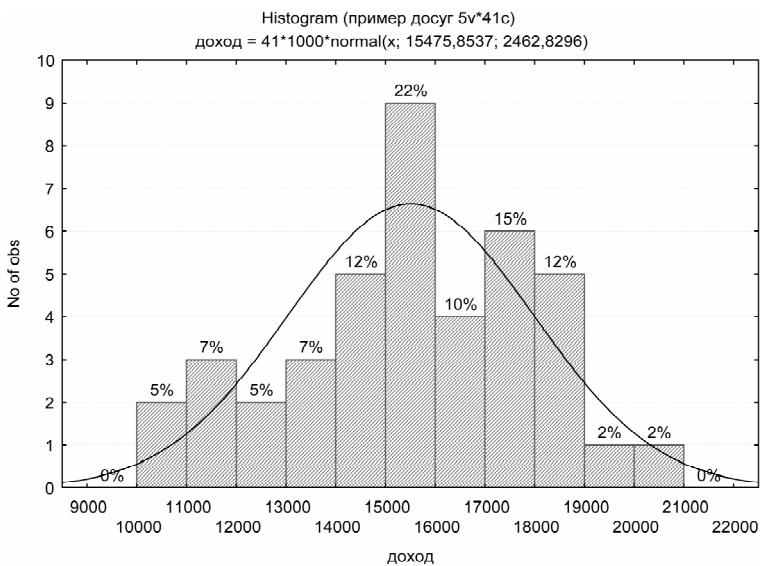


Рис. 2. Распределение респондентов по доходу, %.

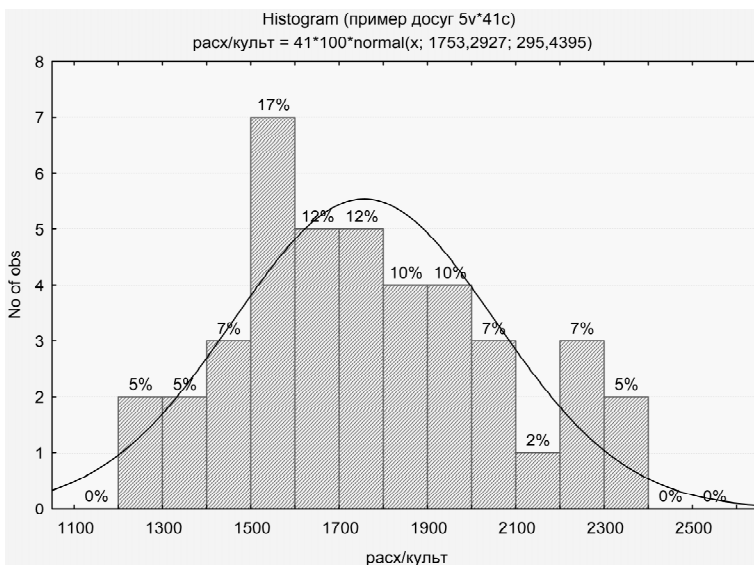


Рис. 3. Распределение респондентов по расходам на культуру, %.

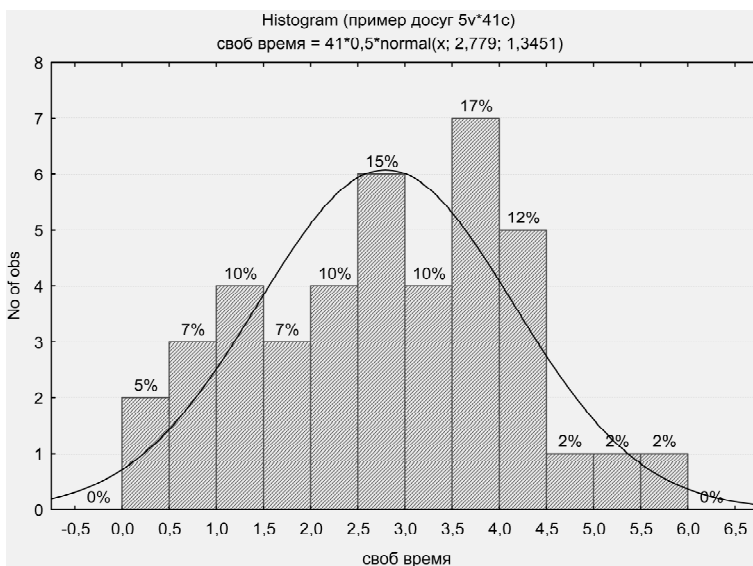


Рис. 4. Распределение респондентов по величине свободного времени, %.

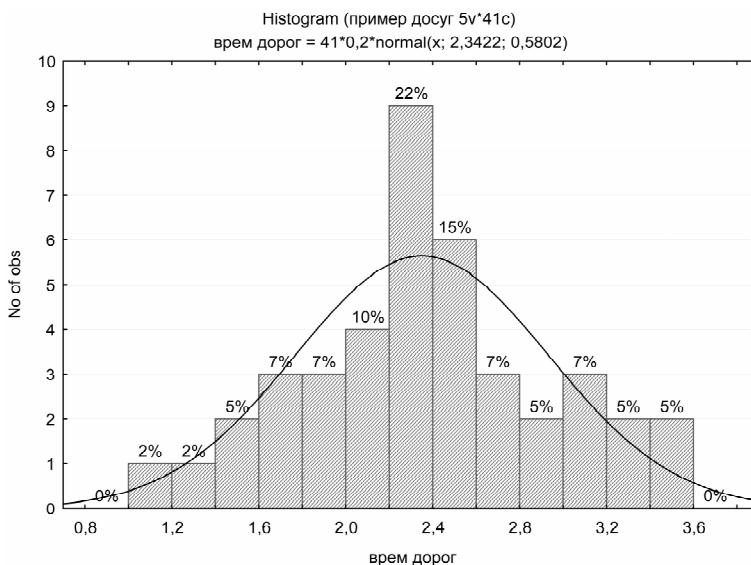


Рис. 5. Распределение респондентов по затратам времени на дорогу, %.

Итак, мы рассмотрели процедуры и команды построения гистограмм. Это требуется не только для наглядного представления о выборочной совокупности, но и для того, чтобы суммировать число ответов на вопросы анкеты. На рис. 5 показано, что 2% опрошенных затрачивают на дорогу от 1,0 до 1,2 часа в день. Столько же респондентов затрачивают не дорогу от 1,2 до 1,4 часа. 5% опрошенных затрачивают от 3,4 до 3,6 часа. Эти данные можно вносить в анкеты, чтобы было видно, сколько человек дали тот или иной ответ.

Обычно студенты, изучавшие информатику и Excel, говорят, зачем осваивать STATISTIC, графики можно получить и с помощью пакета Excel. Дело в том, что для Excel сначала надо сгруппировать все данные в некоторые группы, а затем строить график. ППП STATISTICA сам группирует сотни, тысячи данных в 9 – 11 интервалов и строит график. Число интервалов мы задаем. Этот пакет удобнее, он специально создан для обработки больших массивов данных. Мы экономим время.

Суммарные ответы можно получить и при помощи одномерных таблиц. Это мы рассмотрим далее, когда будем осваивать таблицы.

Круговые диаграммы

1. Курсор на Graphs. Щелчок мышью. На экране меню.
2. Курсор на 2D Graphs. На экране подменю.

3. Курсор на Pie Charts. (Значок – круглая диаграмма). Щелчок мышью. На экране окно с командами. Если стоит Quick, то мы можем установить, сколько интервалов нас интересует. По умолчанию стоит 10 против Categories. Мы стрелками уменьшим до 5, поскольку у нас небольшое число наблюдений.
4. Курсор на Variables. Щелчок мышью.
5. Курсор на Select All. Активизируем (можно курсором) все пять признаков.
6. ОК.
7. Курсор на Advanced. На экране окно с командами. На белом поле курсором активизируем Text and Percent. На графике появятся проценты по каждой доле и номер этой доли.
8. ОК.
9. На экране появится пять круговых диаграмм с надписями: проценты и указание, какой доле они принадлежат.

Исследование вариации признаков

Вариацией называются колеблемость, многообразие, изменчивость величины признака у единиц совокупности. Если мы взглянем на таблицу 1, то увидим, значения признаков отличаются один от другого. Правда, есть и совпадения, но если бы все признаки были бы одинаковыми, то не надо было бы проводить обследования, опросы; достаточно было бы изучить одну единицу, один объект, одного человека и картина была бы ясна. Но в жизни все индивиды отличаются друг от друга. Возраст, заработная плата, виды получаемых доходов, размеры жилища, количество детей в семье, читаемые книги и т.п., – все эти показатели не одинаковы у разных людей. Точно также различаются друг от друга вузы, предприятия, города, страны и пр. При этом, одни индивиды имеют близкие значения количественных показателей, другие очень отличаются один от другого. Например, возраст новобранцев в одной роте может быть совершенно одинаков – все одного года призыва. Возраст студентов заметно отличается даже среди однокурсников, а в целом он колеблется, например, от 17 лет у первокурсников до 25 лет у пятикурсников (есть студенты, которые поступили в вуз на два года позже своих одноклассников). На предприятии же возраст работающих может колебаться от 16 до 75 лет. В этом случае говорят о вариации признаков. *Показатели вариации являются важными при решении задач описания объектов.*

С точки зрения управления, считается, что чем больше вариация, размах значений признака, тем легче изменить этот признак, легче им управлять. Напротив, при малых коэффициентах вариации необходимы большие затраты усилий, чтобы изменить признак в большую или меньшую сторону.

Вариация порождается комплексом экономических, социальных и иных условий, которые вызывают различия в величине тех или иных свойств людей, предприятий, городов и стран. Например, доходы зависят от квалификации работников, от профессии, от района проживания и множества других причин и условий. По степени вариации можно судить о многих сторонах процесса развития изучаемых явлений и объектов, об однородности совокупности, устойчивости индивидуальных значений признака и др.

Меры вариации.

Меры вариации делятся на абсолютные и относительные. К абсолютным относятся размах вариации, среднее линейное отклонение, дисперсия и стандартное (среднее квадратическое) отклонение. К относительным — коэффициенты осцилляции, вариации, относительное линейное отклонение и др.

В пакете STATISTICA предусмотрены такие меры вариации: Minimum and maximum, Variance, Standard Deviation. На основе Minimum and maximum нетрудно подсчитать размах вариации как их разность.

Variance (дисперсия) представляет собой средний квадрат отклонений индивидуальных значений признака от их средней арифметической величины.

Standard Deviation (стандартное отклонение) есть корень квадратный из дисперсии.

Обычно *стандартное отклонение* используют одновременно со средней арифметической. Средняя арифметическая указывает на среднюю величину признака, а стандартное отклонение характеризует степень разброса, вариации, отклонения индивидуальных значений признака от средней. Дисперсия применяется при подсчете различных статистических показателей, например, коэффициентов корреляции и др.

Рассчитаем эти три показателя вариации с помощью пакета STATISTICA.

Команды:

1. Курсор на Statistics. Щелчок мышью. Появится подменю.
2. Курсор на Basic Statistics. Щелчок мышью. Появится очередное подменю.
3. Курсор на Descriptive statistics.
4. ОК. Появляется окно с множеством команд.
5. Курсор на Variables. На экране появляется окно с перечнем наших вариантов (признаков).
6. Курсором активизируем все, или курсор на Select All и щелчок мышью.
7. ОК. Возврат к окну с командами.
8. Курсор на Advanced. Щелчок мышью.

