ЧАСТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«СТАВРОПОЛЬСКИЙ МНОГОПРОФИЛЬНЫЙ КОЛЛЕДЖ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к практическим занятиям и практической подготовке
для студентов специальности
38.02.07 Банковское дело
по дисциплине
«Проведение статистических наблюдений»

Ставрополь, 2024 г.

Методические указания составлены в соответствии с программой дисциплины «Технология продаж банковских продуктов и финансовое консультирование» и ФГОС для студентов по специальности 38.02.07 Банковское дело.

Составители (авторы): Данилов С.В.

СОДЕРЖАНИЕ

Практическая работа № 1

Сбор статистической информации. Использование статистических функций в табличном процессоре MS EXCEL (4 часа)

Цель работы

1. Изучить основные положения и определения, классификацию, вид и тип показателей, используемых при статистических измерениях.
2. Освоить возможности применения электронных таблиц EXCEL для проведения статистического анализа данных на персональном компьютере.

Краткая теория

Статистическое наблюдение

Статистическое наблюдение – планомерный, научно-организованный, систематический сбор данных о явлениях и процессах общественной жизни путем регистрации заранее намеченных существенных признаков.

В **задачи статистического наблюдения** входят: получение достоверной исходной информации, обеспечение полноты информации, проведение стат. наблюдения в короткие сроки.

Объект статистического наблюдения – совокупность, о которой должны быть собраны сведения. Он состоит из отдельных *элементов, единиц*. Характеристика объекта может быть получена лишь посредством характеристики его единиц.

Единица наблюдения – составной элемент объекта наблюдения, который является носителем признаков, подлежащих регистрации. Неточное определение единицы наблюдения влечет за собой погрешности, которые отрицательно сказываются на всем статистическом исследовании. Единицы наблюдения обладают множеством различных свойств, качеств, которые называются **признаками**.

Признак – характерная черта, свойство объекта или явления, которое может быть наблюдаемо или измерено.

Программа статистического наблюдения – перечень признаков, регистрируемых в процессе наблюдения.

Для записи ответов на вопросы программы наблюдения используется *формуляр наблюдения*, который представляет собой разграфленный лист бумаги, содержащий перечень вопросов программы, свободные места для записи ответов на них. Формуляр статистического наблюдения называют *формуляром, бланком, формой, опросным листом, анкетой*. То или иное название дается в зависимости от специфики организуемого наблюдения.

К формуляру наблюдения обычно дается инструкция.

Инструкция – совокупность разъяснений и указаний по программе статистического наблюдения. Она может быть представлена в виде отдельного документа или изложена на формуляре наблюдения.

Таблица 1.1

Классификация признаков

По характеру выражения	По способу измерения	По отношению к характеризуемому объекту	По характеру вариации	По отношению ко времени
Описательные (качественные), количественные	Первичные или учитываемые, вторичные или расчетные	Прямые (непосредственные), косвенные	Альтернативные, дискретные, непрерывные	Моментные, интервальные

Организация статистических работ

При организации статистических работ необходимо решить ряд вопросов: время проведения, продолжительность наблюдения, место наблюдения, установить критический момент наблюдения.

Критический момент – момент времени, по состоянию на который регистрируются данные.

Кроме того, при организации статистических работ необходим оргплан наблюдения.

Оргплан – документ, в котором зафиксированы все важные организационные мероприятия при проведении статистического наблюдения. В оргплане наблюдения указываются цель, объект, место, время, орган наблюдения, программа наблюдения, подготовительные мероприятия – подбор и обучение кадров, подготовка формуляров и пр. В ряде случаев, например, при переписях населения, проводится работа по разъяснению целей, задач, порядка проведения переписи.

Организационные формы

В отечественной статистике выделяют следующие формы:

– ***отчетность*** – форма, при которой сведения поступают в статистические органы от учреждений в виде отчетов об их деятельности; представляется в строго установленные сроки и является основным источником статистических сведений о народнохозяйственном и социальном развитии страны;

– ***специально организованное наблю***дение – наблюдение, организованное со специальной целью на определенную дату для получения данных, которые не вошли в отчетность, или для уточнения данных
отчетности;

– ***регистры*** – такая форма наблюдения, при которой факты состояния отдельных единиц совокупности непрерывно регистрируются. В регистре каждая единица наблюдения характеризуется совокупностью показателей.

Виды статистического наблюдения

Виды статистического наблюдения классифицируются по трем признакам:

а) охват наблюдением единиц совокупности, подлежащей статистическому исследованию;

б) систематичность наблюдения;

в) источник сведений, на основании которого устанавливаются
факты, подлежащие регистрации.

В зависимости от степени охвата изучаемого объекта статистическое наблюдение делится на:

– *сплошное* наблюдение, при котором обследованию подвергаются все без исключения единицы наблюдения;

– *несплошное* наблюдение, при котором обследованию подвергается только часть единиц наблюдения; основными видами такого наблюдения являются;

– *выборочное* наблюдение, при котором исследуется не вся совокупность, а лишь часть ее, отобранная по определенным правилам выборки и обеспечивающая получение данных, характеризующих всю совокупность в целом.

Применение выборочного наблюдения в статистической практике обусловлено рядом причин:

1) выборочное наблюдение проводится быстрее сплошного;

2) при выборочном наблюдении можно провести более глубокое и всестороннее исследование по более обширной программе;

3) к выборочному наблюдению прибегают в тех случаях, когда сплошное наблюдение не может быть осуществлено или нет смысла его проводить (например, если это связано с уничтожением объекта: испытание ткани на разрыв).

– монографическое обследование – детальное изучение и описание характерных в каком-либо отношении отдельных единиц наблюдения; проводится для выявления тенденций в развитии явления или для изучения и распространения передового опыта отдельных предприятий;

– метод основного массива, при котором исследуются наиболее крупные единицы наблюдения.

По признаку систематичности наблюдение может быть: – *текущим*, которое ведется систематически, непрерывно; – *прерывным*, которое подразделяется на:

– периодическое наблюдение, которое повторяется через определенные равные промежутки времени;

– единовременное наблюдение, которое проводится по мере надобности, без соблюдения строгой последовательности.

По источнику сведений наблюдение может быть:

– *непосредственное*, когда факты, подлежащие регистрации, устанавливаются лицами, проводящими наблюдение;

– *документированное*, при котором необходимые сведения берутся из соответствующих документов; – *опрос*, при котором сведения фиксируются со слов опрашиваемого.

Способы статистического наблюдения

В статистике применяют следующие способы наблюдения – опроса:

– *отчетный*, когда организации представляют статистические отчеты о своей деятельности в строго обязательном порядке;

– *экспедиционный*, когда специально обученные работники (счетчики) посещают каждую единицу наблюдения и сами заполняют формуляр наблюдения;

– *саморегистрация*, когда формуляр наблюдения заполняют сами опрашиваемые, но счетчики инструктируют их, проверяют правильность заполнения формуляра;

– *анкетный*, когда сбор статистических данных осуществляется с помощью анкет;

– *корреспондентский*, когда определенные лица ведут наблюдение за процессами и в установленные сроки сообщают результаты наблюдений статистическим органам.

Перечисленные виды, формы и способы статистического наблюдения применяются в зависимости от социально-экономических условий, существующих в стране; от особенностей объекта, который исследуется; от целей и задач, поставленных перед наблюдением; от программы наблюдения; от наличия кадров и средств, которыми располагают статистические органы; от срочности потребности в конкретных статистических данных.

Методические указания по работе со статистическими функциями в табличном процессоре microsoft excel [4]

Для проведения статистической обработки информации табличный процессор **Microsoft Excel** включает в себя программную надстройку

«Пакет анализа» и библиотеку из 78 статистических функций.

Для вставки какой-либо функции, находясь в нужной ячейке, нажмите на пиктограмму f_x (рис. 1.1) в строке формул или ленте Формулы.

Тем самым запускается 1 шаг **Мастера функций**. В нижней части окна появляется краткая справка о данной функции.

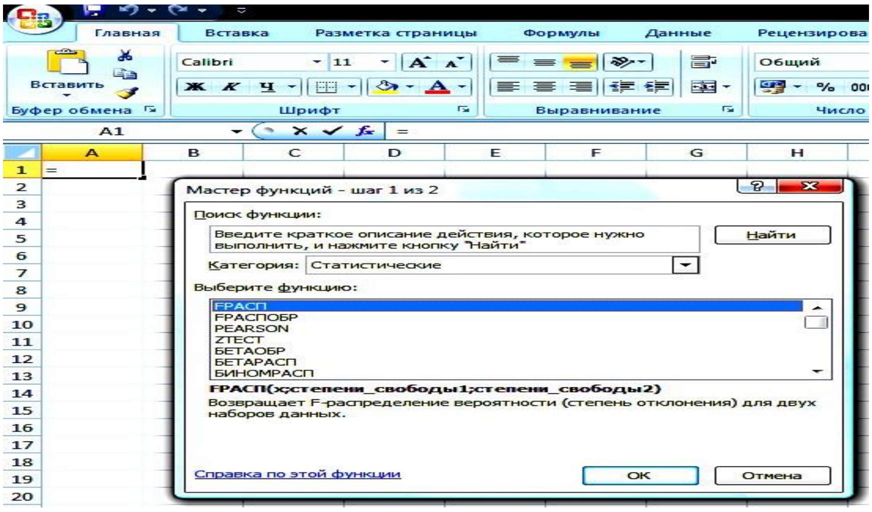


Рис. 1.1. Первый шаг Мастера функций

После выбора функции на 2-ом шаге появляется окно для заполнения аргументов функции (рис. 1.2). Если требуется дополнительная информация по этой функции с примерами ее использования, выберите в этом окне пункт **Справка** по этой функции.

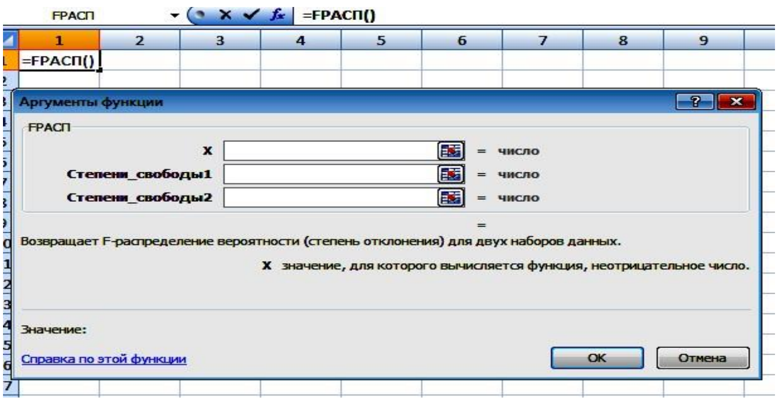


Рис. 1.2. Окно справки Мастера функций

В Excel существуют следующие **Статистические функции:**

СРОТКЛ	– Возвращает среднее абсолютных значений отклонений точек данных от среднего.
СРЗНАЧ	– Возвращает среднее арифметическое аргументов.
СРЗНАЧА	– Возвращает среднее арифметическое аргументов, включая числа, текст и логические значения.
БЕТАРАСП	– Возвращает интегральную функцию плотности бета- вероятности.
ХИ2РАСП	– Возвращает одностороннюю вероятность распределения хи-квадрат.

ХИ2ОБР – Возвращает обратное значение односторонней вероятности распределения хи-квадрат.

ХИ2ТЕСТ – Возвращает тест на независимость.

ДОВЕРИТ – Возвращает доверительный интервал для среднего значения по генеральной совокупности.

КОРРЕЛ – Возвращает коэффициент корреляции между двумя множествами данных.

КОВАР – Возвращает ковариацию, то есть среднее произведений отклонений для каждой пары точек.

КВАДРОТКЛ – Возвращает сумму квадратов отклонений.

ЭКСПРАСП – Возвращает экспоненциальное распределение.

FРАСП – Возвращает *F*-распределение вероятности.

ФИШЕР – Возвращает преобразование Фишера.

ПРЕДСКАЗ – Возвращает значение линейного тренда.

ЧАСТОТА – Возвращает распределение частот в виде вертикального массива.

ГАММАРАСП – Возвращает гамма-распределение.

СРГЕОМ – Возвращает среднее геометрическое.

РОСТ – Возвращает значения в соответствии с экспонен- циальным трендом.

СРГАРМ – Возвращает среднее гармоническое.

ГИПЕРГЕОМЕТ – Возвращает гипергеометрическое распределение

ОТРЕЗОК – Возвращает отрезок, отсекаемый на оси линией линейной регрессии.

ЭКСЦЕСС – Возвращает эксцесс множества данных.

НАИБОЛЬШИЙ – Возвращает *k*-ое наибольшее значение из множества данных.

ЛИНЕЙН – Возвращает параметры линейного тренда.

ЛГРФПРИБЛ – Возвращает параметры экспоненциального тренда.

ЛОГНОРМОБР – Возвращает обратное логарифмическое нормальное распределение.

ЛОГНОРМРАСП – Возвращает интегральное логарифмическое нормальное распределение.

МАКС – Возвращает максимальное значение из списка аргументов.

МАКСА – Возвращает максимальное значение из списка аргументов, включая числа, текст и логические значения.

МЕДИАНА – Возвращает медиану заданных чисел.

МИН – Возвращает минимальное значение из списка аргументов.

МИНА – Возвращает минимальное значение из списка аргументов, включая числа, текст и логические значения.

МОДА – Возвращает значение моды множества данных.

НОРМРАСП – Возвращает нормальную функцию распределения.

НОРМСТРАСП – Возвращает стандартное нормальное интегральное распределение.

ПИРСОН – Возвращает коэффициент корреляции Пирсона.

ПЕРСЕНТИЛЬ – Возвращает *k*-ую персентиль для значений из интервала.

ПУАССОН – Возвращает распределение Пуассона.

ВЕРОЯТНОСТЬ – Возвращает вероятность того, что значение из диапазона находится внутри заданных пределов.

КВАРТИЛЬ – Возвращает квартиль множества данных.

РАНГ – Возвращает ранг числа в списке чисел.

КВПИРСОН – Возвращает квадрат коэффициента корреляции Пирсона.

СКОС – Возвращает асимметрию распределения.

НАКЛОН – Возвращает наклон линии линейной регрессии.

НОРМАЛИЗАЦИЯ – Возвращает нормализованное значение.

СТАНДОТКЛОН – Оценивает стандартное отклонение по выборке.

СТАНДОТКЛОНА – Оценивает стандартное отклонение по выборке, включая числа, текст и логические значения.

СТАНДОТКЛОНП – Вычисляет стандартное отклонение по генеральной совокупности.

СТАНДОТКЛОНПА – Вычисляет стандартное отклонение по генеральной совокупности, включая числа, текст и логические значения.

СТОШYX – Возвращает стандартную ошибку предсказанных значений *y* для каждого значения *x* в регрессии.

СТЬЮДРАСП – Возвращает *t*-распределение Стьюдента.

ТЕНДЕНЦИЯ – Возвращает значения в соответствии с линейным трендом.

ТТЕСТ – Возвращает вероятность, соответствующую критерию Стьюдента.

ДИСП – Оценивает дисперсию по выборке.

ДИСПР – Вычисляет дисперсию для генеральной совокупности.

ВЕЙБУЛЛ – Возвращает распределение Вейбулла.

Пример использования функции ЧАСТОТА

Простейший способ создания распределения частот – использование функции ЧАСТОТА. Эта функция всегда возвращает массив, поэтому она используется в формулах массива, заполняющих диапазон ячеек. Она вычисляет частоту появления значений в интервале значений и возвращает массив чисел. Функцией ЧАСТОТА можно воспользоваться, например, для подсчета количества результатов тестирования, попадающих в интервалы результатов. Поскольку данная функция возвращает массив, она должна задаваться в качестве формулы массива.

Синтаксис

ЧАСТОТА(массив_данных;массив_интервалов)

Массив_данных – массив или ссылка на множество данных, для которых вычисляются частоты. Если аргумент «массив_данных» не содержит значений, функция ЧАСТОТА возвращает массив нулей.

Массив_интервалов – массив или ссылка на множество интервалов, в которые группируются значения аргумента «массив_данных».

Если аргумент «массив_интервалов» не содержит значений, функция ЧАСТОТА возвращает количество элементов в аргументе «массив_данных».

Замечания

- Функция ЧАСТОТА вводится как формула массива после выделения интервала смежных ячеек, в которые требуется вернуть полученный массив распределения.
- Количество элементов в возвращаемом массиве на единицу больше числа элементов в массиве «массив_интервалов». Дополнительный элемент в возвращаемом массиве содержит количество значений, превышающих верхнюю границу интервала, содержащего наибольшие значения. Например, при подсчете трех диапазонов значений (интервалов), введенных в три ячейки, убедитесь в том, что функция ЧАСТОТА возвращает значения в четырех ячейках. Дополнительная ячейка возвращает число значений в аргументе

«массив_данных», превышающих значение верхней границы третьего интервала.

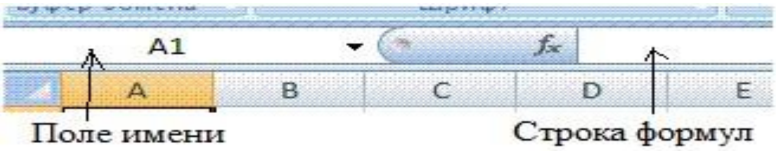
• Функция ЧАСТОТА игнорирует пустые ячейки и текст.

• Формулы, возвращающие массивы, должны быть введены как формулы массива.

В интервал ячеек A1:E25 введите следующие данные:

55	316	223	185	124
124	93	163	213	314
211	41	231	241	212
118	113	400	205	254
262	1	201	12	101
167	479	205	337	118
489	15	89	362	148
179	248	125	197	177
456	153	269	49	127
289	500	198	317	300
126	114	303	314	270
151	279	347	314	170
250	175	93	209	61
166	113	356	124	242
152	384	157	233	99
277	195	436	6	240
147	80	173	211	244
386	93	330	400	141
332	173	129	323	188
338	263	444	84	220
221	402	498	98	2
201	400	3	190	105
35	225	12	265	329
43	302	125	301	444
56	9	135	500	398

Интервалу ячеек A1:E25 присваивается имя «ДАННЫЕ». Для этого выделите этот интервал и в поле ИМЕНИ (крайняя левая часть строки формул) введите это имя.



В интервал ячеек G2:G11 в режиме Автозаполнения вводятся верхние границы интервалов. Для этого в ячейку G2 введите 50, а в ячейку G3–100. Эти ячейки выделяются и за квадратик заполнения протягиваются до ячейки G11.

Выделите интервал ячеек H2:H11 и в строке формул введите формулу:

=ЧАСТОТА(ДАННЫЕ;G2:G11).

Так как эта формула является формулой массива, то ее ввод заканчивается одновременным нажатием трех клавиш Ctrl+Shift+Enter. В результате эта формула будет иметь следующий вид:

{=ЧАСТОТА(ДАННЫЕ;G2:G11)}.

Аналогично в интервале ячеек I2:I11 введите формулу массива для получения распределения частот в процентном формате:

{=ЧАСТОТА(ДАННЫЕ;G2:G11)/СЧЁТ(ДАННЫЕ)}.

Функция СЧЁТ подсчитывает количество ячеек, содержащих числа, и количество чисел в списке аргументов. Функция используется для получения количества числовых ячеек в диапазонах или массивах ячеек.

И, наконец, строится гистограмма распределения частот в процентном формате.

Для построения диаграммы выполните следующую последовательность шагов:

1. Выделите интервал ячеек I1:I11;
2. В ленте «Вставка» выберите «Гистограмма»;
3. Из предложенного набора выберите, например, левую из верхнего ряда;
4. Для того, чтобы уменьшить зазор между столбцами диаграммы щелкните правой кнопкой по области построения (по любому столбику) и в появившемся меню выберите «Формат рядов данных»;

5. В появившемся окне, в пункте «Боковой зазор» переместите бегунок влево примерно до 10 %;

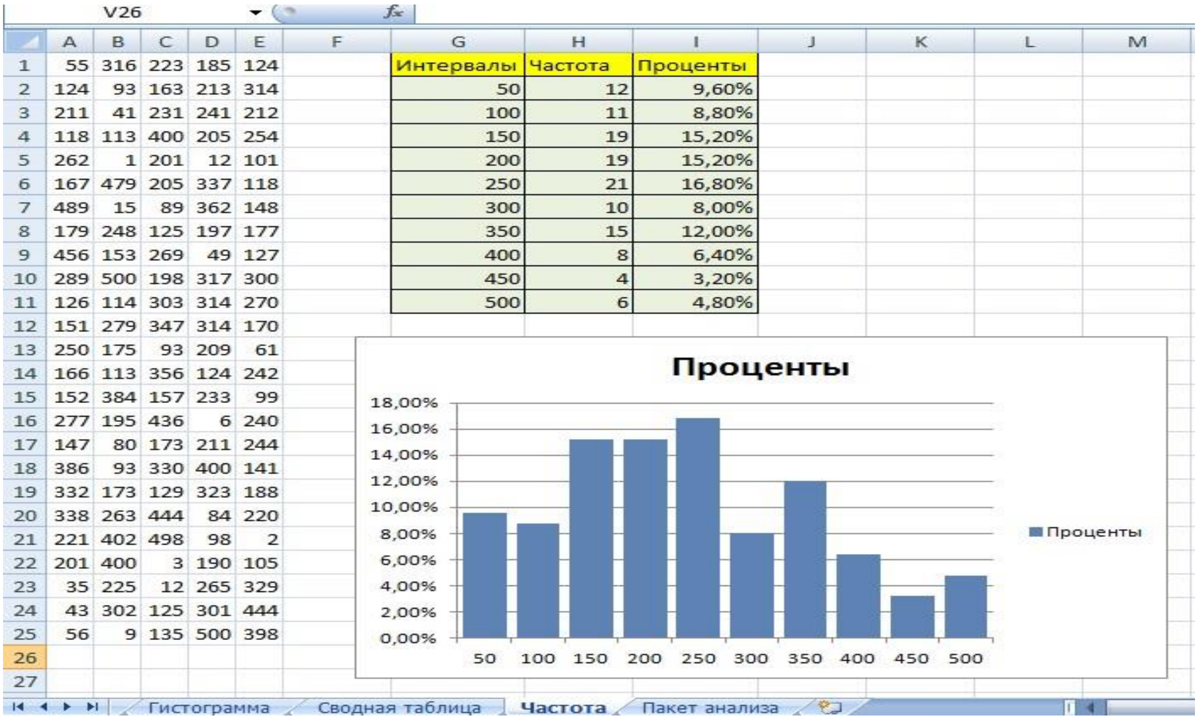


Рис. 1.3. Результат обработки статистических данных

6. Для того, чтобы изменить подписи на оси «Х», щелкните правой кнопкой по области построения, выберите пункт меню «Выбрать данные». В окне выберите в пункте (справа) «Подписи горизонтальной оси

(категории)» «Изменить», и в появившемся окне укажите мышью интервал ячеек G2:G11:

7. Легенду (проценты, справа) следует также убрать. Щелкните по ней правой кнопкой мыши и выберите УДАЛИТЬ.

То же самое распределение частот получаем с помощью надстройки Пакет анализа. Это надстройка находится в ленте Данные.

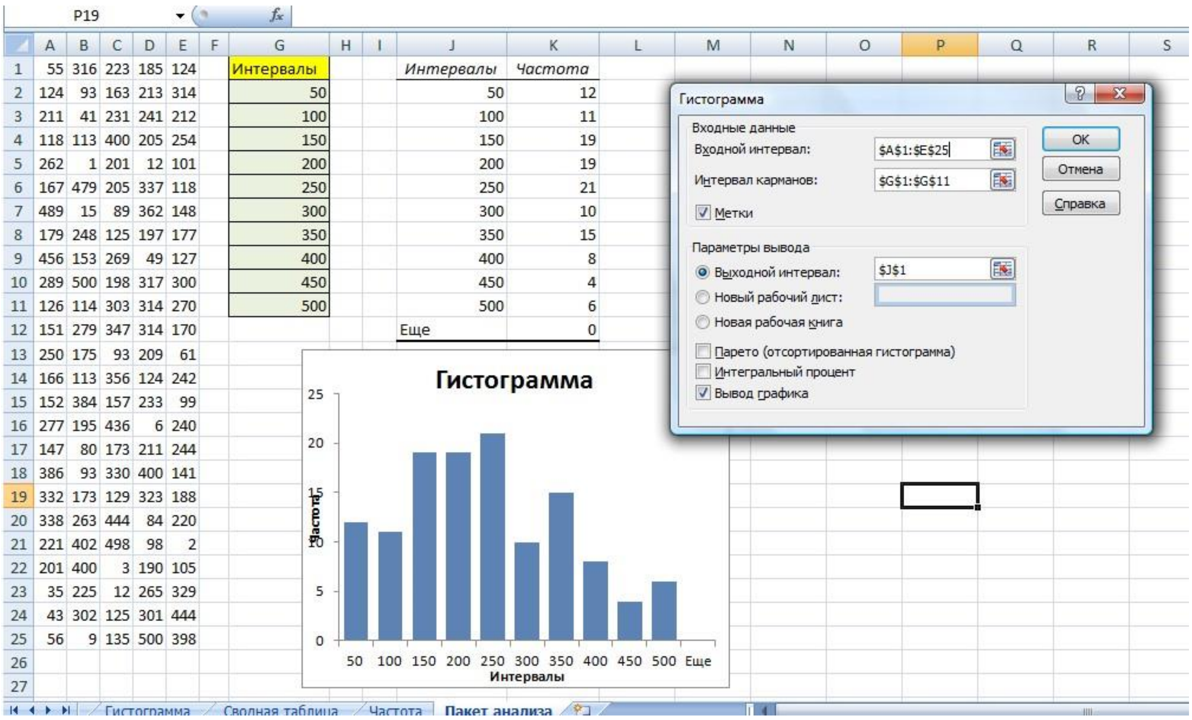


Рис. 1.4. Результат обработки статистических данных в надстройке. Пакет анализа

Необходимые действия, которые следует выполнить, приведены на рисунке.

Задание на выполнение работы

1. Проведите статистическое исследование, например, изучите мнение студентов об организации учебного процесса ВУЗа; проведите среди студентов контроль остаточных знаний по предметам, изучавшимся ими на предыдущих курсах (для студентов 2-го курса – за 1 год обучения, для студентов 3-го курса – за 1, 2 года и так далее). Для этого определите:

1) объект и единицу наблюдения;

2) признаки, подлежащие регистрации;

3) вид и способ наблюдения;

4) разработайте формуляр и напишите краткую инструкцию к его заполнению;

5) составьте оргплан обследования;

6) произведите наблюдение и результаты его представьте в виде статистических таблиц и графиков.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) результаты вычисления индивидуальных заданий;
3) анализ результатов с использованием статистической функции ЧАСТОТА табличного процессора MS EXCEL.

Контрольные вопросы

1. Дайте определение статистического наблюдения. В чем его сущность?
2. Какие вопросы входят в план наблюдения?
3. Что является целью наблюдения?
4. Что такое «объект наблюдения» и как он определяется?
5. Что представляет собой единица наблюдения?
6. Что представляет собой программа наблюдения и как она оформляется?
7. В каких формах осуществляется наблюдение?
8. На какие виды подразделяется наблюдение по времени регистрации и по степени охвата единиц наблюдения?
9. Каким правилам должны удовлетворять вопросы программы статистического наблюдения?
10. Какие статистические функции имеются в табличном процессоре MS EXCEL?

Практическая работа № 2

Сводка и группировка статистических данных (4 часа)

Цель работы

1. Изучить основные положения и определения, используемые при проведении группировки статистических данных.
2. Изучить методику проведения структурной и вторичной группировки статистических данных.
3. Сформировать практические навыки проведения сводки и группировки статистических данных.
4. Построить структурную группировку по статистическим данным.

Краткая теория

Группировка статистических данных

Статистическая группировка – это распределение единиц наблюдения по группам, однородным по одному или нескольким признакам. С помощью группировки решаются следующие задачи:

1) выделение социально-экономических типов (типологическая группировка);

2) изучение структуры явления и структурных сдвигов, происходящих в нем (структурные группировки);

3) выделение связи и зависимости между явлениями (аналитическая группировка).

Перед началом группировки выделяют *группировочный признак* или *основания группировки* – это существенный признак, по которому вся совокупность делится на группы. Выбор основания зависит от цели группировки и всего исследования в целом и от предварительного экономического анализа.

Основанием группировки могут служить следующие признаки: качественный (атрибутный), количественный, пространственный, временной.

Качественные признаки выражают свойства объекта или явления через их наименование без количественного выражения. Например, группировка предприятий по формам собственности – на муниципальную, федеральную, собственность субъектов Федерации и др.

Количественный признак определяет уровень признака. Например, разряд рабочих, уровень образования, возраст и т.д.

Пространственный признак определяет место расположения единицы наблюдения. Например, слесарь в Москве и Казани имеет разную заработную плату.

Временной признак определяет время протекания исследуемого процесса.

После того как определено основание группировки, следует решить вопрос о количестве групп, на которые надо разбить исследуемую совокупность.

При группировке по *качественному признаку* число групп определяется количеством соответствующих наименований, если число этих наименований невелико. Например, по полу – на две группы; по национальному составу – на столько групп, сколько имеется национальностей и т.д.

При группировке по *количественному признаку* число групп определяется в зависимости от характера изменения признака и задач исследования. Если количественный признак меняется *прерывно* (*дискретно*), то есть может принимать только некоторые – чаще целые значения (например, тарифный разряд рабочих), то число групп должно соответствовать количеству значений признака.

При *непрерывном* изменении признак принимает любые значения (например, стаж работы или возраст рабочих), поэтому группы ограничиваются значениями признака в интервале «от – до».

Определение числа групп можно осуществить по *формуле Стерджесса*:

$$k = 1 + 3{,}322\lg(n) = 1 + 1{,}4\ln(n), \qquad (2.1)$$

где n – число единиц (объем) совокупности.

После определения числа групп необходимо определить интервалы группировки. **Интервал** – это значение варьирующего признака, лежащее в определенных границах.

Величина интервала – это разность между максимальным $x_{\max}$ и минимальным $x_{\min}$ значениями признака в каждой группе.

На практике используются три вида интервалов: равные, неравные (постепенно увеличивающиеся или уменьшающиеся) и специализированные.

Группировки с **равными интервалами** целесообразны в тех случаях, когда вариация проявляется в сравнительно узких границах. Формула Стерджесса (2.1) для расчета числа групп применяется при равных интервалах в группах.

Величина равного интервала вычисляется по формуле:

$$b = \frac{R}{k} = \frac{x_{\max} - x_{\min}}{k}, \qquad (2.2)$$

где $x_{\max}$, $x_{\min}$ – соответственно наибольшее и наименьшее значение признака в изучаемой совокупности, k – число групп.

Если размах вариации признака велик и значения признака изменяются не равномерно, то необходимо использовать группировки с **неравными интервалами**. Неравные интервалы могут быть: прогрессивновозрастающими или прогрессивно-убывающими в арифметической или геометрической прогрессии. Величина интервалов определяется как:

$b_{i+1} = b_i + a$, – для арифметической прогрессии; (2.3) $b_{i+1} = b_i \cdot q$, – для геометрической прогрессии, (2.4)

где a – константа, имеющая для прогрессивно-возрастающих интервалов знак «+», а для прогрессивно-убывающих интервалов знак «–»; q – константа (для прогрессивно-убывающих интервалов $q < 1$; в другом случае – $q > 1$).

Специализированные интервалы – это интервалы, различные для разных отраслей и производств, группируемых по одному и тому же признаку, с учетом особенностей каждой отрасли или производства.

Например, группировка предприятий по числу рабочих в автомобильной промышленности будет иметь одни интервалы, в легкой – другие.

Интервалы группировок могут быть *закрытыми* и *открытыми*. **Закрытыми** называются интервалы, у которых имеются верхняя и нижняя границы. У **открытых** интервалов указана только одна граница.

Ряды распределения, построенные по количественному признаку, называются **вариационными**. Например, распределение населения по возрасту, рабочих – по стажу работы, заработной плате и т.д.

Вариационные ряды распределения состоят из двух элементов: **вариантов** и **частот**.

Варианты – упорядоченные значения количественного признака в ряду распределения. Они могут быть положительными и отрицательными, абсолютными (кг, м, руб.) и относительными (доли единицы, проценты). Так, при группировке предприятий по результатам хозяйственной деятельности варианты могут быть положительными (прибыль) и отрицательными (убыток) числами.

Частоты (f_i **)** – это абсолютные числа, показывающие, сколько раз встречаются те или иные варианты в ряду распределения. Сумма всех частот называется **объемом** совокупности и определяет число элементов всей совокупности.

Частосты (p_i **)** – это частоты, выраженные в виде относительных величин. Сумма частостей равна единице или 100 %. Замена частот частостями позволяет сопоставлять вариационные ряды с разным числом наблюдений.

Для вариационных рядов существует еще два типа частотных характеристик: **накопленная частота** и **накопленная частость**. Накопленная частота показывает, какое число единиц имеет величину варианта, не большую данной. Она определяется путем суммирования значения признака по данной группе со всеми частотами предшествующих групп. Накопленная частость характеризует удельный вес единиц наблюдения, у которых значение признака не превосходит верхнюю границу данной группы.

Виды группировок

В статистике различают:

1. **Типологические группировки**.

Типологическая группировка – это разделение качественно разнородной исследуемой совокупности на однородные группы единиц в соответствии с социально-экономическими типами. Они разделяют исследуемые общественные явления на классы, социально-экономические типы, т.е. по атрибутному признаку.

2. **Структурные группировки**.

Структурные группировки – это расчленение однородной в качественном отношении совокупности единиц на группы, которые характеризуют строение этой совокупности, ее структуры. Например, изучение состава населения по полу, возрасту и другим признакам, определение значения каждого вида транспорта в транспортном балансе страны, группировка предприятий по численности работников и т.д.

3. **Аналитические группировки**.

Они устанавливают связь между отдельными признаками изучаемого объекта и выявляют факторы, влияющие на эту связь. Под *факторами* в статистике понимают признаки, которые оказывают влияние на другие, зависимые признаки, называемые *результативными.*

4. **Вторичные группировки**.

Это повторная группировка статистического материала, производимая по тому же признаку, что и первичная группировка с дроблением или укрупнением интервалов группировки. Цель вторичной группировки: получить более наглядную картину развития изучаемого явления.

Вторичная группировка может производиться *двумя способами*:

– *объединение первоначальных интервалов*; в этом случае интервалы первичной группировки укрупняются, что дает более четкую картину развития явления;

– *долевая перегруппировка*; осуществляется в соответствии с удельным весом каждой группы в общей совокупности.

5. **Комбинированные группировки.**

Это группировка, производимая не по одному, а по нескольким группировочным признакам. Поскольку при этом резко возрастает количество групп, то этот вид группировки применяют при большом количестве наблюдений.

Методические указания по проведению структурной группировки

По данным табл. 2.1 произведите группировку 20 коммерческих банков по сумме выданных кредитов (в млн руб.).

Таблица 2.1

Данные о сумме выданных кредитов банками

№ банка	1	2	3	4	5	6	7	8	9	10
Сумма кредитов	20	40	50	50	60	60	130	60	60	70
№ банка	11	12	13	14	15	16	17	18	19	20
Сумма кредитов	70	80	80	90	100	110	40	30	30	40

Проведите группировку коммерческих банков по сумме выданных кредитов.

Решение

1. Постройте ранжированный ряд, то есть, расположите все элементы выборки в порядке возрастания значений:

20, 30, 30, 40, 40, 40, 50, 50, 60, 60, 60, 60, 70, 70, 80, 80, 90, 100, 110, 130.

2. Определите размах варьирования прибыли:

$$R_x = x_{max} - x_{min} = 130 - 20 = 110 \text{ млн. руб.},$$

где x_{max}, x_{min} – соответственно наибольшее и наименьшее значение прибыли.

3. По формуле Стерджесса (2.1) определите число групп, на которое надо разбить исследуемую выборку из 20 банков:

$$k = 1 + 1{,}4 \cdot \ln(n) = 1 + 1{,}4 \cdot \ln(20) \approx 5{,}2 \approx 5,$$

где n – объем выборки.

Полученное значение k округлите до целого числа.

4. Зная число групп, определите величину интервала по формуле (2.2):

$$b = R_x / k = 110/5 = 22{,}0 \text{ млн. руб.}$$

5. Определите левые и правые границы отдельных групп:

$$x_{1лев} = x_{min}; \quad x_{1прав} = x_{1лев} + b; \quad x_{2лев} = x_{1прав}; \quad x_{2прав} = x_{2лев} + b$$

и т.д.

$x_1^{лев} = x_{min} = 20$ млн руб.; $x_1^{прав} = x_1^{лев} + b = 20 + 22{,}0 = 42{,}0$ млн руб.; $x_2^{лев} = x_1^{прав} = 42{,}0$ млн руб.; $x_2^{прав} = x_2^{лев} + b = 42{,}0 + 22{,}0 = 64{,}0$млн руб.; $x_3^{лев} = x_2^{прав} = 64{,}0$ млн руб.; $x_3^{прав} = x_3^{лев} + b = 64{,}0 + 22{,}0 = 86{,}0$ млн руб

$x_4^{лев} = x_3^{прав} = 86{,}0$ млн руб.; $x_4^{прав} = x_4^{лев} + b = 86{,}0 + 22{,}0 = 108{,}0$ млн руб.; $x_5^{лев} = x_4^{прав} = 108{,}0$млн руб.; $x_5^{прав} = x_5^{лев} + b = 108{,}0 + 22{,}0 = 130{,}0$ млн руб.

6. Определите среднее значения кредитов в середине каждой группы:

$$\bar{x}_j = (x_{левj} + x_{правj})/2,$$

где j – номер класса или группы $(1 \le j \le k)$.