9. Помечаем галочками показатели, которые нам нужны: Minimum and maximum, Variance, Standard Deviation.
10. Standard Deviation
11. Variance
12. Minimum and maximum
13. Курсор на Summary. Щелчок мышью. На экране таблица с требуемыми показателями.

Таблица 3

Показатели вариации

	Минимум	Максимум	Дисперсия	Стандартное отклонение
Возраст, лет	17,00	25,00	5,00	2,236
Доход, руб.	10 500,00	20 550,00	6 065 530,00	2462,830
Расходы на культуру, руб.	1230,00	2400,00	87 285,00	295,440
Свободное время, час.	0,33	5,56	2,00	1,345
Время на дорогу, час.	1,12	3,45	0,3364	0,580

Как видно из таблицы, стандартные отклонения существенно отличаются один от другого, поскольку размерности признаков разные. Но нам надо сравнить, какие признаки более однородные, какие имеют большой разброс значений (вариацию). Для того, чтобы свести их к более сопоставимым значениям, подсчитывается, в частности, коэффициенты вариации. Подсчет идет по формуле (2):

$$V_{\square} = (\square_x / M_x) * 100\%, \quad (2)$$

где V_{\square} – коэффициент вариации,

\square_x – стандартное отклонение,

M_x – средняя арифметическая величина.

Рассчитаем коэффициенты вариации.

Стандартные отклонения отражены в таблице 3, средние арифметические – в таблице 2. Средние арифметические: возраст – 21,05 лет, доход – 15475,85 руб., расходы на культуру – 1753,29 руб., свободное время – 2,78 час., время на дорогу – 2,34 час.

Итоги расчетов сведем в таблицу 4.

Коэффициенты вариации (%)

	Средняя арифметическая	Стандартное отклонен	Коэффициент вариации, %
Возраст, лет	21,05	2,236	10,62
Доход, руб.	15 475,85	2462,830	15,91
Расходы на культуру, руб.	1753,29	295,440	16,85
Свободное время, час.	2,78	1,345	48,38
Время на дорогу, час.	2,34	0,580	24,78

Как видно, картина несколько изменилась; самое большое числовое значение стандартного отклонения имеет показатель признака «доход», а коэффициент вариации этого признака по величине находится на предпоследнем месте. Напротив, самые маленькие числовые значения имеют стандартные отклонения признаков «свободное время» и «время на дорогу», а коэффициенты вариации этих признаков самые численно большие. О чем это говорит? Самый большой разброс значений (вариацию) имеют последние два признака, доход же более стабильный признак. Что же касается дохода, то этот признак в обследованной совокупности студентов более однороден и более устойчив. С точки зрения управления, более легкому воздействию поддается признак «свободное время». Для этого надо изучить факторы, влияющие на величину свободного времени студентов. Но это уже следующая исследовательская задача.

Итак, мы рассмотрели задачу описания объекта. Более сложной задачей является задача сопоставления двух и более объектов. Но логика расчетов и процедуры и команды ППП STATISTICA одни и те же для обеих задач. Для описания объектов используются средние величины, а также меры вариации признаков. Для более углубленного изучения и понимания сути этих статистических показателей необходимо обратиться к учебникам по теории статистики, а также к книгам по математической статистике.

2. Задачи поиска причинно-следственных связей

Описание свойств и их количественного выражения — необходимый этап обработки и анализа первичной социологической информации; с описания начинается исследование любого объекта и не только в социологии. За описание объектов присуждают Нобелевские премии; например, это поощрение получили физики, описавшие явление радиоактивности, — супруги Кюри. А.Смитт, одним из первых описавший производство и обращение товаров, является основоположником классической школы в экономической теории. Важную роль сыграли описание, и классификация на этой основе, видов живых организмов в биологии.

Более сложной задачей, которую решают исследователи, является задача поиска причинно-следственных связей. Если решение первой задачи отвечает на вопрос: каково состояние объекта? Чем один объект отличается от другого? Решение второй задачи позволяет ответить на вопрос: почему объект оказался в таком состоянии и что надо сделать, чтобы изменить объект? Зная причины и воздействуя на них, можно изменять состояние объекта. Например, повысить успеваемость студентов или улучшить их профессиональную подготовку; уменьшить масштабы преступности и коррупции; улучшить демографическую ситуацию в стране; удвоить ВВП в России в ближайшие годы.

Вопрос: «почему?» преследует нас не только в обществе, но и в быту, в личной жизни. Например, почему я не получил пятерку на экзамене? Почему мой однокурсник сделал более успешную карьеру? Почему сорвалась обещавшая стать успешной сделка? Теория статистики и ППП STATISTICA помогают решать такого рода задачи и отвечать на вопрос: «почему?». Эти задачи играют важную роль в управлении; воздействуя на причину, мы можем получить нужные нам следствия.

Выявлению причинно-следственных отношений между двумя или большим числом свойств, явлений служат статистические таблицы, разработанные по двум или большему числу признаков.

Одним из инструментов поиска причинно-следственных связей является корреляционно-регрессионный анализ. Кроме этого, в статистике имеются и другие методы, с помощью которых решаются такого рода задачи — факторный, дисперсионный, латентный и др. виды анализа.

Прежде всего, поговорим о причинно-следственных связях, чтобы верно интерпретировать полученные таблицы или уравнения регрессии.

В мире все со всем связано; есть всеобщая связь явлений и процессов. В философии это обозначается категорией «детерминизм». Из нашего личного опыта мы знаем, что любое следствие имеет свою причину или набор причин; ничто не случается без причины. Причинно-след-

ственные отношения достаточно сложны и многоплановы. И наоборот, зная причины, мы обычно ожидаем соответствующее следствие. (Например, черная кошка).

Первая проблема в этом ряду, связанная с математико-статистическими методами анализа, это проблема *ложной корреляции*. Слово «**корреляция**» означает **связь** двух, или более, явлений. Например, доход (зарплата и др.) связан с квалификацией, возрастом и др. факторами. В этом случае говорят, что доход коррелирует с квалификацией, чем выше квалификация тем выше заработная плата. Но не всякая статистически установленная (в статистической таблице или в уравнении регрессии) связь двух, например, признаков свидетельствует о наличии причинно-следственной связи между этими признаками. Например, в одном из учебников по статистике приводится такой пример *ложной* корреляции (связи). Во время второй мировой войны запрещалось пользоваться радиоприемниками, чтобы кто-нибудь ненароком не стал слушать радиопередачи враждебного государства. Но всегда найдется несколько граждан, которые склонны совершать неправомерные, противоправные поступки. Так вот, в Англии была установлена корреляция между количеством незаконно используемых радиоприемников и числом душевно больных. Можно сказать, что это естественно, т.к. нормальные люди не совершают противоправных поступков. Но дело в том, что радиоприемники незаконно использовали одни люди, а душевными заболеваниями страдали совсем другие.

В качестве другого примера ложной корреляции можно использовать известный в статистике анекдот: все 100% людей, умерших по причине онкологических заболеваний, употребляли свежие огурцы. Следовательно, свежие огурцы опасны для жизни, они вызывают неизлечимые болезни. Черная кошка также относится к ложной корреляции.

Все это сказано не для того, чтобы вызвать недоверие к статистическим методам, а для того, чтобы предостеречь исследователей от легкого и бездумного отношения к результатам расчетов, от скоропалительных выводов. Статистический анализ должен дополняться содержательным, качественным анализом сущности изучаемых явлений. Поэтому проведение прикладного социологического исследования предполагает предварительное изучение той теории среднего уровня, в рамках которой проводится это исследование; прежде чем начинать собственное исследование, надо ознакомиться с результатами исследований предшественников.

Ложная корреляция является также следствием обычного предрассудка – что после этого, значит по причине этого. Например, опубликовано много книг о толковании снов; многие люди верят, что сон имеет пророческое значение, если увидеть то-то и то-то, то в реальной жизни

будет именно это. Сначала сон потом реальное событие. Сон — причина будущих счастливых или неприятных событий. Но философия учит — что после этого, не значит, что по причине этого. Бесспорно, причина всегда предшествует следствию; но день не является причиной наступающей после дня ночи; точно также, ночь не является причиной нарождающегося дня, хотя последовательность всегда соблюдается.

Другая проблема поиска причинно-следственных связей — многопричинность явлений в мире. Одна причина может вызывать несколько следствий, а одно следствие может порождаться многими причинами. Например, провал на экзамене может произойти из-за того, что студент работает, чтобы иметь деньги для оплаты обучения в вузе. Другой причиной может быть то, что студент месяц сидел ночами у постели тяжело больной мамы, устал физически и не смог подготовиться к экзамену. Одним словом, сколько провалов на экзамене, столько и их причин.

Далее. Одна причина может вызвать несколько следствий. Например, плохое материальное положение семьи вынуждает студента много работать для того, чтобы обеспечить семью и заработать на оплату учебы. Некоторые студенты оставляют институт, откладывают освоение хорошей профессии на будущее, в связи с этим не складывается карьера. Так что материальное положение семьи тормозит успешную карьеру и нарушает жизненные планы студента.

Но мы знаем множество примеров, когда плохое материальное положение семьи и надежда на лучшее будущее после окончания института выступают мощным стимулом упорного систематического труда, в результате чего небогатый студент добивается выдающихся успехов в науке, политике, бизнесе. Например, видный русский социолог А.П. Шапов родился в семье бедного сельского дьячка и крестьянки; образование получил в православных учебных заведениях; был приглашен заведовать кафедрой русской истории в Казанский университет; выдвинул ряд плодотворных социологических концепций. Другой пример, С.О. Макаров — видный флотоводец, вице-адмирал, океанолог, специалист в области непотопляемости и живучести кораблей — был сыном боцмана (должность мл. командного состава экипажа судна); скорее всего, чело- века не богатого. Ли Яккока — талантливый американский менеджер — эмигрант из Италии, в США попал в поисках лучшей жизни. Таким образом, неудовлетворительное материальное положение порождает два следствия — неудачу в жизни и огромный успех.

Еще одна проблема, связанная с причинно-следственными отношениями, это воздействие причин не только на следствие, но и друг на друга. Некоторые причины усиливают друг друга, другие причины подавляют друг друга. Например, чтобы зерно проросло, необходимы, как

минимум, влага, тепло и солнечный свет. Если один из этих факторов отсутствует, зерно не прорастает. Таким образом, необходимо сочетание ряда факторов-причин, которые усиливают друг друга, чтобы получить искомое следствие.

Другой пример. В социологии известно, что свободное время является простором для всестороннего развития личности. Для того чтобы заниматься физической культурой, посещать театр и концертные залы, читать книги, общаться с интересными людьми, черпать информацию из Интернета и др., необходимо свободное время. Но уже древние греки знали, что избыток свободного времени ведет к аморализму и порокам. Итак, свободное время можно рассматривать как одну из причин развития личности, но другие причины подавляют эту функцию свободного времени и обращают ее в свою противоположность (кто-то увлекается спиртным, кто-то — наркотиками, что ведет к деградации личности). Какие это причины? В одном случае это плохое воспитание в семье; в другом — дурное влияние улицы, дворовых компаний; в третьем — безработица; можно назвать и другие причины, которые подавляют функцию свободного времени как фактора духовного роста личности.