$\bar{x}_1 = (x_1^{лев} + x_1^{прав})/2 = (20 + 42{,}0)/2 = 31{,}0$ млн руб.; $\bar{x}_2 = (x_2^{лев} + x_2^{прав})/2 = (42{,}0 + 64{,}0)/2 = 53{,}0$ млн руб.; $\bar{x}_3 = (x_3^{лев} + x_3^{прав})/2 = (64{,}0 + 86{,}0)/2 =$

75,0 млн руб.; $x_4 = (x_4^{лев} + x_4^{прав})/2 = (86,0 + 108,0)/2 = 97,0$ млн руб.; $x_5 = (x_5^{лев} + x_5^{прав})/2 = (108,0 + 130,0)/2 = 119,0$ млн руб.

7. Определите частоту попадания f_j случайной величины в j-й класс. Для этого элементы ранжированного ряда сравните с границами j-го класса. Если элемент x_i ($1 \le i \le n$) вариационного ряда удовлетворяет условиям:

$$\begin{cases} x_{левj} \le x_i < x_{правj} \text{ при } j < k, \\ x_{левj} \le x_i \le x_{правj} \text{ при } j = k, \end{cases}$$

тогда его отнесите к j-му ($1 \le j \le k$) классу, то есть f_j увеличите на 1.

В результате получите следующий интервальный ряд распределения коммерческих банков по сумме выданных кредитов:

x_j …	20–42,0	42,0–64,0	64,0–86,0	86,0–108,0	108,0–130,0
f_j …	5	7	4	2	2

8. Рассчитайте частости $p_j = f_j / n$ случайной величины в j-м классе:

$$p_1 = \frac{f_1}{n} = \frac{5}{20} = 0,25;$$

$$p_2 = \frac{f_2}{n} = \frac{7}{20} = 0,35;$$

$$p_3 = \frac{f_3}{n} = \frac{4}{20} = 0{,}2;$$

$$p_4 = \frac{f_4}{n} = \frac{2}{20} = 0{,}1;$$

$$p_5 = \frac{f_5}{n} = \frac{2}{20} = 0{,}1.$$

9. Определите сумму накопленной частоты попадания в j-й класс

$$S_{нак} = \sum_{j=1}^{k} f_j .$$

Все расчеты представьте в табличной форме (см. табл. 2.2).

Таблица 2.2

Группировка банков по сумме выданных кредитов

Номер группы j	$x_{лев\,j}$	$x_{прав\,j}$	$\bar{x}_j$	f_j	p_j	$S_{нак}$
1	20,0	42,0	31,0	5	0,25	5
2	42,0	64,0	53,0	7	0,35	12
3	64,0	86,0	75,0	4	0,2	16
4	86,0	108,0	97,0	2	0,1	18
5	108,0	130,0	119,0	2	0,1	20
Сумма:	–	–	–	20	1	–

Методические указания по проведению вторичной группировки

Рассмотрим приемы проведения вторичной группировки на примерах.

Пример 1. Произведите укрупнение интервалов на основе данных табл. 2.3.

Таблица 2.3

Группировка магазинов по размеру товарооборота

Группы магазинов по размеру товарооборота за IV квартал, тыс. руб.	Число магазинов	Товарооборот за IV квартал, тыс. руб.
До 10	15	93
10–15	8	112
15–20	13	200
20–30	3	68
30–50	9	378
50–60	7	385
60–70	3	180
70–100	8	600
100–200	22	2400
Свыше 200	12	3744
Итого	100	8160

Приведенная группировка недостаточно наглядна, потому что не показывает четкой и строгой закономерности в изменении товарооборота по группам.

Уплотните ряды распределения, образовав шесть групп. Новые группы образованы путем суммирования первоначальных групп (табл. 2.4).

Таблица 2.4

Вторичная группировка магазинов по размеру товарооборота

Группы магазинов по размеру товарооборота за IV квартал, тыс. руб.	Число магазинов	Товарооборот за IV квартал, тыс. руб.	Товарооборот в среднем на 1 магазин, тыс. руб.
До 10	15	93	6,2
10–20	21	312	14,9
20–50	12	446	37,2
50–100	18	1165	64,7
100–200	22	2400	109,1
Свыше 200	12	3744	312,0
Итого	100	8160	

Пример 2. Имеются следующие данные о распределении фирм по количеству служащих по двум регионам (табл. 2.5).

Таблица 2.5

Группировка фирм по количеству служащих

№ п/п	Группы фирм по количеству служащих	Удельный вес фирм, в % к итогу	Группы фирм по количеству служащих	Удельный вес фирм, в % к итогу
1	До 100	4,3	до 50	1,0
2	100–200	18,4	50–70	1,0
3	200–300	19,5	70–100	2,0
4	300–500	28,1	100–150	10,0
5	Свыше 500	29,7	150–250	18
6	–	–	250–400	21
7	–	–	400–500	23
8	–	–	свыше 500	24
	Итого	100	Итого	100

Эти данные не позволяют провести сравнение распределения фирм по количеству служащих в двух регионах, так как фирмы первого региона разбиты на 5 групп, а второго – на восемь. Необходимо ряды распределения привести к сопоставимому виду.

За основу сравнения возьмите распределение фирм в первом регионе. Следовательно, произведите перегруппировку фирм второго региона, чтобы образовать такое же число групп и с теми же интервалами, как и в первом регионе. Полученные данные сведены в таблице 2.6.

Таблица 2.6

Вторичная группировка фирм по количеству служащих

Группы фирм по количеству служащих	Удельный вес фирм, в % к итогу		Расчеты
	I регион	II регион	
До 100	4,3	4,0	1 + 1+ 2 = 4
100–200	18,4	19,0	10 + 9 = 19
200–300	19,5	16,0	9 + 7 = 16
300–500	28,1	37,0	21–7 = 14, 14 + 23 = 37
Свыше 500	29,7	24,0	24

Итого	100,0	100,0	–

Для определения числа фирм, которые надо взять из пятой группы, условно примите, что это число фирм, должно быть пропорционально удельному весу отобранных служащих в группе.

Определите удельный вес 50 служащих в пятой группе:

$$(50*18)/(250 - 150) = 9.$$

Определите удельный вес 50 служащих в шестой группе:

$$(50*21)/(400 - 250) = 7 \text{ и т.д.}$$

Порядок выполнения работы

1. Ознакомьтесь с методикой проведения группировки статистических данных и построения статистического вариационного ряда.
2. У преподавателя получите вариант индивидуального задания (приложение).
3. Проведите структурную и вторичную группировку статистических данных и постройте статистический вариационный ряд. Представьте полученный статистический вариационный ряд в табличной форме.
4. Сделайте выводы по работе и оформите отчет.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) алгоритм и результаты вычисления индивидуальных заданий;
3) анализ полученного статистического вариационного ряда в табличной форме.

Контрольные вопросы

1. Что называется статистической группировкой?
2. Какие группировочные признаки вы знаете?

3. Дайте характеристику типологических группировок.
4. Дайте характеристику структурных группировок.
5. Дайте характеристику аналитических группировок.
6. Как можно определить число групп и границ интервалов между ними?
7. Какие бывают интервалы группировок и как точно обозначить их границы? Приведите примеры.
8. Что называется вторичной группировкой, в каких случаях приходится прибегать к ней и как можно получить новые группы на основании уже имеющихся?
9. Что представляют собой статистические ряды распределения ипо каким признакам они могут быть образованы?
10. Какова методика построения дискретных и интервальных рядов распределения? Приведите примеры.

Практическая работа № 3

Обработка статистических данных на основе табличного процессора MS EXCEL (4 часа)

Цель работы

1. Изучить методику построения интервального статистического ряда, гистограммы частот в **EXCEL**.
2. Сформировать практические навыки проведения расчета статистических характеристик в **EXCEL**.
3. Построить интервальный статистический ряд в **EXCEL**.
4. Построить гистограммы в **EXCEL**.

Методические указания по выполнению лабораторной работы

Рассмотрим все этапы обработки статистических данных на примере.

Задание

Для случайной выборки объемом $n = 50$ с несовпадающими числами выполнить следующую последовательность действий:

1. Выведите на лист Excel исходные статистические данные.
2. Постройте вариационный ряд.

3. Вычислите статистические характеристики.
4. Постройте интервальный статистический ряд.
5. Постройте гистограмму частот.
6. Составьте статистическую функцию распределения статистического ряда.
7. Составьте и постройте статистическую функцию распределения группированного статистического ряда.
8. Сделайте выводы.

В качестве примера рассмотрите следующую выборку:

2,183	3,962	1,797	3,206	8,039	4,666	5,497	2,916	5,842	5,410
7,566	6,384	1,269	4,333	8,350	7,973	6,059	4,746	5.536	5,419
9,852	4,925	6,758	4,652	2,567	3,007	2,951	1,706	3,722	5,287
2,907	4,541	2,655	2,461	4,108	5,563	7,190	2,275	4,122	8,949
3,490	5,874	9,576	2,667	2,942	4,797	6,273	6,197	5.577	4,679

1. Ввод исходных статистических данных. Вводите данные в первый столбец таблицы (рис. 3.1).

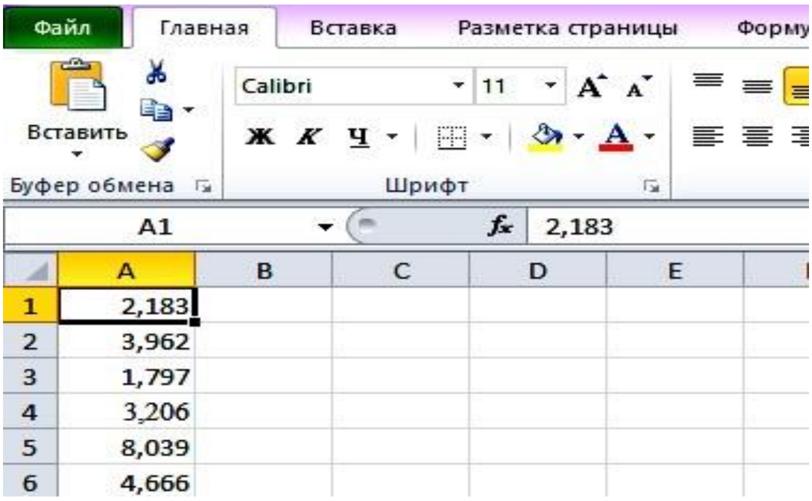


Рис. 3.1. Ввод исходных данных

2. Построение вариационного ряда.

Произведите сортировку данных в порядке возрастания. Для этого: а) выделите первый столбец;

б) на ленте во вкладке «Данные» выберите «Сортировка и фильтр» (рис. 3.2).

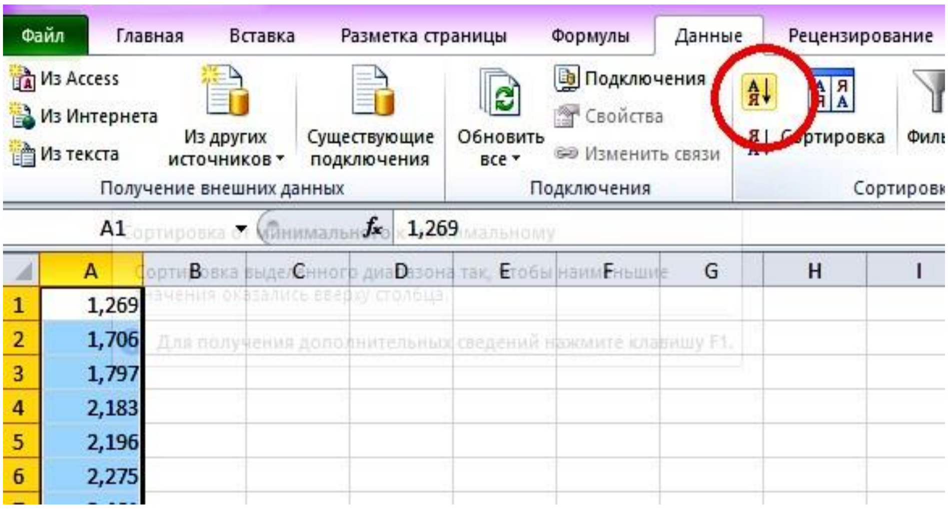


Рис. 3.2. Сортировка данных

3. Вычисление статистических характеристик.

На ленте во вкладке «Данные» выберите «Анализ данных» меню «Описательная статистика» нажмите ОК.

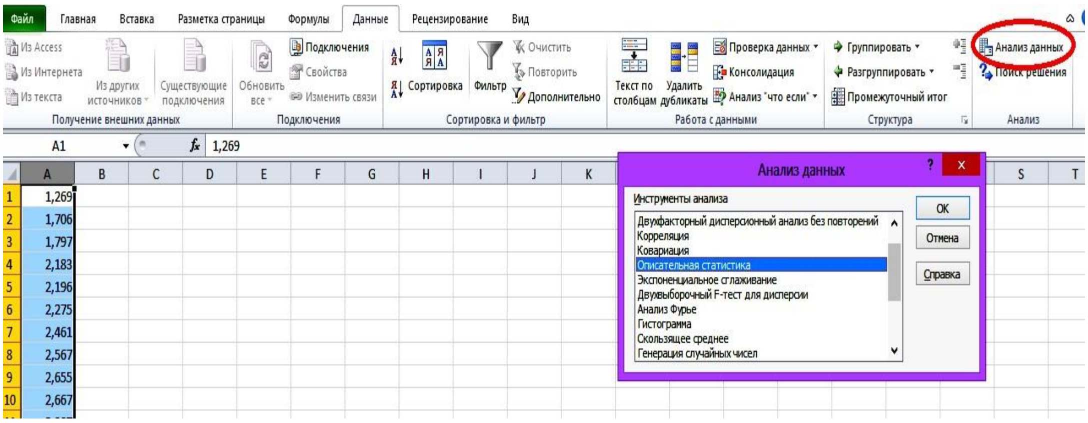


Рис. 3.3. Выбор надстройки «Анализ данных»

В пункт «Входной интервал» введите диапазон ячеек с исходными данными $A$1:$A$50, а в пункте «Выходной интервал» обозначьте первую ячейку для записи результаов $C$1. Поставьте флажок напротив пункта «Итоговая статистика» и нажмите ОК (рис. 3.4).

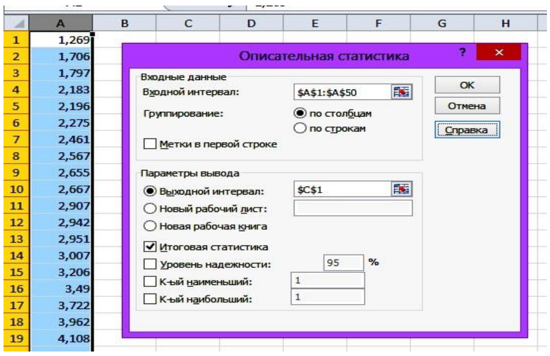


Рис. 3.4. Ввод данных в окно «Описательная статистика»

На рабочем листе появляется таблица с вычисленными значениями числовых характеристик выборки (рис. 3.5).

	A	B	C	D	E
1	1,269		*Столбец1*		
2	1,706				
3	1,797		Среднее	4,85406	
4	2,183		Стандартная ошибка	0,30046143	
5	2,196		Медиана	4,7125	
6	2,275		Мода	#Н/Д	
7	2,461		Стандартное отклонение	2,12458318	
8	2,567		Дисперсия выборки	4,51385369	
9	2,655		Эксцесс	-0,3252774	
10	2,667		Асимметричность	0,47173201	
11	2,907		Интервал	8,583	
12	2,942		Минимум	1,269	
13	2,951		Максимум	9,852	
14	3,007		Сумма	242,703	
15	3,206		Счет	50	
16	3,49				

Рис. 3.5. Результат обработка статистического ряда

Здесь «Среднее» означает математическое ожидание выборки, а «Стандартная ошибка» – погрешность ее значения. «Дисперсия выборки» означает исправленную выборочную дисперсию, а «Стандартное отклонение» – исправленное среднее квадратичное отклонение. Положительное значение «Асимметричности» означает, что «длинная часть» кривой лежит правее моды. Отрицательное значение «Эксцесса» означает, что кривая имеет более низкую и «плоскую» вершину, чем нормальная кривая. «Интервал» равен разности **Xmax − Xmin**. «Сумма» дает результат суммирования всех элементов выборки. «Счет» задает общее число элементов выборки.

4. Построение интервального статистического ряда.

Длину интервала группировки определите по формуле (2.2 лаб. раб. № 2).

Необходимые данные имеются в таблице: Xmax – в ячейке D13, Xmin – в ячейке D12, число элементов выборки *n* – в ячейке D15.

В ячейку C16 введите слово «Интервал», в ячейку D16 введите формулу:

□ (D13□ D12)/(1□ LOG(D15;2)).

В ячейке D16 появится значение числа *b*. В ячейку C17 введите букву *b*. В ячейку D17 введите формулу:

□ ОКРУГЛ(D16;1).

В ячейке D17 получите округленное до одного знака после запятой значение интерала *b*.

Проведите формирование интервалов. Для этого от Xmin отступите влево примерно на *b*/2 и получите начальную точку отсчета. Последовательно прибавляйте к ней целое число отрезков *b*, получите все граничные точки интервалов. В ячейку F1 введите формулу:

□ ОКРУГЛ((D12□D17/2);1).

В этой ячейке появится значение начальной точки отсчета. В ячейку F2 введите формулу:

□ F1□$D$17.

В этой ячейке появится значение второй граничной точки первого интервала. Вернитесь в ячейку F2, поставьте курсор в правый нижний угол рамки и двигайте его вниз, не отпуская левую кнопку мыши. В результате такой процедуры (протяжка) столбец F заполнят граничные точки интервалов. Самый нижний интервал должен включать Xmax (рис. 3.6).

Проведите подсчет числа вариант, попавших в каждый интервал, определите относительные частоты и серединные точки этих интервалов.

Для этого на ленте во вкладке «Данные» выберите «Анализ данных» меню «Гистограмма» (рис. 3.6).

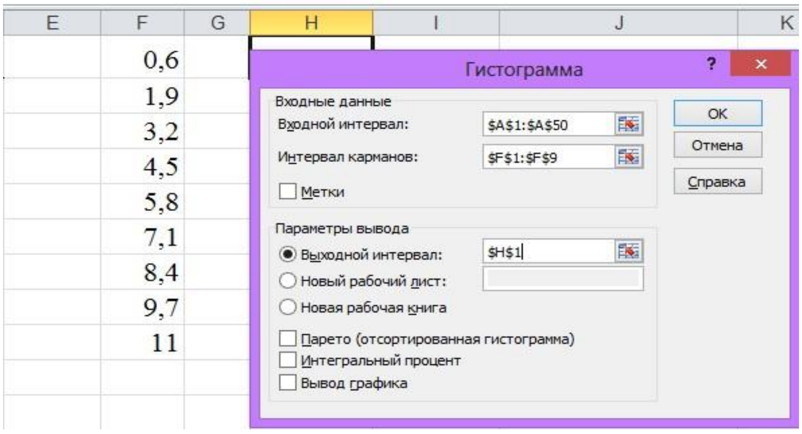


Рис. 3.6. Построение гистограммы

В пункт «Входной интервал» введите диапазон ячеек с исходными данными $A$1:$A$50, в пункт «Интервал карманов» – диапазон ячеек с границами интервалов $F$1:$F$9. Отметьте точкой пункт «Выходной интервал» и введите в него адрес первой ячейки для записи результатов $H$1. Появится таблица из двух столбцов с обозначениями «Карман» и «Частота» (рис. 3.7).

Определите относительные частоты p_i^*, значения серединных точек интервалов:

$$x_i^* = \frac{x_{i-1} + x_i}{2},$$

и высоты прямоугольников:

$$y_i = p_i^* / b.$$

Для этого

1) в ячейку J1 введите заголовок «Относительная частота»;
2) в ячейку J3 введите формулу:

= I3/D$15

и протяните её вниз до ячейки J10. В результате к таблице из двух столбцов добавится третий столбец (рис. 3.7). В этой таблице частота появления случайной величины в каждом интервале записана в одной строке с концом интервала;

3) в ячейку K1 введите заголовок столбца X*;

4) в ячейку K3 введите формулу:

 СРЗНАЧ(H2:H3).

Протяните эту формулу до ячейки K10. В результате в четвертом столбце таблицы (рис. 3.7) появятся значения серединных точек интервалов;

5) в ячейку L1 введите заголовок столбца $У_i$;

6) в ячейку L3 введите формулу:

 J3/4D$17.

Протяните её вниз до ячейки L10.

В результате в пятом столбце таблицы (рис. 3.7) появятся значения $У_i$.

H	I	J	K	L
Карман	***Частота***	***Относит. частота***	**x***	**yi**
0,6	0		0	
1,9	3	0,06	1,25	0,0462
3,2	11	0,22	2,55	0,1692
4,5	7	0,14	3,85	0,1077
5,8	14	0,28	5,15	0,2154
7,1	7	0,14	6,45	0,1077
8,4	5	0,1	7,75	0,0769
9,7	2	0,04	9,05	0,0308
11	1	0,02	10,35	0,0154
Еще	0			

Рис. 3.7. Определение относительной частоты распределения

5. Построение гистограммы частот.

Установите курсор на ячейку, где хотите расположить график и вверху в меню переключитесь на вкладку «Вставка». На ленте во вкладке «Вставка» выберите вид диаграммы «Гистограмма». Нажмите на знак «Вставка гистограммы», выберите «Гистограмма с группировкой». На ленте откроется вкладка работы с диаграммой. На листе Excel появится новый объект – чистый график. Когда он выделен, то верхняя панель с иконками действий имеет другой вид, специально для работы с графиками.

Чтобы заполнить график, нажмите на кнопку «Выбрать данные» (рис. 3.8).

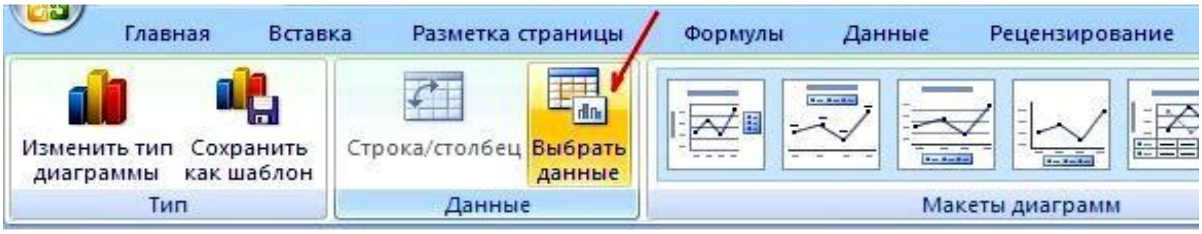


Рис. 3.8. Выбор данных

Отобразится окно выбора данных для графика. В нем имеется поле «Выбор данных для диаграммы». В конце поля нажмите на кнопку выбора диапазона (рис. 3.9).

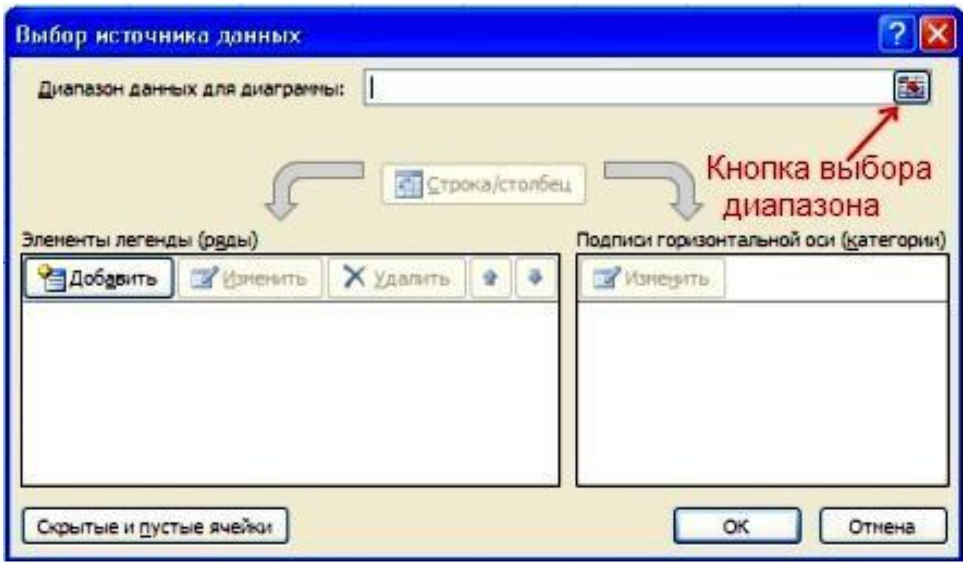


Рис. 3.9. Выбор диапазона данных

В появившееся окно введите диапазон ячеек со значениями относительной частоты (рис. 3.10).

 Лист1!$J$3:$J$10.

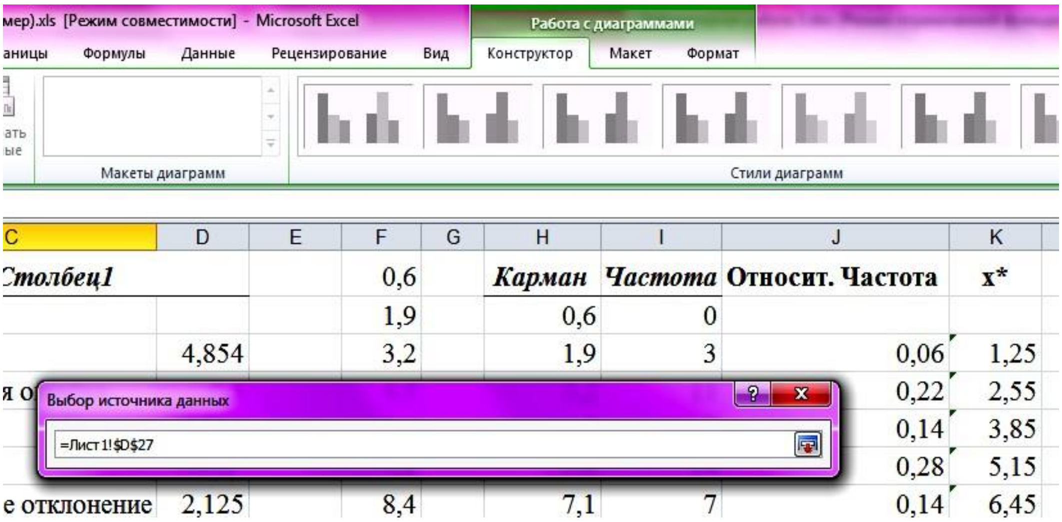


Рис. 3.10. Выбор источника данных

В полученном графике необходимо изменить данные по оси X. Вместо значений 1, 2, 3 и т.д., введите диапазон ячеек с серединными точками. Для этого, в правом столбце «Подписи горизонтальной оси (категории) нажмите на кнопку «Изменить» (рис. 3.11).

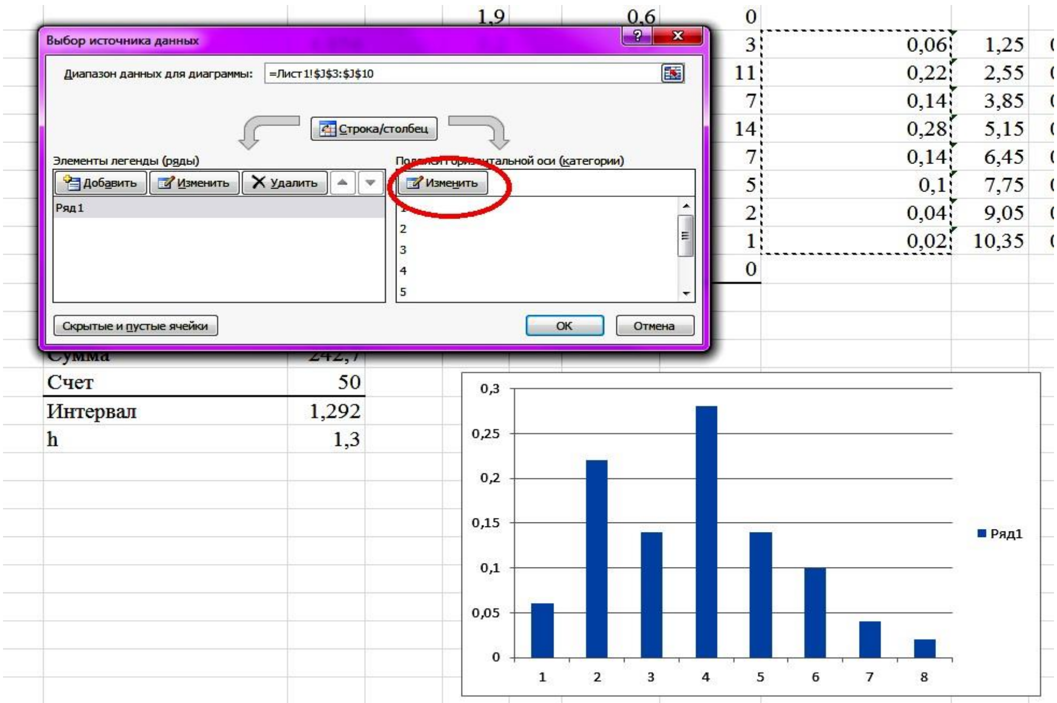


Рис. 3.11. Алгоритм изменения подписи

В окошко «Подписи по оси» введите =Лист1!$K$3:$K$10.

Появляется предварительное изображение диаграммы (рис. 3.12).

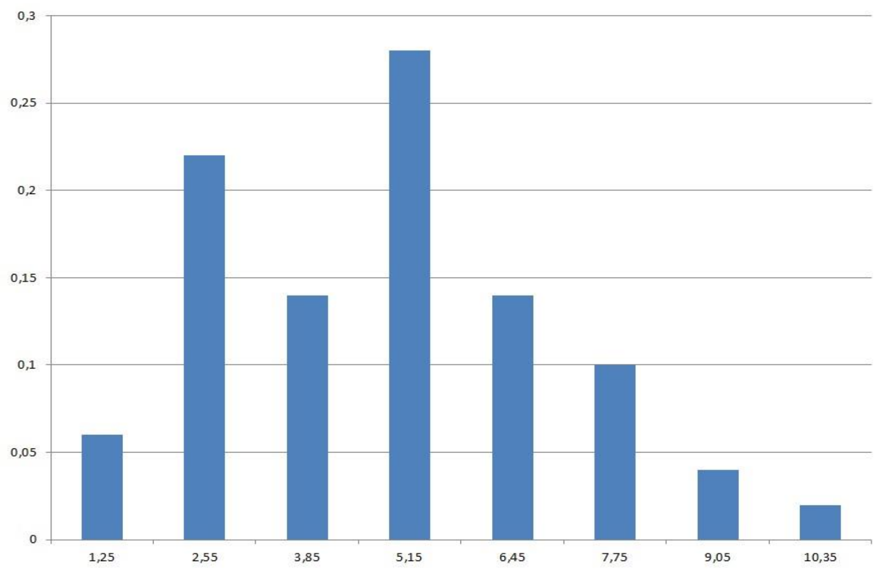


Рис. 3.12. Гистограмма с измененой подписью данных

Оформление графика можно изменить, выделяя нужную его часть, кликая правой клавишей мыши и выбирая «Формат области диаграммы», либо можно изменить тип диаграммы (рис. 3.13).

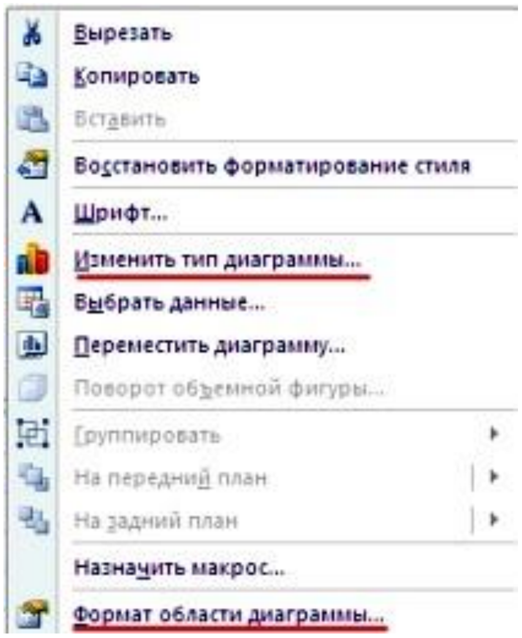


Рис. 3.13. Меню изменения типа диаграммы

Параметры оформления графика, такие как название, подписи осей, сетка и т.п., можно настроить, если **мышкой выделить сам график и в главном меню переключиться на пункт «Макет»**. Например, чтобы включить подпись оси по вертикали, выберите «Макет» → «Названия осей» →

«Название основной вертикальной» и один из появившихся вариантов расположения подписи, например «Вертикальное название» и так далее. Самостоятельно попробуйте различные виды заливок и выбора места размещения (рис. 3.14).

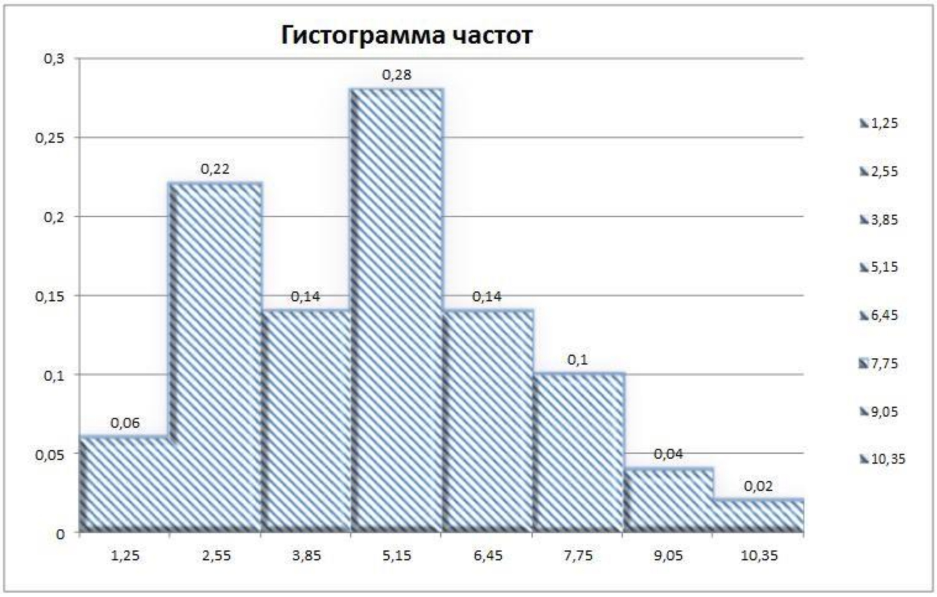


Рис. 3.14. Измененный вид диаграммы

6. Составление статистической функции распределения статистического ряда. Определите значение статистической функции распределения:

$$F^*(x) = \begin{cases} 0 & \text{при } x \le x_1, \\ p_1 & \text{при } x_1 < x \le x_2, \\ p_1 + p_2 & \text{при } x_2 < x \le x_3, \\ \ldots & \\ p_1 + p_2 + \ldots + p_{n-1} & \text{при } x_{n-1} < x \le x_n, \\ 1 & \text{при } x > x_n. \end{cases}$$

В нашем случае все $P_i=1/n$ и график функции образует последовательность из 50 ступенек одинаковой высоты, но разной ширины.

Проведите расчет функции *F*(x)* на Листе 2.

Для этого выделите столбец А (исходные данные), скопируйте на Лист 2. В ячейку В1 (на Листе 2) введите формулу:

=1/Лист3!$D$15.

В ячейку В2 введите формулу:

=В1+1/Лист3!$D$15.

Протяните до ячейки В51 (рис. 3.15).

	A	B
1	Исходные данные	Функция распределения
2	1,269	=1/Лист3!$D$15
3	1,706	=B2+1/Лист3!$D$15
4	1,797	=B3+1/Лист3!$D$15
5	2,183	=B4+1/Лист3!$D$15
6	2,196	=B5+1/Лист3!$D$15
7	2,275	=B6+1/Лист3!$D$15
8	2,461	=B7+1/Лист3!$D$15
9	2,567	=B8+1/Лист3!$D$15
10	2,655	=B9+1/Лист3!$D$15
11	2,667	=B10+1/Лист3!$D$15
12	2,907	=B11+1/Лист3!$D$15
13	2,942	=B12+1/Лист3!$D$15
14	2,951	=B13+1/Лист3!$D$15
15	3,007	=B14+1/Лист3!$D$15
16	3,206	=B15+1/Лист3!$D$15
17	3,49	=B16+1/Лист3!$D$15
18	3,722	=B17+1/Лист3!$D$15
19	3,962	=B18+1/Лист3!$D$15
20	4,108	=B19+1/Лист3!$D$15

Лист1 Лист2 Лист3

Готово

Рис. 3.15. Построение функции распределения

7. Составление и постоение статистической функции распределения группированного статистического ряда.

Менее трудоемко и часто более удобно для наглядного анализа свойств выборки построение статистической функции распределения группированного статистического ряда, который содержит существенно меньшее количество точек.

Составьте и постройте такую функцию распределения.

Для этого перейдите на Лист 1 и в ячейку M1 введите заголовок столбца **F*(статистическая функция распределения).** В ячейку M3 введите формулу: =J3.

В ячейку M4 введите формулу:

=M3+J4.

Вернитесь в ячейку M4 и протяните до ячейки M10. В верхней строке столбцов **X*** и **F*** допишите нули. В результате получите последовательность значений функции распределения **F*** в последнем столбце таблицы (рис. 3. 16).