Иногда, людям свойственно видеть «причины», лежащие на поверхности событий, но при этом люди не видят истинных причин, поскольку они скрыты от нашего взгляда. Например, некоторые обвиняют своих ближних в неудачах, которые их преследуют, но при этом не понимают, что эти неудачи являются следствием социальных событий. Дефолты, приватизация, создание клана олигархов привели к обеднению массы людей, поэтому винить в этом ближних — показатель близорукости индивида. В этом случае говорят о *латентных* событиях (причинах). Латентное — значит не видимое невооруженным глазом, скрытое от наблюдателя. Наука позволяет увидеть латентные события. Статистика способствует этому видению.

Еще одно очень важное положение, особенно важное при интерпретации и понимании корреляции и других математико-статистических методов анализа эмпирических данных. Это понятие функциональных и вероятностных (стохастических) взаимозависимостей.

Функциональные зависимости это жестко детерминированные зависимости двух или более свойств предметов, явлений, признаков. Если один признак увеличивается на единицу, то другой признак всякий раз увеличивается (уменьшается) на строго определенную величину. Например, увеличивая длину стола в два раза, мы получим увеличенную площадь ровно в два раза, не в 1,9 и не в 2,1 раза, а в два раза. Если увеличим и длину, и ширину стола в два раза, площадь стола увеличится в четыре раза, не в 3,9 и не в 4,2 раза.

Корреляционная связь — *вероятностная, стохастическая*. Если мы увеличим возраст рабочего в два раза — с 20 лет до 40 лет, то его зарплата увеличится не в два раза. В общей массе рабочих в возрасте 20 лет и 40 лет зарплата будет выше на какую-то величину у лиц старшего возраста. Но у отдельного индивида зарплата может остаться неизменной, а у некоторых она может даже уменьшиться, у третьих — увеличиться более чем в два раза.

Иногда в газетах можно прочесть, что лица, не имеющие высшего образования, имеют более высокую зарплату в сравнении с лицами, окончившими вузы. Здесь мы имеем стохастическую, вероятностную зависимость. У отдельных индивидов, получивших высшее образование, зарплата может быть выше, ниже или точно такой же, как и у работников со средним образованием. Но если мы возьмем тысячу работников с высшим образованием и тысячу работников со средним образованием, то 80% работников, завершивших высшее образование, будут иметь более высокие доходы, а 20% лиц с высшим образованием будут иметь либо такую же зарплату, как и лица со средним или более низким образованием, либо более низкую зарплату. Таким образом, корреляционная зависимость проявляется в большой массе событий, явлений. Нельзя подсчитать корреляционную зависимость между двумя, тремя, пятью объектами. Для этого надо обследовать десятки, сотни, тысячи объектов, опросить тысячи респондентов.

Все эти философские, общенаучные положения необходимо учитывать при построении таблиц, регрессионных моделей и при интерпретации результатов, полученных при анализе таблиц и моделей.

Начнем с простого, с одномерных таблиц (но они относятся к описательным задачам).

Одномерные таблицы

Команды:

1. Курсор на Statistics. Щелчок мышью. Появится подменю.
2. Курсор на Basic Statistics/ Tables. Появится подменю.
3. Курсор на Frequency tables.
4. ОК. Появится окно с множеством команд.
5. Курсор на Variables. Появится окно; активизируем все признаки.
6. ОК
7. Курсор на Advanced. Щелчок мышью. Появится окно.
8. Курсор на No. Of exact intervals. В окне напротив ставим точку. А цифру 10 (количество интервалов по умолчанию) заменяем при помощи стрелок на цифру 5. Иначе говоря, устанавливаем число интервалов, в которые сгруппируются наши 41 анкет.
9. Курсор на Summary. Щелчок мышью. Появятся 5 таблиц.

Таблица 5

Распределение респондентов по доходам

Интервалы, руб.	Число событий	Накопленные частоты	Проценты	Накопленные проценты
9243,750<x<=11756,25	4	4	9,75610	9,7561
11756,25<x<=14268,75	8	12	19,51220	29,2683
14268,75<x<=16781,25	16	28	39,02439	68,2927
16781,25<x<=19293,75	11	39	26,82927	95,1220
19293,75<x<=21806,25	2	41	4,87805	100,0000
Всего	41	41	100,00000	100,0000

В первой колонке таблицы (см., например, табл. 5) указано, сколько анкет (ответов) оказалось в этом интервале. Во второй колонке даются накопленные частоты, т.е. все ответы суммируются; это называется кумулятой. В третьей колонке указаны проценты числа ответов, попавших в интервал. В четвертой колонке даются накопленные проценты. Приведем отредактированную в WORD таблицу.

Как показано в таблице, наибольшее число респондентов имеют среднемесячный семейный доход от 14268,75 руб. до 16781,25 руб., что составляет 39,02%, несколько более одной трети от всей выборочной совокупности. Накопленные частоты и проценты (кумуляты) применяются редко, поэтому эти колонки можно убирать, чтобы таблица не была перегружена лишними цифрами.

Приведем еще одну таблицу.

Таблица 6

Распределение ответов респондентов на вопрос: «Как Вы оцениваете социальную политику Правительства Москвы?»

	Число событий	Накопленные частоты	Проценты	Накопленные проценты
Положительно	240	244	16,58	16,58
Удовлетворительно	480	724	32,63	49,21
Не удовлетворительно	420	1144	28,56	77,77
Затрудняюсь сказать	320	1464	21,76	99,53
Не ответили	7	1471	0,47	100,00
Всего	1471	1471	100,00	100,00

В таблице 6 приведено распределение ответов респондентов, которых опрашивали в мае—июне 2005 г. по проблемам жилищного строительства в столице. Как видно, почти половина опрошенных одобряют социальную политику в столице (49,21% = 16,58 + 32,63). Проценты, которые приведены в 4-й колонке, показывают, сколько процентов респондентов дали тот или иной ответ. Это важно при подведении итогов опроса, для описания объекта. Выше мы приводили округленные цифры, которые сопровождали графики, а здесь даны более точные не округленные данные. Они вносятся в анкету, чтобы было видно, сколько ответов получено на тот или иной вопрос. Но, довольно часто, приходится округлять до целых чисел.

Одномерные таблицы получают при суммировании ответов сотен или тысяч анкет; они позволяют свести множество индивидуальных ответов к небольшому числу обобщенных, легко обозреваемых данных.

Это не относится к поиску причинно-следственных связей, но мы дали эту процедуру именно в этом месте, чтобы студенты приобрели навыки работы с таблицами и интерпретации полученных таблиц.

Одномерные таблицы некоторые социологи называют «линейка» или простое распределение.

Двумерные таблицы

Команды:

1. Курсор на STATISTIC. Щелчок мышью. Появится подменю.
2. Курсор на Basic Statistics/ Tables. Появится подменю.
3. Курсор на Tables and banners.
4. ОК. Появится окно.
5. Курсор на Specify tables (select variables). Щелчок мышью. Появится окно, в котором имеется 6 колонок, в которых указаны все варианты (признаки). Из их числа мы отбираем два признака, по которым строится таблица.
6. В первой колонке активизируем курсором (затемняем) первый признак, во второй колонке активизируем другой признак (по нумерации они могут быть любыми, например, 34 и 76 и т.д.). Выделим первый и второй (оценка социальной политики Правительства Москвы и удовлетворенность москвичей городскими дорогами).
7. ОК.
8. ОК. Появится окно с множеством процедур. В нижнем левом углу галочками в окнах отмечаем получение в таблице процентов. Либо проценты по столбцу, либо по строке, либо по отношению к общему числу опрошенных. Целесообразнее брать либо по столбцу, либо по строке. Редко бывает необходимость подсчитывать общий процент. Брать ли

- по столбцу или по строке это зависит от содержания задачи. Если это окно не появилось, надо курсор на Options и щелчок мышью.
9. Зададим сначала процент по строке. Курсором ставим галочку против «Percentages of row count». (Процент по столбцам задается командой «Percentages of column count»).
 10. Курсор на Summary. Щелчок мышью. На экране таблица (табл. 7).

Таблица 7

Взаимозависимость ответов на вопрос: «Как Вы оцениваете социальную политику Правительства Москвы» и «Удовлетворены ли Вы качеством городских дорог» (%)

Как Вы оцениваете социальную политику Правительства Москвы	Удовлетворены ли Вы качеством городских дорог			
	Очень удовлетворен	Удовлетворен	Трудно сказать	Не удовлетворен
Положительно	0,00%	45,45%	45,45%	9,09%
Удовлетворительно	11,54%	53,85%	34,62%	0,00%
Не удовлетворительно	0,00%	42,86%	28,57%	28,57%
Затрудняюсь сказать	33,33%	0,00%	66,67%	0,00%

Поскольку проценты взяты по строке, постольку мы должны анализировать именно по строкам. Среди тех респондентов, которые ответили, что они положительно оценивают социальную политику, 45,45% удовлетворены качеством московских дорог, 45,45% затруднились с ответом, и только 9,09% не удовлетворены качеством дорог. Среди тех, кто удовлетворительно оценивает социальную политику, 11,54% очень удовлетворены качеством дорог, 53,85% удовлетворены качеством дорог, 34,62% — затруднились дать какой-либо ответ, и *никто* не сказал, что *не удовлетворен* качеством дорог.

Среди тех, кто *не удовлетворительно* оценил социальную политику, нет ни одного, кто *очень* удовлетворен качеством дорог, 42,86% удовлетворены качеством дорог, а 28,57% затруднились с ответом, и **28,57%** считают, что качество дорог *не удовлетворительное*.

Таким образом, чем больше людей не удовлетворены социальной политикой, тем больше они не удовлетворены качеством дорог.

Казалось бы, что удовлетворенность или неудовлетворенность социальной политикой является **причиной** удовлетворенности или неудовлетворенности качеством московских дорог. Но неудовлетворенность социальной политикой это понятие абстрактное. Не может человек восхищаться

плохими дорогами, если он высоко ценит социальную политику Правительства Москвы. *Значит, эта таблица показывает зависимость по-видимости, а не по существу.* Эта таблица не дает нам истинного знания.

Построим таблицу иначе, в которой в качестве причины рассмотрим качество дорог, а удовлетворенность социальной политикой будет следствием.

Таблица 8

Взаимозависимость ответов на вопрос: «Как Вы оцениваете социальную политику Правительства Москвы» и «Удовлетворены ли Вы качеством городских дорог» (%)

Удовлетворены ли Вы качеством городских дорог	Как Вы оцениваете социальную политику Правительства Москвы			
	Положительно	Удовлетворительно	Не удовлетворительно	Затрудняюсь сказать
Очень удовлетворен	0,00	75,00	0,00	25,00
Удовлетворен	22,73	63,64	13,64	0,00
Трудно сказать	27,78	50,00	11,11	11,11
Не удовлетворен	33,33	0,00	66,67	0,00

В этой таблице мы поменяли местами ответы на два анализируемых вопроса, хотя можно было бы просто задать проценты не по строке, а по столбцу и результат был бы таким же.

Судя по данным этой таблицы, мы можем сказать, что плохие или хорошие дороги могут выступать причиной удовлетворенности или неудовлетворенности социальной политикой. *Значит, чтобы москвичи были довольны социальной политикой, надо строить хорошие дороги и поддерживать в хорошем состоянии уже построенные.* На наш взгляд, этот вывод ближе к истине, или, если хотите, является истиной.