H	I	J	K	L	M
Карман	*Частота*	**Относит. Частота**	**x***	**yi**	**F***
0,6	0				
1,9	3	0,06	1,25	0,0462	0,06
3,2	11	0,22	2,55	0,1692	0,28
4,5	7	0,14	3,85	0,1077	0,42
5,8	14	0,28	5,15	0,2154	0,7
7,1	7	0,14	6,45	0,1077	0,84
8,4	5	0,1	7,75	0,0769	0,94
9,7	2	0,04	9,05	0,0308	0,98
11	1	0,02	10,35	0,0154	1
Еще	0				

Рис. 3.16. Вариант построения функции распределения

Постройте график функции распределения F*(x). Эта процедура полностью описана в разделе 5 «Построение гистограммы частот». Только в окне «Выбор источника данных» вместо интервала относительных частот введите интервал значений функции:

=Лист1!$L$2:$L$10,

а в строку «Подписи горизонтальной оси» введите диапазон ячеек с серединными точками =Лист1!$K$2:$K$10. Результат на рис. 3.17.

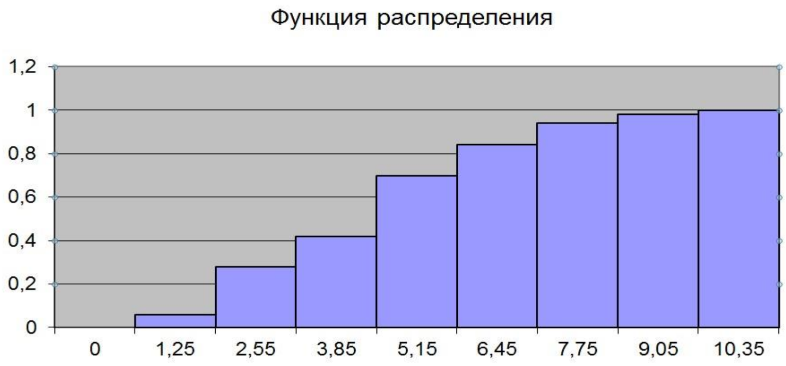


Рис. 3.17. Гистограмма функции распределения

Порядок выполнения работы

1. Ознакомьтесь с методикой проведения статистической обработке в Excel.
2. У преподавателя получите вариант индивидуального задания (приложение № 1).
3. Выполните следующие расчеты:

1) выведите на лист Excel исходные статистические данные;
2) постройте вариационный ряд;
3) вычислите статистические характеристики;
4) постройте интервальный статистический ряд;
5) постройте гистограмму частот;
6) составьте функцию распределения статистического ряда.
7) составьте и постоить функцию распределения группированного статистического ряда;
8) сделайте выводы.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) результаты обработки статистических данных в табличном процессоре Excel;
3) анализ полученного интервального статистического ряда и функции распределения.

Контрольные вопросы для самопроверки

1. Как проводится сортировка статистических данных в EXCEL?
2. Какие статистические характеристики можно вычислить с помощью меню «Описательная статистика»?
3. Опишите этапы построения интервального статистического ряда в EXCEL?
4. Как построить гистограмму частот в MS EXCEL?
5. Как построить функцию распределения статистического ряда в MS EXCEL?

Практическая работа № 4

Расчет средних величин, абсолютных и относительных показателей вариации признака (4 часа)

Цель работы

1. Изучить методику вычисления средних величин по статистическим данным.
2. Изучить методику вычисления абсолютных и относительных показателей вариации признака.
3. Сформировать практические навыки проведения расчета средних величин, абсолютных и относительных показателей вариации признака по статистическим данным.
4. Провести расчет средних величин, абсолютных и относительных показателей вариации признака.

Краткая теория

Средние величины

Средняя величина – обобщающая характеристика варьирующего признака единиц статистической совокупности.

Средняя величина всегда именованная, она имеет ту же размерность, что и признак у отдельных единиц совокупности.

В экономических исследованиях и плановых расчетах применяются две категории средних:

 степенные средние; структурные средние.

К категории степенных средних относятся: средняя арифметическая, средняя гармоническая, средняя квадратическая, средняя геометрическая.

Средняя обозначается через $\bar{x}$.

Общая формула степенной средней записывается следующим образом:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i}^{n} x_i^k \cdot f_i}{\sum_{i}^{n} f_i}},$$

где f_i – частота (повторяемость индивидуальных значений признака); x_i – величины, по которым вычисляется средняя.

В зависимости от степени k получаются различные виды средних величин, их формулы представлены в табл. 4.1.

Таблица 4.1

Формулы расчета различных видов степенных средних величин

Значение k	Наименование средней	Формулы средней	
		простая	взвешенная
–1	Гармоническая	$\bar{x} = \frac{n}{\sum \frac{1}{x_i}}$	$\bar{x} = \frac{\sum W_i}{\sum \frac{1}{x_i} W_i}$; $\bar{x} = \frac{\sum f_i}{\sum \frac{1}{x_i} f_i}$
0	Геометрическая	$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \ldots x_n}$	$\bar{x} = \sqrt[\sum f]{x_1^{f_1} \cdot x_2^{f_2} \ldots x_n^{f_n}}$
1	Арифметическая	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i}$
2	Квадратическая	$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}$
3	Кубическая	$\bar{x} = \sqrt[3]{\frac{\sum x_i^3}{n}}$	$\bar{x} = \sqrt[3]{\frac{\sum x_i^3 f_i}{\sum f_i}}$

Структурные средние величины

Структурные средние – мода и медиана – характеризуют величину варианта, занимающего определенное положение в ранжированном ряду.

Мода – наиболее часто встречающееся значение признака в данном ряду распределения, то есть имеющее наибольшую численность в ряду распределения.

В дискретном ряду мода определяется визуально, то есть отыскивается просмотром численностей, которые имеют варианты признака. Значение признака, имеющее наибольшую численность (в абсолютном или относительном выражении), и есть мода. Если несколько значений признака имеют одинаковую, наибольшую по сравнению с другими численность, то, следовательно, в ряду не одна, а несколько мод, например две. Ряд с двумя модами называется двумодальным.

Мода для интервального ряда распределения с равными интервалами определяется по формуле:

$$x_{mo} = x_{j^*}^{\text{лев}} + b\frac{f_{j^*} - f_{j^*-1}}{(f_{j^*} - f_{j^*-1}) + (f_{j^*} - f_{j^*+1})}, \quad (4.1)$$

где $x^{\text{лев}}_{j^*}$ – левая граница модального интервала; b – величина модального интервала; f_{j^*} – частота модального интервала; f_{j^*-1} – частота интервала, предшествующего модальному; f_{j^*+1} – частота интервала, следующего за модальным.

Медиана x_{me} – это значение признака, приходящееся на середину ранжированного ряда, то есть это вариант, который делит ряд распределения на две равные по объему части. При нормальном распределении статистической величины мода и медиана совпадают со средним значением этой величины: $x_{mo} = x_{me} = \bar{x}$.

Для определения медианы сначала определяют ее место в ранжированном ряду, используя формулу:

$$N_{me} = \frac{n+1}{2}, \qquad (4.2)$$

где n – число членов ряда.

Если совокупность содержит четное число значений варьирующего признака ($n = 2k$, $k = n/2$), то в этом случае за медиану условно принимают значение:

$$x_{me} = \frac{1}{2}(x_k + x_{k+1}),$$

так как в ряду нет члена, который делил бы совокупность на две равные по объему группы.

В дискретном ряду распределения медиана определяется непосредственно по накопленной частоте, соответствующей номеру медианы.

В интервальном ряду распределения сначала указывают интервал, в котором находится медиана.

Медианным является первый интервал, в котором сумма накопленных частот превысит половину общего числа наблюдений.

Медиана для интервального ряда распределения с равными интервалами определяется по формуле:

$$x_{me} = x_{j^{**}} + b\frac{\frac{n}{2} - S_{me-1}}{f_{j^{**}}}, \qquad (4.3)$$

где $x_{j^{**}}$ – нижняя граница медианного интервала; S_{me-1} – накопленная частота интервала, предшествующего медианному, $f_{j^{**}}$ – частота медианного интервала.

Показатели вариации

Вариация – изменения значения признака в пределах совокупности. Показатели вариации делятся на две группы: абсолютные и относительные.

К абсолютным показателям вариации относятся: размах колебаний, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение.

1. **Размах колебаний**, или **размах вариации**, представляет собой разность между максимальным и минимальным значениями признака в изучаемой совокупности.

Размах вариации определяется по формуле:

$$R = x_{max} - x_{min}. \qquad (4.4)$$

Безусловным достоинством этого показателя является простота расчета. Однако размах вариации зависит от величины только крайних значений признака, поэтому область его применения ограничена достаточно однородными совокупностями. В частности, на практике он находит применение в предупредительном контроле качества продукции.

2. **Среднее линейное отклонение *d* и среднее квадратическое отклонение** σ показывают, на сколько в среднем отличаются индивидуальные значения признака от среднего его значения. Среднее линейное отклонение определяется по формуле:

$$d = \frac{\sum_{j=1}^{k} |x_j - \bar{x}| \cdot f_j}{\sum_{j=1}^{k} f_j}. \qquad (4.5)$$

Среднее линейное отклонение имеет ту же размерность, что и признак, для которого оно вычисляется.

3. **Дисперсия** σ^2 представляет собой средний квадрат отклонений вариантов признака от средней величины и определяется по формуле:

$$\sigma^2 = \frac{\sum (x_j - \bar{x})^2 \cdot f_j}{\sum f_j}. \qquad (4.6)$$

Дисперсию можно определить и как разность между средним квадратом вариантов признака и квадратом их средней величины, то есть

$$\sigma^2 = \overline{x^2} - \bar{x}^2, \qquad \text{где } \overline{x^2} = \frac{\sum x_j^2 f_j}{\sum f_j}.$$

4. Среднее квадратическое отклонение σ – корень второй степени из дисперсии и определяется по формуле:

$$\sigma = \sqrt{\frac{\sum (x_j - \bar{x})^2 \cdot f_j}{\sum f_j}}. \qquad (4.7)$$

Для характеристики степени однородности совокупности, типичности, устойчивости средней, а также и для других статистических оценок используются **относительные показатели вариации.** Они вычисляются как отношение абсолютных показателей вариации к средней арифметической (или медиане) и чаще всего выражаются в процентах. Формулы расчета относительных показателей вариации следующие:

– коэффициент осцилляции $K_R = \frac{R}{\bar{x}} \cdot 100\ \%;$ (4.8)

– относительное линейное отклонение $K_d = \frac{\bar{d}}{\bar{x}} \cdot 100\ \%;$ (4.9)

– коэффициент вариации $V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100\ \%.$ (4.10)

Наиболее часто применяется **коэффициент вариации**. Его используют не только для сравнительной оценки вариации, но и для характеристики однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33 %
(для распределений, близких к нормальному).

Методические указания по выполнению лабораторной работы

По полученной ранее группировке определите:

1) среднее значение показателя, модальное и медианное значение;

2) показатели вариации абсолютные и относительные.

Определите средний размер выданных кредитов, модальное и медианное значение (пункт 1).

1. Средний размер суммы выданных кредитов по 20 коммерческим банкам определите по формуле средней арифметической взвешенной (см. табл. 1.1).

$$\bar{x} = \frac{\sum x_j f_j}{\sum f_j} = \frac{31,0 \cdot 5 + 53,0 \cdot 7 + 75,0 \cdot 4 + 97,0 \cdot 2 + 119,0 \cdot 2}{5 + 7 + 4 + 2 + 2} = \frac{12508,0}{20} = 62,9 \text{ млн руб.}$$

Следовательно, средняя сумма выданных кредитов по 20 обследованным нами банкам равна 62,9 млн. руб.

2. Рассчитайте структурные средние:

1) для определения модального размера выданных кредитов сначала определите модальный интервал по данным табл. 4.2. Это тот интервал, которому соответствует наибольшая частота f_j. Наибольшая частота равна 7, что соответствует второму интервалу [42,0–64,0]. Нижняя граница модального интервала $x^{лев}_{j*} = 42,0$; частота модального интервала $f_{j*} = 7$; частота интервала, предшествующего модальному, $f_{j*-1} = 5$; частота интервала, следующего за модальным, $f_{j*+1} = 4$.

Подставьте в формулу (4.1) соответствующие величины, получите:

$$x_{mo} = x^{лев}_{j^*} + b\frac{(f_{j^*} - f_{j^*-1})}{(f_{j^*} - f_{j^*-1}) + (f_{j^*} - f_{j^*+1})} =$$

$$= 42,0 + 22,0 \frac{7 - 5}{(7 - 5) + (7 - 4)} = 42,0 + 8,8 = 50,8 \text{ млн руб.}$$

Следовательно, банки наиболее часто выдают кредиты размером 50,8 млн. руб.;

2) для определения медианы сначала определите медианный интервал по данным табл. 4.2. Это тот интервал, до которого сумма накопленных частот меньше половины всей численности ряда, а с прибавлением его частоты – больше половины. По накопленным частотам половина численности $\frac{n}{2} = \frac{20}{2} = 10$. Сумма накопленных частот в первом интервале равна 5, то есть меньше половины численности ряда. Прибавив частоту второго интервала, равную 7, получите сумму 12, превышающую половину. Следовательно, медианным является второй интервал [42,0–64,0].

Медиану определите по формуле (4.3):

$$Me = x_{Me} + h\frac{\frac{n}{2} - S_{Me-1}}{f_{Me}} = 42{,}0 + 22{,}0\frac{10 - 5}{7} = 42{,}0 + 15{,}712 = 57{,}712 \approx 57{,}71 \text{ млн руб.}$$

Следовательно, у 50 % коммерческих банков сумма выданных кредитов меньше 57,71 млн. руб., а у остальные 50 % – больше.

Определите показатели вариации абсолютные и относительные (пункт 2).

1. Определите абсолютные показатели вариации распределения банков по размеру выданных кредитов (табл. 2.2).

а) найдите разницу между максимальным и минимальным размером кредита и получите величину размаха вариации [формула (4.4)]:

$$R = x_{max} - x_{min} = 130 - 20 = 110 \text{ млн руб.};$$

б) для расчета среднего линейного отклонения найдите абсолютное отклонение значений середины интервала (x_j) от средней величины ($\bar{x}$) по модулю $|x_j - \bar{x}|$. Вычислите произведения отклонений $|x_j - \bar{x}|$ на их веса (f_j), и подсчитайте сумму их произведений. Эта сумма равна 457,6 (графа 5 табл. 1.7).

По формуле (1.9) вычислите среднее линейное отклонение:

$$\bar{d} = \frac{\sum_{j=1}^{k} |x_j - \bar{x}| f_j}{\sum_{j=1}^{k} f_j} = \frac{457{,}6}{20} = 22{,}88 \text{ млн. руб.};$$

в) для определения дисперсия возведите в квадрат отклонения значений середины интервала (x_j) от средней величины ($\bar{x}$) – $(x_j - \bar{x})^2$. Затем квадрат отклонений умножьте на веса (f_j) и подсчитайте сумму, которая равна 14979,8 (графа 6 табл. 4.2).

По формуле (4.6) вычислите дисперсию:

$$\sigma^2 = \frac{\sum (x_j - \bar{x})^2 f_j}{\sum f_j} = \frac{14979{,}8}{20} = 748{,}99 \text{ млн. руб}^2.;$$

г) вычислети корень квадратный из дисперсии (4.7) и получите среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum (x_j - \bar{x})^2 \cdot f_j}{\sum f_j}} = \sqrt{748{,}99} = 27{,}37 \text{ млн. руб.}$$

Таблица 4.2

Расчет абсолютных показателей вариации

Номер группы j	Размер кредита, млн руб.	Середина интервала x_j	Число банков f_j	$\lvert x_j - \bar{x} \rvert \cdot f_j$	$(x_j - \bar{x})^2 \cdot f_j$
1	2	3	4	5	6
1	20,0–42,0	31,0	5	159,5	5088,05

2	42,0–64,0	53,0	7	69,3	686,07
3	64,0–86,0	75,0	4	48,4	585,64
4	86,0–108,0	97,0	2	68,2	2325,62
5	108,0–130,0	119,0	2	112,2	6294,42
Сумма:	–	–	20	457,6	14979,8

Определите относительные показатели вариации: а) коэффициент осцилляции (4.8):

$$K_R = \frac{R}{\bar{x}} \cdot 100\ \% = \frac{110}{62,9} \cdot 100\ \% = 174,88\ \%;$$

б) относительное линейное отклонение (4.9):

$$K_d = \frac{\bar{d}}{\bar{x}} \cdot 100\ \% = \frac{22,88}{62,9} \cdot 100\ \% = 36,37\ \%;$$

в) коэффициент вариации (4.10):

$$V = \frac{\sigma}{\bar{x}} \cdot 100\ \% = \frac{27,37}{62,9} \cdot 100\ \% = 43,51\ \%.$$

Коммерческие банки не однородны по размеру выдаваемых кредитов, так как коэффициент вариации равен 43,51 %, то есть больше 33 %.

Порядок выполнения работы

1. Ознакомьтесь с методикой вычислений средних значений статистических данных и абсолютных и относительных показателей вариации признака.
2. У преподавателя получите вариант индивидуального задания.
3. Проведите вычисления средних значений статистических данных, абсолютных и относительных показателей вариации признака.
4. Сделайте выводы по работе, и оформите отчет.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) алгоритм и результаты вычисления индивидуальных заданий;
3) анализ полученных результатов.

Контрольные вопросы

1. В чем заключается сущность и роль средней величины в статистике?
2. Какие виды средних применяются в статистике?
3. Как исчисляются средняя арифметическая простая и взвешенная?
4. Как исчисляются средняя гармоническая простая и взвешенная?
5. Как исчисляются структурные средние – мода и медиана?
6. Что представляет собой вариация признака, от чего зависят ее размеры?
7. Что представляет собой размах вариации признака?
8. Что такое дисперсия признака?
9. Какие показатели относительного рассеяния применяются в статистике?
10. Как исчисляется коэффициент вариации и его значение для экономического анализа?

Практическая работа № 5

Однофакторный корреляционный и регрессионный анализ (4 часа)

Цель работы

1. Изучить методику проведения однофакторного корреляционного и регрессионного анализа.
2. Сформировать практические навыки проведения однофакторного корреляционного и регрессионного анализа.
3. Построить уравнение регрессии, определить коэффициенты регрессии, коэффициент корреляции, коэффициент детерминации.
4. Оценить значимость коэффициентов регрессии и уравнения регрессии.

Краткая теория

Связь между явлениями классифицируется по ряду признаков, которые делятся на два класса: *факторные*, вызывающие изменения явлений, и *результативные*, изменяющиеся под влиянием факторных. Связи между явлениями и признаками классифицируются по *степени тесноты*, *направлению*, *аналитическому выражению* и *количеству факторов*, действующих на результативный признак.

Рассматривается выборка двух взаимосвязанных дискретных случайных величин X и Y. Пара (x_i, y_i), где $1 \le i \le n$ соответствует i-й точке (i-му опыту). Здесь n – объем парной выборки.

Для удобства последующего использования табличные (опытные) данные моделируют некоторой функцией, которую называют уравнением регрессии:

$$Y = f(X).$$

Процедура построения регрессионной (статистической) модели предусматривает, во-первых, выбор функции $f(X)$.

В качестве функции чаще всего используют полином:

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2 + \cdots + a_k \cdot X^k, \tag{5.1}$$

где a_j – коэффициенты регрессии $(0 \le j \le k)$; k – порядок полинома.