Среди тех, кто очень удовлетворен качеством дорог, **75,00% удовлетворены** социальной политикой. И нет ни одного, кто был бы не удовлетворен социальной политикой. Напротив, среди неудовлетворенных качеством дорог одна треть положительно оценивают социальную политику, а *две трети считают социальную политику неудовлетворительной.*

Представим ту же таблицу, но так чтобы проценты были распределены по столбцам, поскольку «причина» – удовлетворенность дорогами помещена в верхнем заголовке. Помещать ли причину в верхнем заголовке или в боковом, это дело автора; от перестановки мест зависимости не меняются.

Таблица 9

Зависимость ответов на вопрос: «Как Вы оцениваете социальную политику Правительства Москвы» от ответов на вопрос: «Удовлетворены ли Вы качеством городских дорог» (%)

Как Вы оцениваете социальную политику Правительства Москвы	Удовлетворены ли Вы качеством городских дорог			
	Очень удовлетворен	Удовлетворен	Трудно сказать	Не удовлетворен
Положительно	00,00	22,73	27,78	33,33
Удовлетворительно	75,00	63,64	50,00	0,00
Не удовлетворительно	0,00	13,64	11,11	66,67
Затрудняюсь сказать	25,00	0,00	11,11	0,00

Таким образом, двумерные таблицы позволяют нам судить о причинно-следственных связях. Естественно, что на удовлетворенность социальной политикой влияет тысяча причин, поэтому надо строить много таблиц.

Таблицы строятся не только по двум признакам, но и по трем, четырем и большему числу признаков, но для этого надо иметь несколько тысяч анкет, чтобы было меньше пустых клеточек.

Оформление таблиц

В курсовой или в дипломной работе таблицы, которые мы получаем в STATISTICA, должны быть соответствующим образом оформлены в свете тех требований, которые приняты в издательствах. Посмотрите книгу или статью в журнале, в которых имеются таблицы, и вы увидите, как оформлены таблицы и, кстати, графики. Рассмотрим эту проблему подробнее.

В 2007 г. было проведено социологическое исследование по теме: «Общественное мнение москвичей о качестве электроснабжения квартир и офисов в столице». На примере данных, полученных в ходе опроса жителей Москвы, построим одну двумерную таблицу и оформим ее. Одномерные таблицы оформляются так же.

В анкете был следующий вопрос, ответы на который приведены ниже.

«Как Вы оцениваете электроснабжение своих домов?» «Очень хорошее» – 16,9%; «Хорошее» – 52,1%; «Среднее» – 25,0%; «Плохое» – 5,5%;

Не ответили – 0,5%. В STATISTICA была получена двумерная таблица распределения ответов на этот вопрос в зависимости от пола респондентов. При переносе ее из STATISTICA в WORD она приобрела следующий вид (табл. 10).

Таблица 10

Зависимость ответов от пола респондентов

	N=1220	Var40	Var40	Row
Count	0	3	1	4
Column Percent		0,46%	0,18%	
Count	1	129	77	206
Column Percent		19,55%	13,75%	
Count	2	339	297	636
Column Percent		51,36%	53,04%	
Count	3	161	144	305
Column Percent		24,39%	25,71%	
Count	4	26	41	67
Column Percent		3,94%	7,32%	
Count	7	2	0	2
Column Percent		0,30%	0,00%	
Count	All Grps	660	560	1220
Table Percent		54,10%	45,90%	100,00%

Если таблица будет опубликована в таком виде, или представлена заказчику в отчете, ее не поймут и информация не будет воспринята. Точно также и в курсовой или дипломной работе таблица должна быть понятной. Поэтому надо оформить название таблицы, верхние и боковые заголовки – названия строк и столбцов, указать единицы измерения. В колонках и строках, т.е. в ячейках таблицы целесообразно ограничиться данными, выраженными в процентах. Абсолютные данные (Count) дополнительной информации не несут, поэтому эти строки следует удалить. Таблица станет компактнее и будет легче читаться. Таким образом, в таблице останется 5 или 6 строк и два столбца (ответы женщин и мужчин). Затем, мы добавляем две строки сверху для заголовков и один столбец для боковика (бокового заголовка, обозначающего названия строк) (см. табл.11).

Таким образом мы получаем пустую таблицу, которая называется макетом таблицы, которую затем заполняем цифрами из вышеприведенной таблицы 10 (см. табл.11).

Распределение ответов на вопрос: «Как Вы оцениваете электроснабжение своих домов?» в зависимости от пола респондентов, %

	Пол респондентов	
	Женщины	Мужчины
Очень хорошее	19,55	13,75
Хорошее	51,36	53,04
Среднее	24,39	25,71
Плохое	3,94	7,32
Не ответили	0,76	0,18
Доля женщин и мужчин в выборке	54,10	45,90

Вот окончательный вид таблицы, которая должна фигурировать в отчете, в курсовой или дипломной работе.

Здесь все имеет значение: место слова «Таблица» – справа, выравнивание бокового заголовка по левому краю, верхнего заголовка (головки) – по центру, цифр в ячейках – по центру, названия таблицы – по центру. Знак процента – % – в ячейках не указывается; он указывается в названии таблицы после запятой. Сколько знаков должно быть после запятой? Может быть, как в приведенной таблице, два знака, может быть один знак (19,5), могут быть только целые величины (19). Некоторые заказчики не любят цифры, боятся их, поэтому у них рябит в глазах от обилия цифр; в этих случаях приводятся только целые числа. Но в любом случае надо округлять сотые и десятые доли так, чтобы получилось ровно 100 %.

Как видно, женщины в 1,5 раза чаще одобрительно относятся к качеству электроснабжения квартир. Мужчины, напротив, более критически оценивают качество этой услуги. Следует изучить причины этих различий в общественном мнении мужчин и женщин.

Последнюю строку в таблице (доля женщин и мужчин) приводить не обязательно, но авторы пакета STATISTICA по каким-то соображениям ее заложили в программу и она получается автоматически. Будете ли Вы ее оставлять или убирать, это личный выбор автора исследования и отчета.

Двумерные графики

Выше было показано, что одномерные таблицы могут быть представлены в виде гистограмм или круговых диаграмм. Точно также, двумерные таблицы можно представить в виде двумерных диаграмм (гистограмм).

Двумерные гистограммы

Команды:

1. Курсор на Graphs. Щелчок мышью. Появится меню.
2. Курсор на 3 D. Появится подменю.
3. Курсор на Bivariate Histograms. Щелчок мышью. Появится окно.
4. Курсор на Variables. Появится окно с признаками.
5. Активизируем курсором два признака для двумерной гистограммы.
6. ОК.
7. ОК. На экране изображение двумерного графика. Его можно поворачивать, чтобы наиболее наглядно видеть столбики.

Bivariate Histogram (мэси доp 1710 63v*1710c)

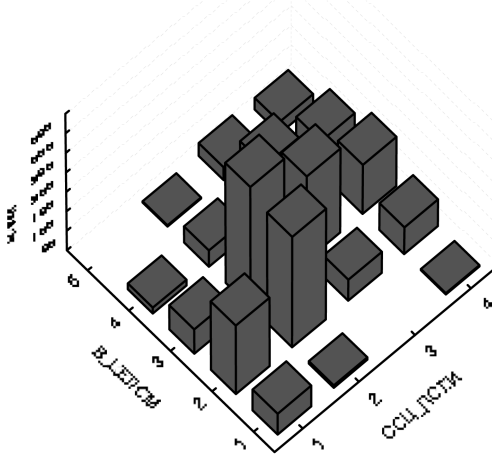


Рис. 6. Двумерная гистограмма зависимости ответов на вопрос анкеты от возраста респондентов, %

8. Вверху (STATISTICA) в самой нижней строчке, рядом с Normal Graph... есть кнопка со стрелкой. Курсор на эту стрелку. Щелчок мышью. Изменится вид графика. Сверху, снизу и слева появятся стрелки, при их помощи можно изменять положение столбиков.

9. ОК. Положение графика зафиксировано в удобной для вас позиции.

Мы привели три вида одного графика (рис. 6, 7, 8), чтобы показать возможности обзора графической взаимосвязи двух признаков с разных точек зрения. Если на втором рисунке не видны задние столбики, то на третьем рисунке видны эти, но не видны другие. На первом рисунке (рис. 6) мы попытались представить столбики как бы сверху. Этот ракурс позволяет охватить в целом все столбики.

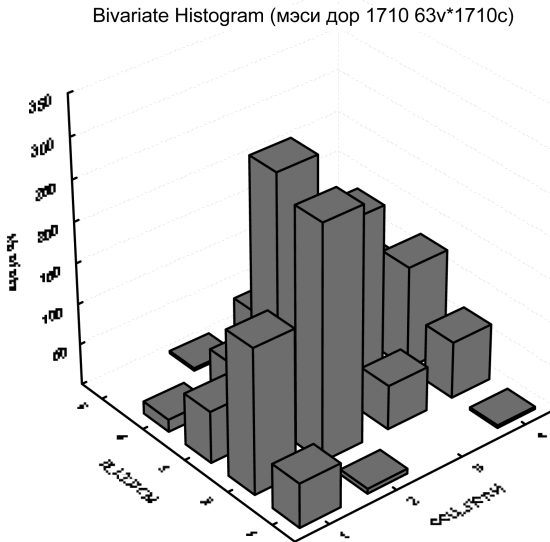


Рис. 7. Двумерная гистограмма зависимости ответов на вопрос анкеты от возраста респондентов, %

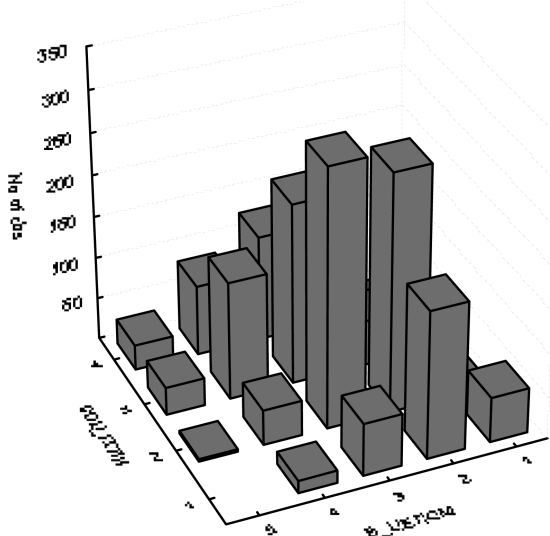


Рис. 8. Двумерная гистограмма зависимости ответов на вопрос анкеты от возраста респондентов, %

Отметим, что, если мы построили таблицу, то ее можно перенести в Excel и строить другого формата графики; в этом отношении, в Excel больше возможностей.

Корреляция

Работа с таблицами представляет некоторую трудность. Во-первых, в любой таблице есть колебания в процентах, поэтому всегда есть соблазн сказать, что есть зависимость двух или большего числа признаков; хотя эти колебания могут быть не очень большие. Во-вторых, в таблице достаточно много цифр, поэтому их утомительно читать. Есть более экономные методы выявления связей между двумя, тремя и даже девятью признаками. В частности, корреляционный анализ, который позволяет вместо десятка данных одной таблицы получить один коэффициент корреляции, что, бесспорно, нагляднее. А вместо десятка таблиц, в которых надо просмотреть сотни цифр, мы получаем десяток коэффициентов, что позволяет затрачивать меньше времени на их просмотр и анализ.

Выше мы говорили о характере корреляционной связи. Известно, что признаки, свойства предметов бывают количественные и качественные. Заработная плата, свободное время, возраст и др. являются количе-

ственными признаками; они имеют единицы измерения (час, рубль, метр, килограмм и т.д.). Удовлетворенность работой, пол респондента, его профессия и др. являются качественными признаками; они не имеют единиц измерения; например, удовлетворенность учебой или учебным заведением не имеет единицы измерения. Соответственно, есть методы подсчета корреляции для количественных признаков и для качественных признаков. В связи с тем, что социологи чаще используют качественные шкалы для оценки свойств изучаемых объектов, начнем с подсчета качественных коэффициентов корреляции (связи). К ним относятся коэффициенты Чупрова, Крамера, Спирмена, Кендэлла.

Любые коэффициенты корреляции колеблются от -1 до $+1$. Если мы получили значение коэффициента корреляции равным $+1$ или -1 , это означает, что мы имеем дело с *функциональной* зависимостью. Если получен коэффициент равный 0 , это означает полное *отсутствие* какой-либо связи между признаками. Вероятностные, стохастические связи никогда не могут измеряться коэффициентами равными $+1$ или -1 . В социологии вообще редки значения коэффициентов $0,9$ или $0,8$; чаще всего эти значения будут $-0,5$ или $+0,5$. Если значение коэффициента будет $0,09$, $0,15$, то говорят об очень незначительной связи признаков или об отсутствии этой связи. В социологии значимыми коэффициентами являются коэффициенты равные от $0,25$ до $0,75$. Коэффициенты близкие к 1 вызывают недоверие. Они вообще крайне редки.

Как интерпретировать плюс или минус при численном значении коэффициента корреляции? Плюс означает положительную связь — при увеличении одного признака второй признак также увеличивается. Например, с возрастом растет квалификация и, тем самым, заработная плата.

Минус характеризует отрицательную связь. При увеличении одного признака другой уменьшается. Например, с увеличением числа пропущенных занятий уменьшается объем знаний студента. Чем больше занятий пропустил студент, тем ниже оценка на экзамене.

Итак, рассчитаем коэффициент корреляции (связи) качественных признаков — коэффициент Спирмена. По методам расчета он относится к коэффициентам ранговой корреляции. Для более полного понимания этого коэффициента, равно как и всего корреляционного анализа, следует обратиться к учебникам по теории статистики.

Коэффициент Спирмена

Команды:

1. Курсор на Statistics. Щелчок мышью. Появится меню.
2. Курсор на Nonparametrics. Появится подменю.

3. Курсор на Correlations (Spearman, Kendall tau, gamma). Активируем (затемняем).
4. ОК.
5. Курсор на Variables. Появится окно с перечнем показателей. Выделяем два или более признаков. Для начала выделим два. Один в первой колонке, другой во второй колонке (как и для парных таблиц).
6. ОК. Появится окно. В правом верхнем углу есть: Spearman. Щелчок мышью и на экране — таблица с указанием значения коэффициента связи Спирмена. Если исследователя интересует коэффициент Кендалла, то даем следующие команды.
7. На этом же подменю. Курсор на Advanced. Появится подменю. Отметим курсором Кендалла. Щелчок мышью и на экране появится значение этого коэффициента.

Выше мы строили двумерную таблицу взаимной зависимости удовлетворенности качеством дорог в столице и удовлетворенности социальной политикой Правительства Москвы. (См. таблицы 7, 8 и 9). Из таблиц видно, что качество дорог влияет на удовлетворенность социальной политикой. Подсчет величины этого коэффициента для характеристики корреляции (связи) двух признаков, что было отражено в двумерной таблице, дал нам возможность получить численное значение коэффициента. Он равен +0,475. Для социологии это довольно значимая величина. Значит, мы уверенно можем говорить о причинно-следственной связи между удовлетворенностью дорогами и удовлетворенностью социальной политикой Правительства Москвы.

Как видно, вместо таблицы, в которой 16 цифр, получена одна цифра +0,475. Иначе говоря, мы сжали информацию в 16 раз. Очевидно, что коэффициент корреляции в более компактной форме дает нам ту же информацию, что и таблица с множеством цифр.

Если в нашей анкете есть ряд факторов (причин), которые согласно нашим гипотезам влияют на следствие, то можно выделить в одном столбике все эти факторы, а в другом столбике выделить только следствие; тогда мы получим в итоговой таблице все коэффициенты Спирмена (Кендалла), отражающие зависимости следствия от всех выделенных факторов. Из этой итоговой таблицы мы выделяем для дальнейшего анализа наибольшие значения факторов, а наименьшие факторы откладываем в сторону. Хотя нас может заинтересовать, почему какой-либо фактор имеет низкий коэффициент Спирмена, хотя по нашим гипотезам он должен быть более значимым. Это тоже может быть поводом для дальнейших размышлений.

Еще один метод выявления причинно-следственных связей — **корреляционно-регрессионный анализ**. *Корреляция* в переводе на русский язык

означает *связь*. Часто мы можем ее интерпретировать как причинно-следственную связь, но это не всегда. Иногда корреляционная связь и причинная не совпадают. И, тем не менее, установление корреляционной связи наталкивает на размышления и становится поводом для дальнейшего поиска действительных причин, влияющих на изменение изучаемого признака. Мерой интенсивности связи выступает *коэффициент корреляции*. Он колеблется от -1 до 0 и +1. Чем ближе он к +1, тем сильнее связь. Отрицательное значение коэффициента корреляции означает обратную зависимость – с увеличением фактора («причины») результативный признак («следствие») уменьшается. Положительный коэффициент означает положительную связь – с увеличением фактора также увеличивается результативный признак.

Уравнение регрессии может быть линейным и степенным. *Линейное* уравнение на графике изображается прямой линией, направленной снизу слева вверх направо, при положительной связи; или сверху вниз слева направо, при отрицательной связи. *Степенное* уравнение может изображаться на графике в виде гиперболы, параболы и др.

Линейное уравнение регрессии имеет вид:

$$Y = a + vX, \quad (3)$$

где Y – результативный признак, зависимая переменная («следствие»);
 X – фактор, независимая переменная («причина»);
 a – значение результативного признака при нулевом значении фактора;
 v – тангенс угла, показывает угол падения или подъема прямой; показывает, на сколько единиц изменится результативный признак при изменении фактора на единицу.

Подсчитаем коэффициент корреляции и уравнение регрессии по данным таблицы 12.

Таблица 12

Распределение респондентов по пяти признакам

Возраст, лет	Доход, руб.	Расходы на культуру, руб.	Свободное время, час.	Время на дорогу, час.
17	10 500	1520	2,55	2,49
17	11 000	1530	3,98	2,46
18	12 000	1420	4,23	2,31
18	11 150	1540	4,12	2,25
18	15 220	1545	4,11	2,12
18	13 100	1510	3,85	3,21
19	11 320	1430	4,02	3,12
19	12 320	1490	4,01	2,84
19	15 200	1340	3,81	2,72

Продолжение таблицы

19	14 650	1600	5,56	2,61
19	15 110	1620	5,25	2,36
20	15 200	1330	4,01	2,45
20	14 000	1650	3,71	2,23
20	15 100	1680	3,62	2,12
20	13 540	1720	3,51	1,82
20	15 900	1730	2,21	1,73
20	16 100	1740	2,32	1,65
21	14 200	1810	2,41	1,44
21	17 120	1710	3,42	3,45
21	16 450	1820	3,32	3,14
21	14 500	1830	1,91	3,32
21	15 110	1230	2,23	2,81
21	16 050	1290	1,73	2,75
21	17 120	1920	2,55	2,41
22	13 000	1930	2,88	2,23
22	17 620	1950	2,76	2,53
22	17 700	1800	2,51	2,22
22	15 650	2020	3,21	1,84
22	18 000	2030	1,23	1,93
22	18 150	2210	1,01	1,45
23	18 250	2120	2,99	3,42
23	19 000	1810	0,99	3,11
23	16 220	1910	0,85	1,12
23	17 200	2210	0,64	1,22
24	14 220	1520	1,66	2,23
24	18 120	2230	1,22	1,71
24	16 000	2100	3,22	2,11
24	14 400	1630	0,44	2,22
25	18 910	2360	1,02	2,32
25	19 560	1650	0,33	2,12
25	20 550	2400	4,55	2,44

Мы специально привели в таблице признаки, которые измеряются в годах, рублях, часах, чтобы получить линейное уравнение регрессии, так как для качественных признаков (типа: 1-мужской, 2-женский пол; или 1-да, 2-нет и т.д.) линейное уравнение регрессии нельзя получить. Выше были подсчитаны коэффициенты ранговой корреляции Спирмена и Кендэла, но не подсчитывались уравнения регрессии.

Команды в ППП STATISTICA:

Регрессия

Команды:

1. Курсор на STATISTICS.
2. В появившемся подменю выделяем курсором Multiple Regression.
3. Щелчек мышью. Появится меню.
4. Курсор на Variables (зеленая стрелка; всегда означает выбор вариантов для обработки и анализа).
5. Щелчок мышью и появятся две колонки наших признаков.
6. Выбираем в левой колонке зависимый признак («следствие»), а правой колонке – независимый признак («причину»). Выберем, например, в левой колонке третий признак «траты денег на потребление культурных благ» а в правой колонке – второй признак «доход», полагая, что траты денег на культуру зависят от дохода.
7. ОК. На экране появился результат.

Посреди белого поля красными буквами и цифрой обозначен коэффициент корреляции. В белом поле он обозначен как $beta = 0,643$. Этот коэффициент говорит о том, что наблюдается довольно плотная связь между доходом и затратами на культуру.

Другой показатель, который нам необходим для подстановки в уравнение $y = a + vx$, обозначен в этой итоговой таблице как Intercept, он равен 559,711. Это коэффициент a , который означает, что при нулевом значении x (дохода) затраты на культуру равны 559,711 рублей (видимо, в месяц).

Итак, наше уравнение примет вид:

$$y = 559,711 + vx, \quad (4)$$

где y – затраты денег на культуру,

x – доход индивида.

В самой нижней строке мы читаем: Summary: Regression results (результаты регрессионного анализа).

Команды:

1. Курсор на Summary: Regression results.
2. Щелчок мышью и на экране появится еще одна картинка, на которой под надписью Beta стоит 0,6429 (при округлении последней цифры мы получаем 0,643, т.е. то же, что и выше).
3. На этой же картинке мы видим букву B , под которой мы видим две цифры: первая означает обозначает коэффициент a в уравнении $y = a + vx$. Он равен 559,711; то же значение, что и выше.

А под ним есть другая цифра, она равна 0,0771. Это есть коэффициент b в уравнении $y = a + bx$.

Этот коэффициент означает, что с увеличением дохода на 1 рубль расходы на культуру возрастают на 7,7 копеек.

Есть еще один путь получения уравнения регрессии.