На втором этапе построения модели определяют коэффициенты регрессии a_k. Это осуществляется путем аппроксимации опытных точек.

Уравнение регрессии позволяет вычислить ожидаемое значение функции Y для опытных значений x_i:

$$Y(x_i) = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2 + \cdots + a_k \cdot x_i^k. \tag{5.2}$$

Разность между опытным значением y_i и ожидаемым значением $Y(x_i)$ составляет ошибку или погрешность функции:

$$\Delta Y_i = y_i - Y(x_i). \tag{5.3}$$

Аппроксимация может быть произведена при разных требованиях к величине ΔY_i. Наиболее распространенным является требование минимизации суммы квадратов отклонений опытных точек от линии регрессии. Это требование называют принципом Лежандра, согласно которому коэффициенты регрессии a_j должны быть подобраны так, чтобы сумма:

$$U = \sum_{i=1}^{n} (\Delta Y_i)^2 = \sum_{i=1}^{n} \left[y_i - a_0 - a_1 \cdot x_i - a_2 \cdot x_i^2 - \cdots - a_k \cdot x_i^k \right]^2 \tag{5.4}$$

принимала минимальное значение.

Метод определения коэффициентов регрессии по принципу Лежандра называют методом наименьших квадратов.

Искомые коэффициенты регрессии находятся из решения системы уравнений:

$$\frac{\partial U}{\partial a_j} = 0,$$

или

$$\frac{\partial U}{\partial a_0} = \sum_{i=1}^{n} (-2)(y_i - a_0 - a_1 x_i - \cdots - a_k x_i^k) = 0,$$

$$\frac{\partial U}{\partial a_1} = \sum_{i=1}^{n} (-2)(y_i - a_0 - a_1 x_i - \cdots - a_k x_i^k) x_i = 0,$$

…………………………………………………………

$$\frac{\partial U}{\partial a_k} \quad n$$

$$\frac{\partial \Phi}{\partial a_k} = -2\sum_{i=1}^{n}\left[y_i - \left(a_0 + a_1 x_i + \cdots + a_k x_i^k\right)\right] x_i^k = 0.$$

Отсюда получается система нормальных уравнений:

$$\begin{cases} n a_0 + a_1 \sum\limits_{i=1}^{n} x_i + \cdots + a_k \sum\limits_{i=1}^{n} x_i^k = \sum\limits_{i=1}^{n} y_i, \\ a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} x_i^2 + \cdots + a_k \sum\limits_{i=1}^{n} x_i^{k+1} = \sum\limits_{i=1}^{n} x_i y_i, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0 \sum\limits_{i=1}^{n} x_i^k + a_1 \sum\limits_{i=1}^{n} x_i^{k+1} + \cdots + a_k \sum\limits_{i=1}^{n} x_i^{2k} = \sum\limits_{i=1}^{n} x_i^k y_i. \end{cases} \tag{5.5}$$

В простейшем случае $k = 1$, то есть полинома первой степени, уравнение регрессии принимает вид:

$$Y = a_0 + a_1 \cdot X. \tag{5.6}$$

Система (5.5) также упрощается:

$$\begin{cases} n a_0 + a_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i, \\ a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases} \tag{5.7}$$

Уравнение (5.6) с коэффициентами регрессии учитывает погрешность функции и не учитывает погрешность фактора. Его называют уравнением прямой регрессии.

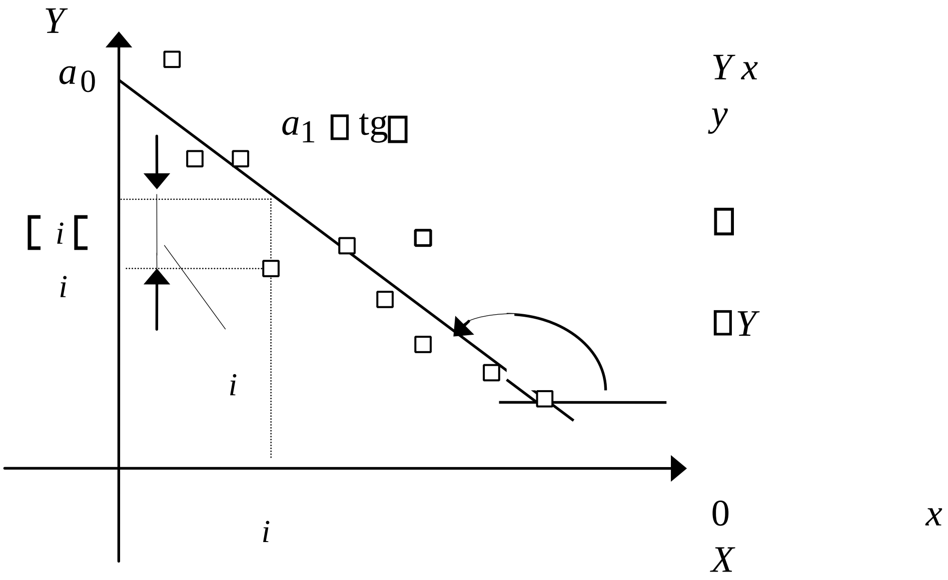


Рис. 5.1. Аппроксимация опытных данных линейным уравнением прямой регрессии

Решим систему (5.7) двух уравнений с двумя неизвестными a_0 и a_1:

$$a_1 = \frac{n\sum y_i x_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}; \qquad (5.8)$$

$$a_0 = \frac{1}{n}\left(\sum y_i - a_1 \sum x_i\right). \qquad (5.9)$$

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициента регрессии (коэффициента a_1).

Если знак при коэффициенте регрессии − положительный, связь зависимой переменной с независимой будет положительной. Если знак при коэффициенте регрессии − отрицательный, связь зависимой переменной с независимой является отрицательной (обратной).

Для анализа общего качества уравнения регрессии используют обычно *множественный коэффициент детерминации* R^2, называемый также квадратом коэффициента множественной корреляции R. R^2 (мера определенности) всегда находится в пределах интервала [0; 1].

Если значение R^2 близко к единице, это означает, что построенная модель объясняет почти всю изменчивость соответствующих переменных. И наоборот, значение R-квадрата, близкое к нулю, означает плохое качество построенной модели.

Коэффициент детерминации R^2 показывает, на сколько процентов ($R^2 \cdot 100$ %) найденная функция регрессии описывает связь между исходными значениями факторов X и Y:

$$R^2 = \frac{\sum_{i=1}^{n} (y_i^T - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

где $\sum (y_i^T - \bar{y})^2$ – объясненная вариация; $\sum (y_i - \bar{y})^2$ – общая вариация.

Соответственно, величина $(1 - R^2) \cdot 100\%$ показывает, сколько процентов вариации параметра Y обусловлены факторами, не включенными в регрессионную модель. При высоком ($R^2 \geq 75\%$) значении коэффициента детерминации можно делать прогноз $y^* = f(x^*)$ для конкретного значения x^*.

Множественный R − коэффициент множественной корреляции R − выражает степень зависимости независимых переменных (X) и зависимой переменной (Y) и равен квадратному корню из коэффициента детерминации, эта величина принимает значения в интервале от нуля до единицы. В простом линейном регрессионном анализе множественный R равен коэффициенту корреляции Пирсона, который вычисляется по формуле:

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}. \qquad (5.10)$$

Коэффициент корреляции может принимать значения в пределах

□1□r□1. Функциональной связи отвечает значение r □1. При r = 0 величины X и Y не зависят друг от друга. При □1□r□1 связь является вероятностной. Интерпретация значений r представлена в табл. 5.1, 5.2.

Таблица 5.1

Оценка линейного коэффициента корреляции r по характеру связи

Значение линейного коэффициента связи	Характер связи	Интерпретация связи
r = 0	Отсутствует	–
0 < r < 1	Вероятностная, прямая	С увеличением X увеличивается Y
□1□r□1	Вероятностная, обратная	С увеличением X уменьшается Y и наоборот
r = +1	Функциональная, прямая	Каждому значению факторного признака строго соответствует одно значение функции, с увеличением X увеличивается Y
r = -1	Функциональная, обратная	Каждому значению факторного признака строго соответствует одно значение функции, с увеличением X уменьшается Y и наоборот

Таблица 5. 2

Оценка коэффициента корреляции r по степени тесноты связи

Значение линейного коэффициента связи	Характер связи
До □ □ 0,3 □	Практически отсутствует
□ □ 0,3 □ – □ □ □0,5 □	Слабая
□ □ 0,5 □– □ □ 0,7 □	Умеренная
□ □ 0,7 □– □ □ 1,0 □	Сильная

Для практического использования моделей регрессии очень важна их адекватность, т.е. соответствие фактическим статистическим данным. Значимость коэффициентов простой линейной регрессии осуществляется с помощью t-критерия Стьюдента. При этом вычисляют расчетные значения t-критерия:

– для параметра a_0 $$t_{a_0} = a_0 \frac{\sqrt{n-2}}{\sigma_{ост}};\qquad (5.11)$$

– для параметра a_1 $$t_{a_1} = a_1 \frac{\sqrt{n-2}}{\sigma_{ост}} \sigma_x,\qquad (5.12)$$

где n – объем выборки;

$\sigma_{ост} = \sqrt{\frac{\sum_{i=1}^{n} (y - y^T)^2}{n}}$ – среднее квадратическое отклонение результативного признака y от выравненных значений y_T;

$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$ – среднее квадратическое отклонение факторного признака x от общей средней $\bar{x}$.

Вычисленные по формулам (5.11) и (5.12) значения, сравнивают с критическими $t_{\alpha,\nu}$, которые определяются по таблице Стьюдента (табл. 5.3) с учетом принятого уровня значимости α и числом степеней свободы вариации $\nu = n - m - 1$ (m – число факторных признаков в уравнении). Обычно в социально-экономических расчетах уровень значимости α принимается равным 0,05. При $t_{рас} > t_{\alpha,\nu}$ параметр является значимым (существенным). Если в уравнении все коэффициенты регрессии значимы, то данное уравнение признают окончательным и применяют в качестве модели изучаемого показателя для последующего анализа.

Таблица 5.3

Квантили распределения Стьюдента $t_{\alpha,\nu}$

□	Уровни значимости α			
	0,20	0,10	0,05	0,01
1	3,08	6,31	12,71	63,66
2	1,89	2,92	4,30	9,93
3	1,64	2,35	3,18	5,84
4	1,53	2,13	2,78	4,60
5	1,48	2,02	2,57	4,03
6	1,44	1,94	2,45	3,71
7	1,42	1,90	2,37	3,50
8	1,40	1,86	2,31	3,36
9	1,38	1,83	2,26	3,25
10	1,37	1,81	2,23	3,17
15	1,34	1,75	2,13	2,95
20	1,33	1,73	2,09	2,85
30	1,31	1,70	2,04	2,75
40	1,30	1,68	2,02	2,70

Проверка *значимости уравнения регрессии* производится на основе вычисления F-критерия Фишера:

$$F_{расч} = \frac{\sigma_y^2}{\sigma_{ост}^2} \cdot \frac{n - m - 1}{m},$$

где $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y - \bar{y})^2}{n}}$ – среднее квадратическое отклонение результативного признака y от общей средней $\bar{y}$.

Полученное значение – критерий $F_{расч}$ сравнивают с критическим (табличным) для принятого уровня значимости α и чисел степеней свободы $\nu_1 = m - 1$ и $\nu_2 = n - m$. Величины $F_{табл}$ при различных значениях ν_1, ν_1 и уровнях значимости α приведены в табл. 5.4. Уравнение регрессии значимо, если $F_{расч} > F_{табл}$.

Это означает, что доля вариации, обусловленная регрессией, намного превышает случайную ошибку. Принято считать, что уравнение регрессии пригодно для практического использования в том случае, если $F_{расч}$ *превышает* табличное *не менее чем в 4 раза*.

Таблица 5.4

Значения $F_{табл}$ по распределению Фишера при уровне значимости $\alpha = 0,05$

ν_2 \ ν_1	1	2	3	4	5	6
1	161,4	199,5	215,7	224,6	230,2	234,0
2	18,51	19,00	19,16	19,25	19,30	19,33
3	10,13	9,55	9,28	9,12	9,01	8,94
4	7,71	6,94	6,59	6,39	6,26	6,16
5	6,61	5,79	5,41	5,19	5,05	4,95
6	5,99	5,14	4,76	4,53	4,39	4,28
7	5,59	4,74	4,35	4,12	3,97	3,87
8	5,32	4,46	4,07	3,84	3,69	3,58
9	5,12	4,26	3,86	3,63	3,48	3,37
10	4,96	4,10	3,71	3,48	3,33	3,22
11	4,84	3,98	3,59	3,36	3,20	3,09
12	4,75	3,88	3,49	3,26	3,11	3,00
13	4,67	3,80	3,41	3,18	3,02	2,92
14	4,60	3,74	3,34	3,11	2,96	2,85

Методические рекомендации по выполнению лабораторной работы

Для проведения регрессионного анализа и прогнозирования необходимо:

1) ***построить график*** исходных данных и попытаться зрительно, приближенно определить характер зависимости;

2) ***выбрать вид функции*** регрессии, которая может описывать связь исходных данных;

3) ***определить численные коэффициенты*** функции регрессии методом наименьших квадратов;

4) ***оценить силу*** найденной регрессионной зависимости на основе коэффициента детерминации R^2;

5) ***сделать прогноз*** (при (R^2 □□75%) или сделать вывод о невозможности прогнозирования с помощью найденной регрессионной зависимости. При этом не рекомендуется использовать модель регрессии для тех значений независимого параметра X, которые не принадлежат интервалу, заданному в исходных данных.

Справочная информацця по технологии работы с режимом «Регрессия» надстройки Пакет анализа MS Excel

Режим работы «Регрессия» служит для расчета параметров уравнения линейной регрессии и проверки его адекватности исследуемому процессу.

Для решения задачи регрессионного анализа в MS Excel выберите в меню **Сервис** команду **Анализ данных** и инструмент анализа «**Регрессия**».

В появившемся диалоговом окне задаем следующие параметры:

1. *Входной интервал Y* – это диапазон данных по результативному признаку. Он должен состоять из одного столбца.

2. *Входной интервал X* – это диапазон ячеек, содержащих значения факторов (независимых переменных). Число входных диапазонов (столбцов) должно быть не больше 16.

3. Флажок *Метки*, установите в том случае, если в первой строке диапазона стоит заголовок.

4. Флажок *Уровень надежности* активизируется, если в поле, находящееся рядом с ним введете уровень надежности, отличный от установленного по умолчанию. Используется для проверки значимости коэффициента детерминации R^2 и коэффициентов регрессии.

5. *Константа ноль.* Данный флажок установите, если линия регрессии должна пройти через начало координат ($a_0 = 0$).

6. *Выходной интервал/ Новый рабочий лист/ Новая рабочая книга* – укажите адрес верхней левой ячейки выходного диапазона.

7. Флажки в группе *Остатки* установите, если необходимо включить в выходной диапазон соответствующие столбцы или графики.

8. Флажок График нормальной вероятности сделайте активным, если требуется вывести на лист точечный график зависимости наблюдаемых значений *Y* от автоматически формируемых интервалов персентилей.

После нажатия кнопки ОК в выходном диапазоне получите отчет.

Методические указания по проведению однофакторного корреляционного и регрессионного анализа

Задача

Некоторая фирма занимается поставками различных грузов на короткие расстояния внутри города. Оценить стоимость таких услуг, зависящую от затрачиваемого на поставку времени. В качестве наиболее важного фактора, влияющего на время поставки, выбрано пройденное расстояние. Исходные данные о десяти поставках приведены в табл. 5.5.

Таблица 5.5

Данные о времени поставок и пройденном расстоянии

Расстояние, км	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Время, мин	16	13	19	18	12	11	8	14	9	16

Определите характер зависимости между расстоянием и затраченным временем, используя мастер диаграмм MS Excel, проанализируйте применимость метода наименьших квадратов, постройте уравнение регрессии, используя МНК, проанализируйте силу регрессионной связи. Проведите регрессионный анализ, используя режим работы «Регрессия» в MS Excel и сравните с результатами, полученными ранее. Посчитать и построить графически меру ошибки регрессионной модели, используя табличный процессор Excel.

Решение

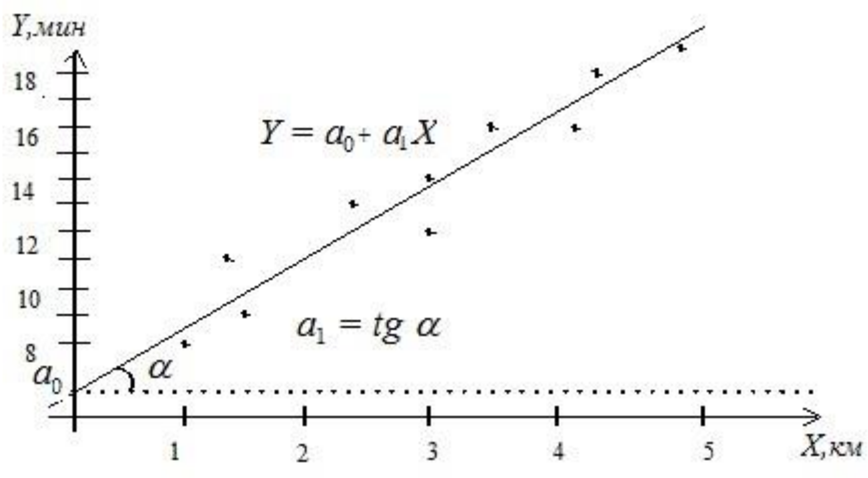


Рис. 5.3. График исходных данных и предполагаемая линия регрессии

На графике постройте исходные данные по десяти поездкам.

Помимо расстояния на время поставки влияют пробки на дорогах, время суток, дорожные работы, погода, квалификация водителя, вид транспорта. Построенные точки не находятся точно на линии, что обусловлено описанными выше факторами. Но эти точки собраны вокруг прямой линии, поэтому можно предположить линейную связь между параметрами. Все исходные точки равномерно распределены вдоль предполагаемой прямой линии, что позволяет применить метод наименьших квадратов.

Вычислите суммы, необходимые для расчета коэффициентов уравнения линейной регрессии и коэффициента детерминации R^2 с помощью вспомогательной таблицы (табл. 5.6).