Команды:

1. Курсор на Graphs. Щелчок мышью.
2. Появляется подменю.
3. Курсор на Scatterplots. Щелчок мышью.
4. Появляется еще одно подменю.
5. Курсор на Variables. Появятся две колонки с нашими признаками.
6. Выделяем в левой колонке вариант 2, а в правой колонке – вариант 3.
7. ОК. И на появившейся картинке мы видим, что X это признак 2 («причина»), а Y это признак 3 («следствие»). X – независимая величина, а Y – зависимая величина.
8. ОК.
9. На экране появился график. На нем изображена прямая, которая обозначает линейную положительную связь между анализируемыми признаками (доходом и затратами денег на культуру). С ростом доходов увеличиваются расходы на культуру.

Для нас представляет интерес уравнение регрессии, которое изображено над графиком.

$$Var3 = 559,711 + 0,0771x. \quad (5)$$

В уравнении вместо x записан вариант 3. Значит, с увеличение дохода на 1 рубль расходы на культуру возрастают на 7,7 копеек. А рост дохода на 1000 руб. приведет к увеличению трат на культуру на 77 руб. Соответственно, если доход вырастет на 10 000 руб., расходы на культуру увеличатся на 770 руб.

Итак, мы рассмотрели методы решения второй задачи – изучение причинно-следственных связей. Для более глубокого изучения этих проблем следует читать учебники по статистике.

3. Прогнозирование социальных процессов

Третий тип задач – прогнозирование социальных процессов. Для решения этих задач сначала рассмотрим обработку временных рядов (рядов динамики), а затем построение уравнений регрессии (подсчет тренда).

Процесс развития, движения социально-экономических явлений во времени в статистике принято называть **динамикой**. Для отображения динамики строят **ряды динамики** (хронологические, временные), которые представляют собой ряды изменяющихся во времени значений статистического показателя, расположенных в хронологическом порядке ¹.

Временной ряд – это последовательность измерений в последовательные моменты времени. Анализ временных рядов включает широкий спектр разведочных процедур и исследовательских методов, которые ставят две основные цели: (а) определение природы временного ряда и (б) прогнозирование (предсказание будущих значений временного ряда по настоящим и прошлым значениям). Обе эти цели требуют, чтобы модель ряда была идентифицирована и, более или менее, формально описана. Как только модель определена, вы можете с ее помощью интерпретировать рассматриваемые данные (например, использовать в вашей теории для понимания сезонного изменения цен на товары, если занимаетесь экономикой). Не обращая внимания на глубину понимания и справедливость теории, вы можете экстраполировать затем ряд на основе найденной модели, т.е. *предсказать его будущие значения* ².

Исследование динамики социально-экономических явлений, выявление взаимосвязи дают основание для прогнозирования – определения будущих размеров уровня изучаемого явления. Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится в будущем, т.е. прогноз основан на *экстраполяции* ³.

Экстраполяция в статистике – распространение установленных в прошлом тенденций на будущий период (экстраполяция во времени применяется, например, для перспективных расчетов населения); распространение выборочных данных на другую часть совокупности, не подвергнутую наблюдению (экстраполяция в пространстве), это прогнозирование неизвестных значений путем продолжения функций за

¹ Теория статистики: учебник. М.: Финансы и статистика, 1998. С. 344.

² <http://www.statsoft.ru/home/portal/glossary/GlossaryTwo/T/TimeSeries.htm>

³ Теория статистики: учебник / Под ред. проф. Р.А. Шмойловой. М.: Финансы и статистика, 1998. С. 400.

границы области известных значений. Методы экстраполяции во многих случаях сходны с методами интерполяции.

Теоретической основой распространения тенденции на будущее является известное свойство социально-экономических явлений, называемое *инерционностью*. На основе рядов динамики получаются весьма надежные прогнозы. При этом прогнозисты надеются, что общая тенденция развития явления в прошлом и настоящем не должна претерпевать серьезных изменений в будущем. Следует заметить, что, чем короче срок экстраполяции (*период упреждения*), тем более надежные и точные результаты дает прогноз.

В зависимости от того, какие принципы и исходные данные положены в основу прогноза, выделяют следующие элементарные методы экстраполяции: *среднего абсолютного прироста, среднего темпа роста и экстраполяцию на основе выравнивания рядов по какой-либо аналитической формуле*. Наиболее распространенным методом прогнозирования считается **аналитическое выражение тренда**.

Итак, надо выделить тренд, и прогнозировать по тренду. Тренд (**trend**) – это общее направление, тенденция развития рынка; это закономерность, характеризующая общую долгосрочную тенденцию в изменениях показателей временного ряда¹. Для повышения надёжности прогноза рекомендуется строить все возможные сценарии прогноза и рассчитывать доверительный интервал прогноза.

Для прогнозирования необходимо подсчитать *ценные* темпы роста. Рассмотрим основные категории (понятия), используемые при анализе временных рядов. Любой временной ряд (ряд динамики) показывает изменения по годам (месяцам, кварталам и др.). Для большей наглядности можно привести ряд динамики из «Московского статистического ежегодника». Например, ряд динамики, показывающий сумму вкладов населения в учреждениях Сбербанка РФ (на конец года, млн. руб.)²:

Таблица 13

1999 г.	2000 г.	2001 г.	2002 г.	2003 г.	2004 г.
31330	45924	69212	96034	167216	229695

¹ Бокс Дж., Дженкис Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1976.

² Московский статистический ежегодник. 2005: стат. сб. / Мосгорстат. М., 2005. С. 47.

Основные показатели динамики

На практике для количественной оценки динамики явлений широко применяют следующие основные аналитические показатели:

- абсолютные приросты;
- темпы роста;
- темпы прироста;
- коэффициенты роста.

Причем каждый из них может быть трех видов:

- базисный;
- цепной;
- средний.

Приведем формулы для их подсчета.

Абсолютный прирост характеризует размер увеличения или уменьшения изучаемого явления за определенный период времени. Он определяется как разность между данным уровнем (значением признака, например, ввод в эксплуатацию жилых домов по годам в 1990–2004 гг.) и предыдущим или первоначальным. Уровень (значение признака), который сравнивается, называется *текущим*, а уровень, с которым производится сравнение, называется *базисным*, так как является базой для сравнения. Если уровень (значение признака), который сравнивается, является *текущим*, а уровень, с которым производится сравнение, является *предыдущим*, в этом случае говорят о *цепных* показателях.

Темпом роста называется отношение данного уровня явления (значения признака) к предыдущему (*цепной* темп роста) или начальному (*базисный* темп роста), выраженное в процентах. Темп роста это наиболее распространенный показатель ряда динамики. Но поскольку нас интересует прогноз на основе ряда динамики, постольку мы покажем не только расчет темпов роста, но также темпов прироста, которые и позволяют прогнозировать социальные явления. Хотя и темпы роста также дают некоторое представление о тенденции развития изучаемого процесса.

Темп роста рассчитывается в процентах по формуле:

$$Tб = Ут * 100 / Уб, \quad (6)$$

где $Tб$ – базисный темп роста, %,

$Ут$ – уровень ряда в текущем периоде,

$Уб$ – уровень ряда в начальном периоде, принимаемым за базу.

В таблице 14 в третьем столбце приведены базовые темпы роста. $Tб1 = 2468 * 100 / 2256 = 109,40\%$; $Tб2 = 2481 * 100 / 2256 = 109,97\%$; и т.д. (см. табл. 14). Как видно, каждое значение ячейки делим на 2256 тыс. кв. м.

Если каждый уровень ряда сравнивается с предыдущим, то получают *цепные* показатели. В таблице 14 в пятой колонке (столбце) приведены базисные темпы роста (уровни каждого года мы делим на уровень 1990 года), а в шестой колонке приведены цепные темпы роста введенных кв.м.

$$T_{ц} = Y_m * 100 / Y_{np}, \quad (7)$$

где $T_{ц}$ – цепной темп роста, %,

Y_m – уровень ряда в текущем периоде,

Y_{np} – уровень ряда в предыдущем периоде.

$T_{ц1} = (2468/2256) * 100 = 109,40\%$; $T_{ц2} = (2481/2468) * 100 = 100,53\%$ и т.д.

База для сравнения должна выбираться обосновано, в зависимости от исторических и экономических особенностей изучаемого явления. В наши дни удобно в качестве базы брать 1990 год, поскольку в 1991 году завершилась эра социализма и Россия пошла по пути рыночных отношений. Но чтобы не удлинять временной ряд и сократить количество расчетов, за базу берут любой другой год. В таблице 14 показано, что в 1991 году количество введенных кв. м. увеличилось в сравнении с *базисным* 1990 годом на 109,40%, а в 1992 году на 109,97%.

Цепной темп роста в 1991 г. в сравнении с 1990 г. составил 109,40% и в 1992 году в сравнении с предыдущим периодом (1991 г.) был только 100,53%, а в 1993 г. в сравнении с 1992 г. составил 91,9%, т.е. снизился на 8,1%.

Таблица 14

Ввод в эксплуатацию жилых домов. Построено жилых домов на территории Москвы общей площадью тыс. кв. м.

Годы (1990–2004 гг.)	Тыс. кв.м.	Абсолютный прирост (тыс. кв. м.)		Темпы роста, %	
		Базисный	Цепной	Базисные темпы роста, %	Цепные темпы роста, %
Столбец (колонка) № 1	№ 2	№ 3	№ 4	№ 5	№ 6
1 (1990 г.)	256	-	-		
2	2468	212	212	109,40	109,40
3	2481	225	13	109,97	100,53
4	2280	24	-201	101,06	91,90
5	2474	218	194	109,66	108,51

Продолжение таблицы

6	2548	292	74	112,94	102,99
7	2856	600	308	126,60	112,09
8	3083	827	227	136,66	107,95
9	3113	857	30	137,99	100,97
10	3053	797	-60	135,33	98,07
11	3342	1086	289	148,14	109,47
12	3691	1435	349	163,61	110,44
13	4274	2018	583	189,45	115,80
14	4443	2187	169	196,94	103,95
15 (2004 г.)	4579	2323	136	202,97	103,06

Наряду с темпом роста рассчитывают показатель *темпа прироста*, характеризующий относительную скорость изменения уровня ряда в единицу времени. Темп прироста показывает, на какую долю (или процент) уровень данного периода или момента времени больше (или меньше) базисного уровня.

Темпом прироста называется отношение абсолютного прироста к *предыдущему* или *начальному* уровню, выраженное в процентах.

Темпы прироста вычисляются просто:

$$T_{пр б} = Tб - 100\%; \quad (8)$$

$$T_{пр ц} = Tц - 100\%; \quad (9)$$

где $T_{пр б}$ – базисный темп прироста,

$T_{пр ц}$ – темп прироста цепной,

$Tб$ – базисный темп роста,

$Tц$ – темп роста цепной.

В таблице 15 в колонках 3 и 4 приведены темпы прироста базовые и цепные.

В 1991 г. темп прироста *базисный* был $109,40 - 100 = 9,40\%$; в 1992 г. – $9,97\%$; в 1993 г. – $1,06\%$.

Цепной темп прироста в 1991 г. составил $9,40\%$; в 1992 г. $100,53 - 100 = 0,53\%$; в 1993 г. темп прироста был отрицательным $-8,1\%$. Иначе говоря, в 1993 г. ввод жилой площади в Москве уменьшился на $8,1\%$.

Таблица 15

**Ввод в эксплуатацию жилых домов. Построено жилых домов
на территории Москвы общей площадью тыс. кв. м.**

Годы (1990–2004 гг.)	Тыс. кв.м.	Темпы прироста, %		Коэффициент роста с переменной базой (цепной)
		Базисные темпы прироста,	Цепные темпы прироста,	
Столбец (колонка) № 1	№ 2	№ 3	№ 4	№ 5
1(1990 г.)	256	-	-	-
2	2468	9,40	9,40	1,0940
3	2481	9,97	0,53	1,0053
4	2280	1,06	-8,1	0,9190
5	2474	9,66	8,51	1,0851
6	2548	12,94	2,99	1,0299
7	2856	26,60	12,09	1,1209
8	3083	36,66	7,95	1,0795
9	3113	37,99	0,97	1,0097
10	3053	35,33	-1,9	0,9807
11	3342	48,14	9,47	1,0947
12	3691	63,61	10,44	1,1044
13	4274	89,45	15,80	1,1580
14	4443	96,94	3,95	1,0395
15 (2004 г.)	4579	102,97	3,06	1,0306

Для прогнозирования явлений используется *цепной коэффициент роста* или коэффициент роста с переменной базой. Если темпы роста выражены в виде простых отношений, т.е. база сравнения принимается за 1, а не за 100%, то полученные показатели называются *коэффициентами роста*. Их значения приведены в колонке 5 таблицы 15.

Этот коэффициент подсчитывается по формуле:

$$Kp = Tц / 100, \quad (10)$$

где Kp – коэффициент роста,

$Tц$ – темп роста цепной.

Или по формуле:

$$Kp = Y2 / Y1, \quad (11)$$

где $Y1$ – значение признака в текущем периоде,

$Y2$ – значение признака в предыдущем периоде.

Коэффициент роста цепной в 1991 г. составил: $109,40 / 100 = 1,0940$; в 1992 г. он был: $100,53 / 100 = 1,0053$; в 1993 г.: $91,90 / 100 = 0,9190$.

Или коэффициент роста цепной в 1991 г. составил: $2468 / 2256 = 1,0940$; в 1992 г. он был: $2481 / 2468 = 1,0053$; в 1993 г.: $2280 / 2481 = 0,9190$.

Именно коэффициенты роста используются для прогнозирования изменений динамического ряда в будущие годы. Кстати, так подсчитывается сложный процент вкладов в банк. Если процент равен 4, то первый вклад мы умножаем на 1,04; полученный итог также умножаем на 1,04 и так далее до конца периода, который нас интересует. Если мы вложили деньги на период 10 лет, то нетрудно подсчитать во сколько раз возрастет наш вклад через 10 лет. Предположим, мы вложили 100 руб. В 1-й год наш вклад возрастет в 1,04 раза и составит 104 руб. В конце второго года он будет равен 108,16 руб.; в конце третьего года – 112,49 руб.; за четвертый год он возрастет также в 1,04 раза и составит 116,98 руб. Через 10 лет наш вклад возрастет до 148,02 руб.

Для того, чтобы наш вклад удвоился через 10 лет, годовой процент должен быть приблизительно 7%.

При прогнозировании социальных явлений надо подсчитать этот средний коэффициент роста.

Таблица 16

Подсчет цепных коэффициентов роста

Годы	№№	Тыс. кв. м.	Цепные темпы роста, %	Коэффициенты роста с переменной базой (цепной)
1990	1	2256	-	-
1991	2	2468	109,40	1,0940
1992	3	2481	100,53	1,0053
1993	4	2280	91,90	0,9190
1994	5	2474	108,51	1,0851
1995	6	2548	102,99	1,0299
1996	7	2856	112,09	1,1209
1997	8	3083	107,95	1,0795

Продолжение таблицы

1998	9	3113	100,97	1,0097
1999	10	3053	98,10	0,9810
2000	11	3342	109,47	1,0947
2001	12	3691	110,44	1,1044
2002	13	4274	115,80	1,1580
2003	14	4443	103,95	1,0395
2004	15	4579	103,06	1,0306

Как видно в таблицах 14–16, будем ли мы рассчитывать темпы прироста на основе темпов роста или коэффициентов роста, результат будет одинаков.

Если темп роста и коэффициент роста всегда положительные числа, то темп прироста может быть положительным, отрицательным и равным нулю.

Сводной обобщающей характеристикой интенсивности изменения уровней ряда динамики служит *средний* темп роста, показывающий, во сколько раз в среднем за единицу времени изменяется уровень динамического ряда.

Необходимость исчисления среднего темпа роста возникает вследствие того, что темпы роста из года в год колеблются. Кроме того, средний темп роста часто следует определить в тех случаях, когда имеются данные об уровне в начале какого-нибудь периода и в конце его, а промежуточные данные отсутствуют.

Для иллюстрации приведем еще одну таблицу (табл. 17).

Такого рода средний темп роста можно исчислять, если положить в основу расчетов рост не в арифметической прогрессии, которая характеризуется постоянной разностью, а геометрической, характеризующейся постоянным отношением, называемым знаменателем прогрессии. Следовательно, вопрос состоит в том, чтобы найти этот знаменатель. Знаменатель геометрической прогрессии определяется делением последующего уровня прогрессии на его предыдущий ¹.

Выше было сказано, что средний темп роста вычисляется по формуле *средней геометрической* из цепных коэффициентов роста.

$$T = \sqrt[6]{K2/1 * K3/2 * K4/3 * K5/4 * K6/5 * K7/6} \quad (12)$$

¹ Теория статистики: учебник. М., 1998. С. 356–357.

Таблица 17

Темпы прироста дохода (пример условный)

Годы	№№	Доход (тыс. руб.)	Абсолют- ный прирост (тыс.руб.)	Базовые темпы прироста (%)	Цепные темпы прироста (%)	Цепные темпы роста (%)	Коэф- фициенты роста
1995	1	3					
1996	2	4	$4 - 3 = 1$	33,3	33,3	133,3	1,333
1997	3	3,5	$3,5 - 3 = 0,5$	16,6	-12,5	87,5	0,875
1998	4	5	$5 - 3 = 2$	66,6	42,8	142,8	1,428
1999	5	4	$4 - 3 = 1$	33,3	-20,0	80,0	0,8
2000	6	7	$7 - 3 = 4$	133,3	75,0	175,0	1,75
2001	7	5	$5 - 3 = 2$	66,6	-28,6	71,4	0,714
2002	8	7	$7 - 3 = 4$	133,3	40,0	140,0	1,40
2003	9	8	$8 - 3 = 5$	166,6	14,3	114,2	1,142
2004	10	6	$6 - 3 = 3$	100,0	-25,0	75,0	0,75
2005	11	9	$9 - 3 = 6$	200,0	50,0	150,0	1,5
2006	12	10	$10 - 3 = 7$	233,3	11,1	111,1	1,111

Использование ППП STATISTICA для подсчета средней геометрической

Рассчитаем этот показатель с помощью пакета STATISTICA.

Команды:

1. Курсор на Statistics. Щелчок мышью. Появится подменю.
2. Курсор на Basic Statistics. Щелчок мышью. Появится очередное подменю.
3. Курсор на Descriptive statistics.
4. ОК. Появляется окно с множеством команд.
5. Курсор на Variables. На экране появляется окно с перечнем наших вариантов (признаков).
6. Курсором активизируем столбец с коэффициентами цепных темпов роста и щелчок мышью.
7. ОК. Возврат к окну с командами.
8. Курсор на Advanced. Щелчок мышью.
9. Помечаем галочкой показатели, которые нам нужны: Geom. Mean.
10. Курсор на Summary. Щелчок мышью. На экране таблица с требуемым показателем.

В результате мы получили средний темп роста дохода (табл. 18) равный **1,115**. Далее мы экстраполируем рост доходов до 2010 года. Эта процедура осуществляется путем умножения значения дохода в каждом году, начиная с 2006 г., на 1,115. Так что в 2007 году можно ожидать доход в

размере 11,15 тыс. руб., в 2008 г. – 12,43 тыс. руб., в 2009 г. – 13,86 тыс. руб., в 2010 г. – 15,45 тыс. руб. (табл. 18).

Разумеется, наши прогнозные значения дохода не совпадут с фактическим доходом в эти годы. Они будут либо больше, либо меньше. Но таким образом, определяется область прогноза, т.е. реальные будущие значения дохода будут близки этим данным; они не будут равны 100 тыс. руб., или 700 руб., но возможно они будут больше или меньше на 0,5 или 3 тыс. руб. Для этого подсчитываются ошибки средней геометрической и берутся поправки. Но для этого надо читать литературу по теории статистики.

Таблица 18

Подсчет средней геометрической на основе цепных коэффициентов роста (пример, приведенный в таблице 17)

Годы	№№	Доход (тыс.руб.)	Цепные коэффициенты роста	Средняя геометри- ческая
1995	1	3		
1996	2	4	1,333	
1997	3	3,5	0,875	
1998	4	5	1,428	
1999	5	4	0,800	
2000	6	7	1,750	
2001	7	5	0,714	
2002	8	7	1,400	
2003	9	8	1,142	
2004	10	6	0,750	
2005	11	9	1,500	
2006	12	10	1,111	1,115
Экстраполяция роста дохода до 2010 г.				
2007	13	11,15		
2008	14	12,43		
2009	15	13,86		
2010	16	15,45		

Интерпретация полученных данных дело весьма сложное и трудное. Мы это видим на примере прогноза и планирования инфляции в нашей стране. Например, планируется инфляция в 8,5%, а фактически она колеблется до 15%. Но не 30%, не 2% и не 50%.

Очевидно, что на величину инфляции влияют многие факторы, например, цены на нефть и др. Так что, одного динамического ряда не

достаточно; надо строить множество динамических рядов. Но это выходит за пределы нашего пособия. Студенты должны усвоить технику обработки рядов динамики, а для решения проблем прогнозирования надо писать другое пособие. Желаящие изучить методы прогнозирования могут обратиться к специальным трудам (Бестужева-Лады И.В. и др.)¹.

Рассчитаем средний коэффициент роста для данных, приведенных в таблице 14. Разумеется, сначала в память компьютера вводим наши данные. Но это мы рассматривали в 1-й главе.

Рассчитаем этот показатель с помощью пакета STATISTICA.

Команды:

1. Курсор на Statistics. Щелчок мышью. Появится подменю.
2. Курсор на Basic Statistics. Щелчок мышью. Появится очередное подменю.
3. Курсор на Descriptive statistics.
4. ОК. Появляется окно с множеством команд.
5. Курсор на Variables. На экране появляется окно с перечнем наших вариантов (признаков).
6. Курсором активизируем столбец с коэффициентами цепных темпов роста и щелчок мышью. Таблица
7. ОК. Возврат к окну с командами.
8. Курсор на Advanced. Щелчок мышью.
9. Помечаем галочкой показатели, которые нам нужны: Geom. Mean.
10. Курсор на Summary. Щелчок мышью. На экране таблица с требуемым показателем.

Таблица 19

Подсчет средней геометрической

Годы (1990 2004гг.)	Тыс. кв. м.	Коэффициент роста с переменной базой	Средняя геометричес- кая, G
1 (1990 г.)	2256	-	
2	2468	1,0940	
3	2481	1,0053	
4	2280	0,9190	
5	2474	1,0851	

¹Дуброва Т.А. Статистические методы прогнозирования. М., 2003. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде WINDOWS. М., 2000.