Таблица 5.6

Расчетная таблица

x_i	y_i	x_i^2	$x_i y_i$	y_i^T	$(y_i^T - \bar{y})^2$	$(y_i - \bar{y})^2$
3,5	16	12,25	56,00	15,223	2,634129	5,76
2,4	13	5,76	31,2	12,297	1,697809	0,36
4,9	19	24,01	93,1	18,947	28,59041	29,16
4,2	18	17,64	75,60	17,085	12,14523	19,36
3,0	12	9,00	36,00	13,893	0,085849	2,56
1,3	11	1,69	14,30	9,371	17,88444	6,76
1,0	8	1,00	8,00	8,573	25,27073	31,36

3,0	14	9,00	42,00	13,893	0,085849	0,16
1,5	9	2,25	13,50	9,903	13,66781	21,16
4,1	16	16,81	65,60	16,819	10,36196	5,76
Σ=28,9	Σ=136	Σ=99,41	Σ=435,30	–	112,4242	122,4

Среднее значение *y* вычислите по формуле:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{16+13+19+18+12+11+8+14+9+16}{10} = 13,6.$$

Вычислите коэффициенты линейной регрессии по формулам (5.8) и (5.9):

$$a_1 = \frac{10 \cdot 435,30 - 136 \cdot 28,9}{10 \cdot 99,41 - 835,21} = 2,660;$$

$$a_0 = 0,1(136 - 2,660 \cdot 28,9) = 5,913.$$

Таким образом, искомая регрессионная зависимость имеет вид:

$$y^T = 5,913 + 2,660x.$$

Наклон линии регрессии $a_1 = 2,66$ минут на км. – это количество минут, приходящееся на один км расстояния. Координата точки пересечения прямой с осью *Y* $a_0 = 5,913$ минут – это время, которое не зависит от пройденного расстояния, а обуславливается всеми остальными возможными факторами, явно не учтенными при анализе. Вычислите коэффициент детерминации:

$$R^2 = \frac{112,424}{122,400} = 0,916 \text{ или } 91,8\ \%.$$

Проведите регрессионный анализ с использованием режима Регрессия MS Excel. Значения параметров, установленных в одноименном диалоговом окне, представлены на рис. 5.4.

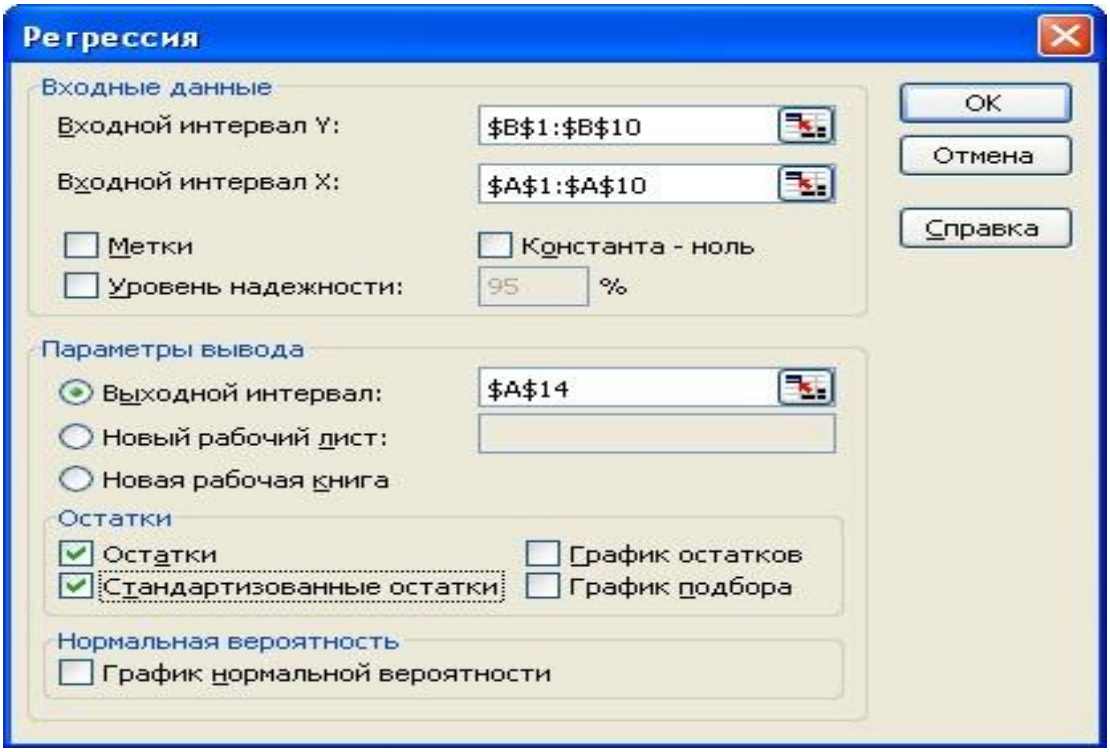


Рис. 5.4 Окно входных данных

Сгенерируются результаты по регрессионной статистике, представленные в таблице 5.7.

Таблица 5.7

Вывод итогов

Регрессионная статистика	
Множественный *R*	0,958275757
R-квадрат	0,918292427
Нормированный *R*-квадрат	0,90807898
Стандартная ошибка	1,11809028
Наблюдения	10

Рассмотрите представленную в таблице 5.7 регрессионную статистику.

Величина *R*-квадрат, называемая также мерой определенности, характеризует качество полученной регрессионной прямой. Это качество выражается степенью соответствия между исходными данными и регрессионной моделью (расчетными данными). Мера определенности всегда находится в пределах интервала [0; 1]. Мера определенности равная 0,91829 говорит об очень хорошей подгонке регрессионной прямой к исходным данным и совпадает с коэффициентом детерминации R^2, вычисленным по формуле.

Таким образом, линейная модель объясняет 91,8 % вариации времени доставки, что означает правильность выбора фактора (расстояния). Не объясняется 100 % − 91,8 % = 8,2 % вариации времени поездки, которые обусловлены остальными факторами, влияющими на время поставки, но не включенными в линейную модель регрессии.

Рассчитанный уровень значимости α_p = 1,26E-05 < 0,05 (показатель значимость F в таблице Дисперсионный анализ подтверждает значимость R^2.

Множественный R – коэффициент множественной корреляции R – выражает степень зависимости независимых переменных (X) и зависимой переменной (Y) и равен квадратному корню из коэффициента детерминации, эта величина принимает значения в интервале от нуля до единицы. В простом линейном регрессионном анализе множественный R равен коэффициенту корреляции Пирсона. Действительно, множественный R в нашем случае равен коэффициенту корреляции Пирсона (0,95827), который вычисляется по формуле:

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - \left(\sum x\right)^2\right)\left(n\sum y^2 - \left(\sum y\right)^2\right)}}.$$

Теперь рассмотрите среднюю часть расчетов, представленную в таблице 5.8 (приведена в сокращенном варианте). Здесь даны коэффициент регрессии a_1 (2,65970168) и смещение по оси ординат, то есть константа a_0 (5,913462144).

Таблица 5.8

Результаты регрессионального анализа

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*
Y-пересечение	5,913462144	0,884389599	6,686489927	0,00015485
Переменная *X*	2,65970168	0,280497238	9,482095791	1,26072E-05

Исходя из расчетов, запишите уравнение регрессии таким образом: y^{T}□ 5,913□ 2,660x.

Это уравнение, совпадает с уравнением, полученным при расчете по МНК вручную с точностью до ошибки округления.

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициента регрессии

(коэффициента $a1$). Знак коэффициента регрессии положительный (+2,660), следовательно, связь также является положительной.

Далее проверьте значимость коэффициентов регрессии: $a0$ и $a1$. Сравните попарно значения столбцов *Коэффициенты* и *Стандартная ошибка* в таблице 5.8, видно, что абсолютные значения коэффициентов больше чем их стандартные ошибки. К тому же эти коэффициенты являются значимыми, о чем можно судить по значениям показателя

Р-значение в таблице 5.8, которые меньше заданного уровня значимости α = 0,05.

В таблице 5.9 представлены результаты вывода остатков. При помощи этой части отчета определите отклонения каждой точки от построенной линии регрессии. Наибольшее абсолютное значение остатка – 1,89256, наименьшее – 0,05399. Для лучшей интерпретации этих данных воспользуйтесь графиком исходных данных и построенной линией регрессии, представленными на рис. 5.5. Как видно, линия регрессии хорошо «подогнана» под значения исходных данных.

Таблица 5.9

Результаты анализа остатков

Вывод остатка			
Наблюдение	*Предсказанное Y*	*Остатки*	*Стандартные остатки*
1	15,22241803	0,777581975	0,737641894
2	12,29674618	0,703253823	0,667131568
3	18,94600038	0,053999622	0,051225961
4	17,0842092	0,915790799	0,868751695
5	13,89256718	– 1,892567185	–1,795356486
6	9,371074328	1,628925672	1,545256778
7	8,573163824	–0,573163824	– 0,543723571

8	13,89256718	0,107432815	0,101914586
9	9,903014664	– 0,903014664	– 0,8566318
10	16,81823903	– 0,818239033	– 0,776210624

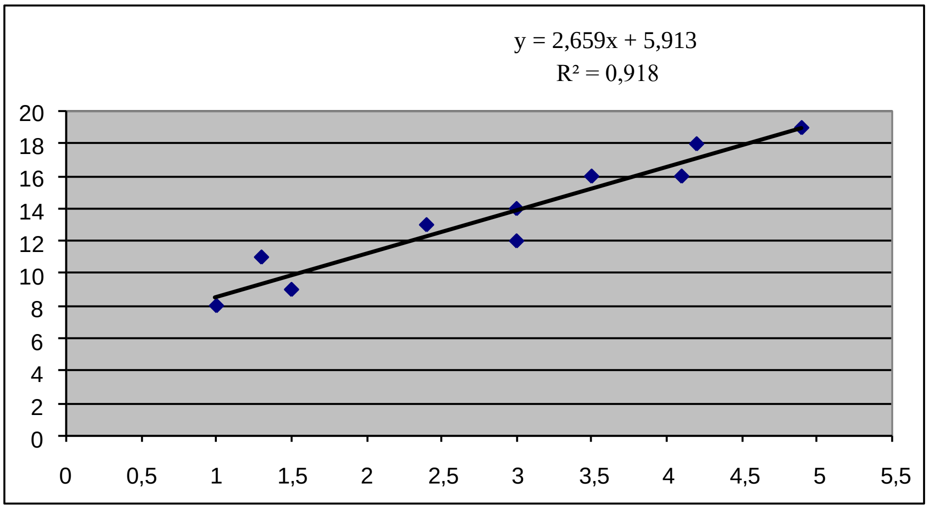


Рис. 5.5. Исходные данные и линия регрессии

Приблизительным, но самым простым и наглядным способом проверки удовлетворительности регрессионной модели является графическое представление отклонений.

Отложите отклонения $(y_i^T - y_i)$ по оси Y, для каждого значения y_i. Если регрессионная модель близка к реальной зависимости, то отклонения будут носить случайный характер и их сумма будет близка к нулю.

В рассмотренном примере $\sum_{i=1}^{n} (y_i^T - y_i) = 0,004.$

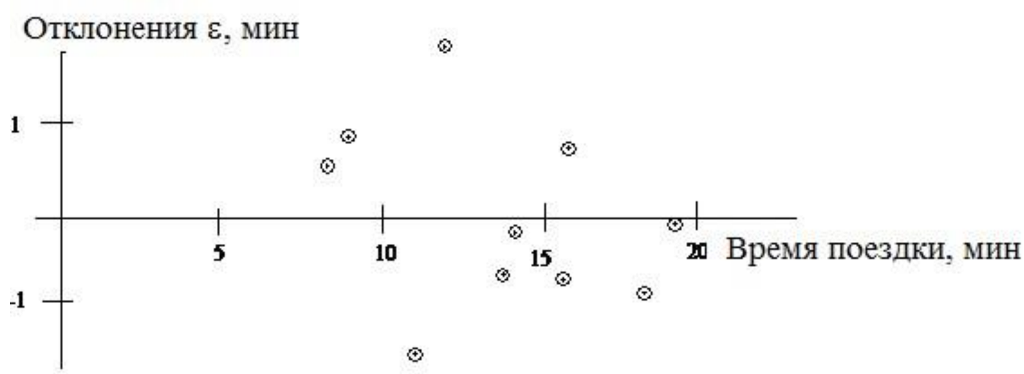


Рис. 5.6. График отклонений

Таким образом, в результате использования регрессионного анализа в табличном процессоре MS Excel:

- построили уравнение регрессии;
- установили форму зависимости и направление связи между переменными – положительная линейная регрессия, которая выражается в равномерном росте функции;
- установили направление связи между переменными;
- оценили качество полученной регрессионной прямой;
- смогли увидеть отклонения расчетных данных от данных исходного набора.

Порядок выполнения работы

1. Ознакомьтесь с методикой проведения корреляционного и регрессионного анализа в Excel.
2. У преподавателя получите вариант индивидуального задания.
3. Парную выборку опытных данных (x_i, y_i) нанесите на график и визуально оцените применимость линейного уравнения регрессии.
4. Вычислите коэффициенты прямой линейной регрессий и коэффициент корреляции.
5. На график опытных точек нанесите рассчитанную линию прямой линейной регрессии. Визуально оцените близость уравнения регрессии к функциональной связи.
6. Сделайте выводы по работе и оформить отчет.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) результаты вычислений индивидуальных заданий;
3) анализ полученных результатов табличном процессоре MS Excel.

Варианты индивидуальных заданий

Постройте регрессионную модель (линейную) для исходных данных, приведенных в таблице 5.10.

Таблица 5.10

Индивидуальные задания

i	Номера контрольных задач									
	1		**2**		**3**		**4**		**5**	
	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i
1	2,62	0,90	15,57	75,53	– 44,2	5,85	45,11	43,20	93,53	30,76
2	2,44	0,71	15,41	75,38	– 44,6	5,70	44,90	43,03	93,37	30,59
3	2,25	0,56	15,22	75,20	– 44,7	5,52	44,75	42,86	93,16	30,43
4	2,06	0,38	15,03	74,04	– 44,8	5,36	44,56	42,72	92,99	30,29
5	1,89	0,21	14,85	74,87	– 45,0	5,19	44,38	42,54	92,81	30,10
6	1,71	0,06	14,68	74,71	– 45,2	5,03	44,20	42,38	92,63	29,94
7	1,53	– 0,11	14,50	74,53	– 45,4	4,86	44,02	42,21	92,45	29,77
8	1,40	– 0,13	14,36	74,52	– 45,5	4,84	43,89	42,19	92,32	29,75
9	1,26	– 0,14	14,23	74,51	– 45,6	4,83	43,75	42,18	92,18	29,74
10	1,13	– 0,16	14,11	74,48	– 45,8	4,81	43,62	42,16	92,05	29,72
11	1,00	– 0,18	13,97	74,47	– 45,9	4,79	43,49	42,14	91,92	29,70
12	0,86	– 0,19	13,83	74,46	– 46,0	4,78	43,35	42,13	91,78	29,69
13	0,73	– 0,21	13,70	74,44	– 46,2	4,76	43,22	42,11	91,65	29,67
14	0,62	– 0,15	13,59	74,50	– 46,3	4,82	43,11	42,17	91,54	29,73

i	Номера контрольных задач									
	6		**7**		**8**		**9**		**10**	
	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i
1	– 17,9	– 69,9	173,2	80,9	14,6	11,8	27,2	7,6	– 10,2	51,5
2	– 18,0	– 70,0	179,5	81,8	14,4	11,6	20,3	7,2	– 10,4	51,4
3	– 18,2	– 70,1	185,8	82,6	14,3	11,4	13,4	6,8	– 10,6	51,2
4	– 18,5	– 70,3	176,1	87,4	14,1	11,2	16,5	13,4	– 10,9	51,0
5	– 18,7	– 70,5	179,0	93,9	13,9	11,0	19,7	20,1	– 11,1	50,9
6	– 18,9	– 70,6	180,0	86,7	13,7	10,9	20,8	12,3	– 11,3	50,7
7	– 19,0	– 70,8	181,1	79,5	13,5	10,7	21,9	4,5	– 11,4	50,5
8	– 19,2	– 70,9	180,0	79,8	13,4	10,6	20,2	4,6	– 11,6	50,4
9	– 19,3	– 71,0	178,9	80,1	13,3	10,5	18,7	4,7	– 11,7	50,3
10	– 19,4	– 71,1	181,0	86,9	13,1	10,4	20,9	12,5	– 11,8	50,1
11	– 19,6	– 71,2	183,2	93,8	13,0	10,2	23,0	20,2	– 12,0	49,9
12	– 19,7	– 71,4	181,0	89,3	12,9	10,0	21,1	14,9	– 12,1	49,7
13	– 19,8	– 71,4	178,9	84,8	12,7	9,9	19,0	9,9	– 12,2	49,5
14	– 20,0	– 71,5	183,5	87,5	12,6	9,7	23,9	12,0	– 12,4	49,4

Окончание табл. 5.10

	Номера контрольных задач

i	11		12		13		14		15	
	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i
1	12,9	62,9	17,2	80,7	– 14,6	11,9	26,2	– 7,6	10,2	61,5
2	13,0	63,0	17,5	81,6	– 14,4	11,7	19,3	– 7,2	10,4	61,4
3	13,2	63,1	18,8	82,4	– 14,3	11,5	12,4	– 6,8	10,6	61,2
4	13,5	63,3	17,1	87,2	– 14,1	11,3	15,5	– 13,4	10,9	61,0
5	13,7	63,5	17,0	93,7	– 13,9	11,1	18,7	– 20,1	11,1	60,9
6	13,9	63,6	18,0	86,5	– 13,7	10,9	19,8	– 12,3	11,3	60,7
7	14,0	63,8	18,1	79,3	– 13,5	10,8	20,9	– 4,5	11,4	60,5
8	14,2	63,9	18,0	79,6	– 13,4	10,7	19,2	– 4,6	11,6	60,4
9	14,3	64,0	17,9	80,0	– 13,3	10,5	17,7	– 4,7	11,7	60,3
10	14,4	64,1	18,0	86,7	– 13,1	10,4	19,9	– 12,5	11,8	60,1
11	14,6	64,2	18,2	93,6	– 13,0	10,2	22,0	– 20,2	12,0	59,9
12	14,7	64,4	18,0	89,1	– 12,9	10,0	20,1	– 14,9	12,1	59,7
13	14,8	64,4	17,9	84,6	– 12,7	9,9	18,0	– 9,9	12,2	59,5
14	15,0	64,5	18,5	87,3	– 12,6	9,7	22,9	– 12,0	12,4	59,4

Контрольные вопросы

1. Какие основные задачи решают с помощью корреляционного и регрессионного анализа?
2. Сформулируйте принцип Лежандра.
3. Какими показателями измеряется теснота корреляционной связи?
4. В чем отличие стохастической связи от функциональной?
5. В чем состоит значение уравнения регрессии? Что характеризуют коэффициенты регрессии?
6. Для чего нужен коэффициент корреляции? В каких пределах он изменяется?
7. Как осуществляется проверка значимости коэффициентов регрессии?
8. Как проверить адекватность уравнения в целом?

9. В каких случаях применяется модель множественной регрессии?
10. Как проводится корреляционный и регрессионный анализ в MS Excel?

Практическая работа № 6

Многофакторный корреляционно-регрессионный анализ в пакете Statistica (4 часа)

Цель работы

1. Познакомиться с методикой выполнения корреляционнорегрессионного анализа в пакете Statistica.
2. Определить парные, частные и множественные коэффициенты корреляций и на их основе выявить факторы, оказывающие наибольшее влияние на результативный показатель.
3. Получить регрессионную модель и провести ее полный статистический анализ.