Продолжение таблицы

6	2548	1,0299	
7	2856	1,1209	
8	3083	1,0795	
9	3113	1,0097	
10	3053	0,9807	
11	3342	1,0947	
12	3691	1,1044	
13	4274	1,1580	
14	4443	1,0395	
15 (2004 г.)	4579	1,0306	1,051890
2005 г.	4808		
2006 г.	5048		
2007 г.	5301		
2008 г.	5566		
2009 г.	5844		
2010 г.	6136		

Как видно, в 1997 г. в эксплуатацию было введено 3083 тыс. кв. м. жилья, а, согласно прогнозу, в 2010 г. будет введено **6136** тыс. кв. м., т.е. за 13 лет будет удвоение построенных жилых домов на территории Москвы. Это при 5,2% прироста в год.

Наиболее распространенным методом прогнозирования считают **аналитическое выражение тренда**.

Надо выделить тренд, и прогнозировать по тренду. Тренд (**trend**) – это общее направление, тенденция развития рынка; это закономерность, характеризующая общую долгосрочную тенденцию в изменениях показателей временного ряда¹. Для повышения надёжности прогноза рекомендуется строить все возможные сценарии прогноза и рассчитывать доверительный интервал прогноза. Для описания тенденции развития явления (тренда) в будущем широко используется регрессионный анализ. При этом изменение исследуемого показателя связывают лишь с течением времени; считается, что влияние других факторов косвенно проявляется через фактор времени. Иначе говоря, исследователь допускает, что все факторы сохраняют свое влияние неизменно и с одной и

¹ Бокс Дж., Дженкис Г. Анализ временных рядов. Прогноз и управление. М.: Мир, 1976.

той же силой. Бесспорно, это лишь допущение, которое иногда имеет место. Для более точных прогнозов надо изучать и влияние самих факторов, которые, согласно гипотезам, влияют на изучаемое явление. Например, можно предположить, что причины инфляции постоянны, поэтому можно считать только тенденцию во времени. Но можно также выявить факторы инфляции (например, излишняя эмиссия денег, взвинчивание цен монополистами и др.) и дополнить временной тренд уравнениями регрессии, отражающими влияние факторов на инфляцию. Это позволит получить более точные прогнозы, если это надо заказчику исследования.

При расчете трендовых моделей в качестве значений зависимой переменной выступают фактические уровни ряда — y , а в роли независимой переменной (x) — время t .

Уравнение линейного тренда имеет вид:

$$Y = a + vx, \quad (13)$$

где y — значения изучаемого признака,

x — время (t),

a и v — коэффициенты уравнения.

Тренд может быть и нелинейным.

В современных пакетах статистической обработки, в частности в ППП STATISTICA, имеется богатый арсенал преобразования данных и широкие возможности для графического изображения, в том числе в различных масштабах. Но мы ограничимся линейным трендом. Для глубокого уяснения сущности трендов необходимо изучить общую теорию статистики, а также ознакомиться со специальной литературой¹.

В главе о причинно-следственных связях была показана последовательность команд для подсчета параметров линейного уравнения регрессии. Здесь порядок расчетов тот же и последовательность такая же. Для подсчета параметров уравнения регрессии (тренда) формируется исходный массив данных, в котором заполняются два столбика — годы и ввод жилья (кв.м.). Годы вводятся как натуральные числа с 1-го по 15-й (например) (см. табл. 20).

Вводим эти данные в исходную матрицу в STATISTICA.

Выполняя команды, которые были приведены в главе о причинно-следственных связях, мы получаем следующее уравнение.

¹ См.: Дуброва Т.А. Статистические методы прогнозирования. М., 2008.

Массив исходных данных для подсчета параметров уравнения регрессии

Годы (1990–2004 гг.)	Тыс. кв. м.
1 (1990 г.)	2256
2	2468
3	2481
4	2280
5	2474
6	2548
7	2856
8	3083
9	3113
10	3053
11	3342
12	3691
13	4274
14	4443
15 (2004 г.)	4579

Команды в ППП STATISTICA:

Регрессия**Команды:**

1. Курсор на STATISTICS.
2. В появившемся подменю выделяем курсором Multiple Regression.
3. Щелчек мышью. Появится меню.
4. Курсор на Variables (зеленая стрелка; всегда означает выбор вариантов для обработки и анализа).
5. Щелчок мышью и появятся две колонки наших признаков.
6. Выбираем в левой колонке зависимый признак («следствие»), а правой колонке – независимый признак («причину»). Выберем, например, в левой колонке признак «количество кв. м.» а в правой колонке – признак «годы», полагая, что количество введенных в эксплуатацию кв. м. жилья зависит от времени.
7. ОК. На экране появился результат.

На основе этих результатов формируем уравнение тренда.

$$Y = 1798,2286 + 166,3964 * x = 1798,2286 + 166,3964 * 21 = 1798,2286 + 3494,3244 = 5292,553 \text{ тыс. кв. м.}$$

Подсчет будущих количеств введенных в эксплуатацию кв. м. жилой площади в столице в 2010 г. на основании среднего темпа роста показал, что в 2010 г. будет введено 6136 тыс. кв. м.

В 2010 г. будет введено в эксплуатацию от 5292 тыс. кв. м. до 6136 тыс. кв. м. Это можно рассматривать как *прогнозное поле*. При существующих темпах прироста будут достигнуты такие значения – от 5292 тыс. кв. м. до 6136 тыс. кв. м. Разумеется, условия могут измениться и в 2010 г. будет построено либо менее 5292 тыс. кв. м., либо более 6136 тыс. кв. м.

Приведем еще один метод получения уравнения регрессии, который был также использован выше в главе о причинно-следственных связях.

1. На основе введенного массива данных (годы и кв. м.) строим график.
2. Курсор на Graph.
3. В появившемся подменю курсор на Scatterplots.
4. В появившейся иконке – курсор на Variables. Появляется два столбика. В левом отмечаем курсором годы, а в правом – кв. м.
5. ОК.

На экране появляется график (рис. 9).

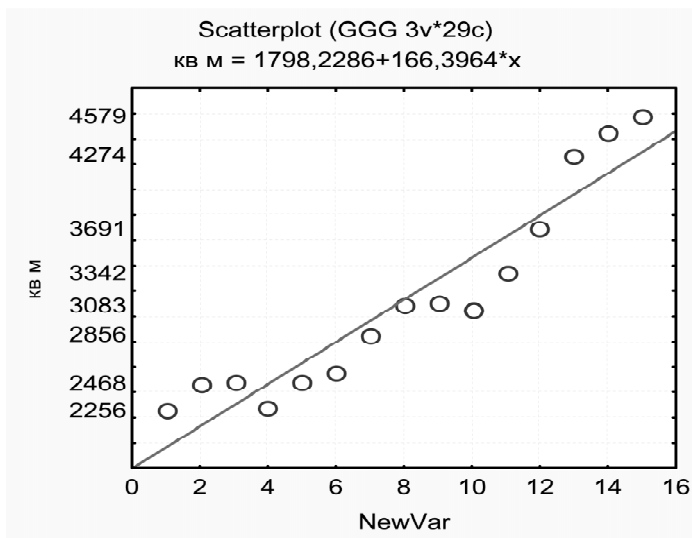


Рис. 9. Линейное уравнение регрессии

В верхней части графика приводится линейное уравнение регрессии (линейный тренд), которое имеет вид:

$$Y(\text{кв.м.}) = 1798,2286 + 166,3964 X, \quad (14)$$

где X – годы.

2010 год идет под номером 21, подсчитаем сколько жилья будет введено в 2010 г.

$$Y = 1798,2286 + 166,3964 * 21,$$

$$Y = 1798,2286 + 3494,3244,$$

$$Y = 5292,553.$$

Выше, при подсчете управления регрессии роста были получены такие же данные. Этот метод позволяет быстро оценить вид зависимости и его величину. А затем можно более глубоко анализировать эти зависимости, используя приведенный метод подсчета параметров уравнения регрессии.

Таким образом, использование обоих методов ППП STATISTICA (Multiple Regression и Scatterplots) дает одинаковые результаты.

Заключение

Предлагаемые студентам методические рекомендации в доступной форме знакомят с методами обработки и анализа первичной социологической информации. Но этот пакет можно использовать для работы с любой информацией, если она выражена в каком-либо количественном виде (руб., метры, штуки, время, порядковые или интервальные шкалы и др.). В этой методичке даются лишь первоначальные знания и навыки работы с пакетом, они далеко не исчерпывают возможности пакета STATISTICA. Цель методички – показать возможности компьютера и пакета, убедить студентов, что эти методы доступны, ими можно и нужно овладеть при систематических занятиях. Но одной этой методички недостаточно для полного овладения методами статистики и компьютерной технологией. В наши дни вышло в свет достаточно много книг, которые помогут студентам получить знания и навыки работы с этим пакетом, приобрести умения использовать статистические методы анализа социологических данных, стать квалифицированным социологом. Добавим, что, овладев пакетом STATISTICA, студенты без труда овладеют и пакетом SPSS.

Рекомендуемая литература

1. Лезина М.Л. Статистика: учебное пособие. М.: Российская Академия предпринимательства, 2008.
2. Теория статистики: учебник. М.: МЭСИ; Финансы и статистика, 1998.
3. Боровиков В. STATISTICA: Искусство анализа данных на компьютере. М.; СПб., 2003.
4. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде WINDOWS. М., 2000.
5. Бьюль Ф., Цёфель П. SPSS: искусство обработки информации. М.; СПб., 2002.
6. Вуколов Э.А. Основы статистического анализа: Практикум по статистическим методам. М., 2004.
7. Дубнов П.Ю. Обработка статистической информации с помощью SPSS. М., 2004.
8. Курицкий Б. Поиск оптимальных решений средствами Excel7.0. М.; СПб., 1997.
9. Методика и техника статистической обработки первичной социологической информации. М.: Наука, 1968.
10. Статистические методы анализа информации в социологических исследованиях. М.: Наука, 1979.
11. Пациорковский В.В., Пациорковская В.В. SPSS для социологов: учебное пособие. М., 2005.
12. Плис А.И., Сливина Н.А. Практикум по прикладной статистике в среде SPSS. Ч. 1. М., 2004.
13. Таганов Д. SPSS: Статистический анализ в маркетинговых исследованиях. М.; СПб., 2005.
14. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. М., 2003.
15. Хили Дж. Статистика: Социологические и маркетинговые исследования. СПб.; Киев, 2005.
16. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 1998.
17. Наследов А. SPSS: компьютерный анализ данных в психологии и социальных науках. СПб.: Питера, 2005.

Содержание

Введение	3
1. Задачи описания объектов	5
2. Задачи поиска причинно-следственных связей	20
3. Прогнозирование социальных процессов	41
Заключение	57
Рекомендуемая литература	58

Учебное издание

Журавлев Геннадий Терентьевич

**Использование ППП STATISTICA
для обработки и анализа
первичной социологической информации**

Учебно-практическое пособие (УПП)

Подписано в печать 02.12.2008
Формат бумаги 60x90 ¹/₁₆. Гарнитура «Ньютон7С»
Объем 3,75 усл. печ. л. Тираж 500 экз. Заказ