Краткая теория

Статистические данные всегда являются приближенными, усредненными. Поэтому они носят оценочный характер и для достоверности результатов необходимо большое число исходных данных. Существует несколько видов статистического анализа данных: корреляционный, регрессионный, дисперсионный, факторный, кластерный и др. Рассмотрим некоторые из них.

Корреляционный анализ

Иногда корреляцию и регрессию рассматривают как совокупный процесс статистического исследования. Корреляционно-регрессионный анализ является одним из значимых методов построения математических моделей в экономике и считается одним из главных методов в маркетинге.

Корреляция в широком смысле слова означает связь между объективно существующими явлениями. Корреляционный анализ – вид статистического анализа, который состоит в количественной оценке силы и направления связи между двумя (парная корреляция) или несколькими (множественная корреляция) наборами данных. В эконометрике различают следующие варианты зависимостей:

- ***парная корреляция*** – связь между двумя признаками,
один из которых результативный, а другой факторный;
- ***частная корреляция*** – зависимость между результативным и одним факторным признаком, при фиксированном значении других факторных признаков;
- ***множественная корреляция*** – зависимость результативного признака от нескольких факторных признаков.
- ***каноническая корреляция*** – зависимость группы результативного признака от группы факторных признаков.

Для наглядности измерения всех связей в случае множественной корреляции целесообразно использовать корреляционную матрицу – матрицу из попарных коэффициентов корреляции.

Использование методов корреляции позволяет решить следующие задачи:

– установить абсолютное изменение результативного признака за счет изменения одного или комплекса факторов;

– определить меру зависимости между результативным признаком и одним из факторов при постоянном значении других;

– установить меру относительного изменения зависимой переменной на единицу относительного изменения фактора или факторов;

– изучить общий объем вариации результативного признака и определить роль каждого фактора в объяснении этого изменения;

– оценить статистическую надежность выборочных показателей корреляционной связи.

Таким образом, коэффициент корреляции – это инструмент, с помощью которого можно проверить гипотезу о зависимости и измерить силу зависимости двух переменных. Если распределение переменных нормальное или несущественно отличается от нормального, применяют линейный коэффициент корреляции Пирсона. Для порядковых (ранговых) переменных или переменных, чье распределение существенно отличается от нормального, используется коэффициент корреляции Спирмена или Кендалла.

Коэффициенты корреляции рангов Спирмена и Кендалла могут быть использованы для определения связи как между *количественными*, так и между *качественными* признаками при условии, если их значения можно упорядочить или проранжировать.

Коэффициент корреляции рангов Спирмена (□) рассчитывается по формуле:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}, \qquad (6.1)$$

где $d = N_x - N_y$, то есть разности рангов каждой пары значений x и y, а n – число наблюдений.

Коэффициент корреляции рангов Кендалла (τ) определяется по формуле

$$\tau = \frac{2S}{n(n - 1)}, \qquad (6.2)$$

где S – сумма положительных и отрицательных баллов.

Во многих экономических задачах, например, при выборе инвестиционных проектов, достаточно именно монотонной зависимости одной переменной от другой. Отметим, что коэффициент ранговой корреляции Спирмена остается постоянным при любом строго возрастающем преобразовании шкалы измерения результатов наблюдений. Другими словами, он является адекватным в порядковой шкале, как и другие ранговые статистики.

Коэффициенты ранговой корреляции Спирмена и Кендалла – мера линейной связи между случайными величинами. Для оценки силы связи между величинами используются не численные значения, а соответствующие им ранги. Эти коэффициенты определяют степень тесноты и направленность связи признаков. Величина коэффициентов лежит в интервале от +1 до –1. Абсолютное значение характеризует тесноту связи, а знак «минус» направленность связи между двумя признаками.

Преимущество – можно ранжировать по признакам, которые нельзя выразить численно: субъективные оценки, предпочтения и т.д. При экспертных оценках можно ранжировать оценки разных экспертов и найти их корреляции друг с другом, чтобы затем исключить из рассмотрения оценки эксперта, слабо коррелирующие с оценками других. Коэффициент корреляции рангов применяется для оценки устойчивости тенденции динамики.

Коэффициенты ранговой корреляции (как и большинство непараметрических оценок) менее чувствительны к выбросам и погрешностям в результатах наблюдений и, в этом смысле, являются более устойчивыми и надежными мерами взаимозависимости по сравнению с коэффициентом Пирсона. Замечено, что в большинстве случаев коэффициент Спирмена больше коэффициента Кендалла.

Регрессионный анализ

Регрессионный анализ – вид статистического анализа, который состоит в представлении зависимости одних факторов от других в виде некоторой функции (уравнения регрессии), с помощью которой осуществляется прогнозирование и поиск ответа на вопросы «Что будет через какое-то время?» или «Что будет, если…?».

В случае парной регрессии уравнение определяется по двум наборам данных, один из которых представляет значения зависимой переменной Y, а другой – независимой переменной X. В случае множественной регрессии уравнение определяется по нескольким наборам данных, один из которых представляет значения зависимой переменной Y, а другие – независимыми переменными $X1$, $X2,…, Xn$. Получение уравнения регрессии происходит в два этапа: подбор вида функции и вычисление параметров функции.

Выбор функции делается на основе некоторых известных физических, химических, экономических (и т.п.) свойств рассматриваемого процесса или на основе иных соображений. В частности, если изучается зависимость между двумя величинами, т.е. ищется аппроксимирующая функция $Y = f(X)$, то можно нанести экспериментальные точки (X_i, Y_i), $i = 1, 2, \ldots, n$, на координатную плоскость и по характеру расположения этих точек сделать предположения о структуре аппроксимирующей функции. Выбор функции, в большинстве случаев, производится среди линейной, квадратичной, степенной и других видов (табл. 6.1).

Таблица 6.1

Виды функций

Парная (простая) регрессия	Множественная регрессия
Линейная регрессия	
$y = ax + b$,	$y = a_0 + a_1x_1 + \ldots + a_mx_m$
Квадратичная (параболическая)	
$y = ax^2 + bx + c$	$y = a_0 + a_1x_1^2 + \ldots + a_m x_m^2$
Степенная	
$y = ax^b$	$y = a0\ x1\ a1\ x2\ a2\ldots\ xm\ am$
Логарифмическая $y = a\ln x + b$	Гиперболическая

Экспоненциальная $y=ae^{bx}$	$y = a_0 + a_1 (1/x_1) + \ldots + a_m(1/x_m)$
где a, b, c – коэффициенты парной регрессии	где a_0, a_1, $a_2, \ldots$, a_m – коэффициенты множественной регрессии, n – объем совокупности, m – количество факторных признаков.

К функции предъявляются следующие требования:

- она должна быть достаточно простой для использования ее в дальнейших вычислениях;
- график этой функции должен проходить вблизи экспериментальных точек так, чтобы сумма квадратов отклонений y – координаты всех экспериментальных точек от y – координат графика функции была бы минимальной (метод наименьших квадратов).

На втором этапе определяют неизвестные коэффициенты аппроксимирующей функции из условия минимума среднего квадрата ошибки прогнозирования. Основным методом решения задачи нахождения параметров $a0$ и $a1$ уравнения регрессии является метод наименьших квадратов (МНК). Он состоит в минимизации суммы квадратов отклонений фактических значений от значений, вычисленных по уравнению регрессии.

Основным параметром парного уравнения регрессии является параметр a_1 (в случая множественной регрессии a_j, где $j = 1, 2, \ldots, n$), который характеризует силу связи между вариацией факторного признака X и вариацией результативного признака Y.

Иногда в эконометрических исследованиях возникают ситуации, в которых использование параметров a_j не дает желаемого результата, так как коэффициент имеет размерность, совпадающую с анализируемым показателем и не пригоден для выявления наибольшего (наименьшего) влияния той или иной независимой переменной. В этом случае используют β – коэффициент или коэффициент эластичности.

С помощью β-*коэффициентов* определяют факторы, в которых заложены наиболее крупные резервы улучшения изучаемого показателя. β-коэффициенты учитывают различия в степени варьирования вошедших в уравнение факторов и вычисляются по формуле:

$$\beta_j = a_j \frac{\sigma_{x_j}}{\sigma_y}, \qquad (6.3)$$

где σ_{x_i} – среднее квадратическое отклонение j-го фактора; σ_y – среднее квадратическое отклонение результативного признака.

β-*коэффициены показывают*, на какую часть среднего квадратического отклонения изменяется результативный признак с изменением соответствующего факторного признака на величину его среднего квадратического отклонения.

Частные коэффициенты эластичности, устраняющие различия в единицах измерения факторов, определяются как:

$$Э_j(\%) = a_j \frac{\bar{x}_j}{\bar{y}}, \qquad (6.4)$$

где a_j – коэффициент регрессии при j-м факторе; $\bar{x}_j$ – среднее значение j-го фактора; $\bar{y}$ – среднее значение результативного признака.

Частные коэффициенты эластичности показывают, на сколько процентов в среднем изменяется анализируемый показатель при изменении на 1 % каждого фактора при фиксированном значении других факторов.

В большинстве случаев при построении модели приходится пользоваться выборочными данными, поэтому прежде чем приступать к использованию модели необходимо убедится в ее адекватности фактическим данным (анализируемому явлению). Для этих целей используют t-критерий Стьюдента и F-критерий Фишера.

Для количественной оценки точности построения уравнения регрессии предназначен коэффициент детерминации R^2, равный квадрату коэффициента множественной корреляции и указывающий, какой процент изменения функции Y объясняется воздействием факторов X_i. Чем его значение ближе к 1, тем уравнение точнее описывает исследуемую зависимость.

Значимое уравнение (с R^2 близким к 1*)* используется, как правило, для прогнозирования изучаемого явления.

Прогноз – это вероятностное суждение о будущем, полученное путем использования совокупности научных методов. Например, прогнозирование финансового состояния выполняется для того, чтобы получить ответы на два вопроса: «как это может быть (какими могут стать финансовые показатели, если не будут приняты меры по их изменению)» и «как это должно быть (какими должны стать финансовые показатели фирмы для того, чтобы ее финансовое состояние обеспечивало высокий уровень конкурентоспособности)». Прогнозирование с целью получения ответа на первый вопрос принято называть исследовательским, на второй – нормативным.

Существует два способа прогнозов по уравнению регрессии: в пределах экспериментальных значений (интерполяция) и за пределами (экстраполяция). Применимость всякой регрессионной модели ограничена, особенно за пределами экспериментальной области, так как характер зависимости может существенно измениться. Поэтому достоверность исследовательского прогноза может быть невысокой. Однако его выполнение полностью обосновано.

Задание

Руководство компании по результатам производственной деятельности 15 своих филиалов в различных городах России за год (табл. 6.2) анализирует факторы, влияющие на производительность труда (y) и предполагает, что важнейшими из них являются следующие:

$x1$ – среднегодовая стоимость основных фондов, тыс. руб.; $x2$ – удельный вес рабочих высокой квалификации в общей численности рабочих, %; $x3$ – трудоемкость единицы продукции; $x4$ – среднегодовая численность рабочих; $x5$ – коэффициент сменности оборудования; $x6$ – удельный вес потерь от брака; $x7$ – среднегодовой фонд заработной платы, тыс. руб.

Таблица 6.2

Исходные данные

	Город	y	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x7$
1	Москва	14	101,03	35	0,4	15780	2,01	0,22	13002
2	С.-Петербург	14,02	98,54	36	0,42	14760	1,86	0,25	10145,6
3	Н. Новгород	7,03	49	17	1,83	630	0,95	0,5	5040,9
4	Ульяновск	7,01	50	17	1,85	633	0,93	0,52	5027,39

5	Пенза	8,21	57,42	19	1,43	752	1,08	0,44	5903,3
6	Самара	10	70	24	1,01	920	1,33	0,35	7100
7	Чебоксары	9,02	61,03	22	1,23	830	1,19	0,39	6494,6
8	Саранск	11	78,09	26	0,82	1028	1,44	0,37	7500
9	Челябинск	9,05	63,31	28	1,2	804	1,2	0,38	6516,5
10	Тольятти	11	77,05	29	0,81	1028	1,46	0,32	7940
11	Волгоград	12	84,03	27	0,64	1126	1,6	0,29	8900
12	Рязань	12	83	29	0,66	1127	1,59	0,25	8668
13	Красноярск	12	84	30	0,68	1096	1,59	0,29	8670,91
14	Тула	7,26	50,81	17	1,75	657	0,96	0,49	5209,8
15	Казань	7,01	55,01	16	1,85	631	0,93	0,51	5027,3

Таблица 6.3

Варианты индивидуальных заданий

Вариант №	Независимые переменные (факторные признаки)	Задания по прогнозированию
		Как изменится производительность труда на московском предприятии, если:
1	x_1, x_3, x_4, x_5	среднегодовую стоимость основных фондов увеличить на 80 тыс. руб., а и трудоемкость единицы продукции на 0,6?
2	x_3, x_4, x_5, x_6	трудоемкость единицы продукции сократить в 4 раза, а коэффициент сменности оборудования снизить в 2 раза?
3	x_1, x_2, x_3, x_5	среднегодовую стоимость основных фондов увеличить на 60 тыс. руб., а коэффициент сменности оборудования – на 0,9?
4	x_1, x_2, x_6, x_7	среднегодовую стоимость основных фондов сократить до 90 тыс. руб., а удельный вес потерь от брака понизить в 2 раза?
5	x_1, x_3, x_4, x_7	среднегодовую стоимость основных фондов сократить до 95 тыс. руб., а трудоемкость единицы продукции понизить на 0,1?

6	x_1, x_2, x_5, x_7	коэффициент сменности оборудования увеличить в 2 раза, а среднегодовой фонд заработной платы уменьшить на 92 тыс. руб.?
7	x_4, x_5, x_6, x_7	коэффициент сменности оборудования уменьшить в 2 раза, а среднегодовой фонд заработной платы увеличить на 92 тыс. руб.
8	x_2, x_3, x_5, x_7	коэффициент сменности оборудования увеличить на 1,5, а среднегодовой фонд заработной платы уменьшить на 32 тыс. руб.?
9	x_1, x_3, x_5, x_7	коэффициент сменности оборудования уменьшить на 1,5, а среднегодовой фонд заработной платы увеличить на 32 тыс. руб.?
10	x_1, x_2, x_4, x_5	среднегодовую численность рабочих сократить на 780 человек, а коэффициент сменности оборудования повысить до 3?

Порядок выполнения задания:

1. **Корреляционный анализ:**

– постройте диаграммы рассеяния;

– получите матрицу парных коэффициентов корреляции (Пирсона);

– вычислите частные коэффициенты корреляции;

– постройте матричную диаграмму рассеяния;

– вычислите ранговые коэффициенты корреляции Спирмена и Кендалла;

– выявите факторы, оказывающие наибольшее влияние на результативный показатель.

2. **Регрессионный анализ**

– получите линейное уравнение множественной регрессии, выбрав в качестве зависимой переменной – Y, в качестве независимых переменных X_i, соответствующие варианту;

– проанализируйте множественные коэффициенты корреляции и детерминации полученной модели;

– проверьте значимость построенной модели, используя уровень значимости $\alpha = 0{,}05$;

– если модель значима, дайте оценку коэффициентов множественной регрессии на основе t-критерия Стьюдента;

– пересчитайте уравнение множественной регрессии, используя только значимые факторы;

– проверьте адекватность полученной регрессионной модели; – выполните прогнозирование в соответствии с вариантом задания.

Методические указания по выполнению задания

Выберите исходные данные для своего варианта из таблиц 6.2, 6.3, скопируйте и вставьте в электронную таблицу пакета *Statistica* [5].

Статистический пакет *Statistica* использует стандартный интерфейс электронных таблиц. Текущий файл данных всегда отображается в виде электронной таблицы. Данные организованы в виде наблюдений и переменных. Наблюдения можно рассматривать как эквивалент столбцов электронной таблицы. Каждое наблюдение состоит из набора значений переменной. Система состоит из ряда модулей, работающих независимо. Каждый модуль включает определенный класс процедур. Почти все процедуры являются интерактивными, то есть для запуска, обработки необходимо выбрать из меню переменные и ответить на ряд вопросов системы.

В электронной таблице установите имена своих переменных. Для этого правой кнопкой мыши щелкните по имени переменной в заголовке таблицы. Откроется контекстное меню, в котором выберите «*Спецификация переменной*». В открывшемся окне введите имя своей переменой в поле «*Имя*» → «*ОК*». Например, вместо *Var*3 поставьте *Y*, *Var*4 замените на *X*1, *Var*5 – на *X*2 и так далее.

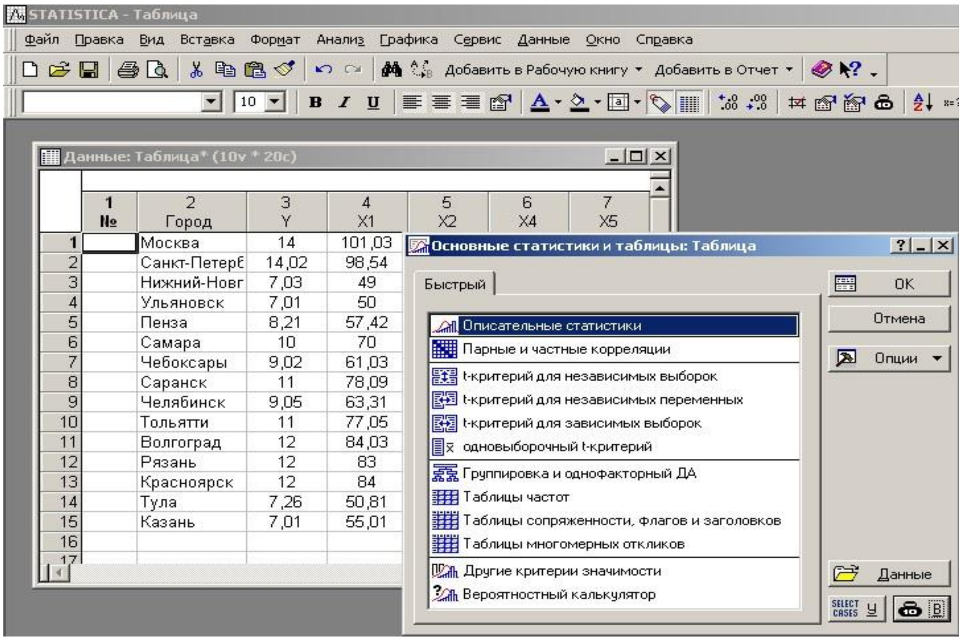


Рис. 6.1. Стартовое окно программы

Корреляционный анализ

1. Построение диаграмм рассеяния

В главном меню программы выберите «*Анализ*» → « *Основные статистики и таблицы*»→ «*Парные* и *частные корреляции*» (рис. 6.1).

В окне «*Парные* и *частные корреляции*» на вкладке «*Частные корреляции*» (рис. 6.2) выберите «*2М рассеяния*» → «*ОК*». В появившемся окне для ввода переменных слева выберите значения исследуемых факторов ($X1$, $X2$, $X4$, $X5$), в окне справа укажите Y и нажмите «*ОК*». В результате получите диаграммы рассеяния для каждой переменной X_i относительно Y.

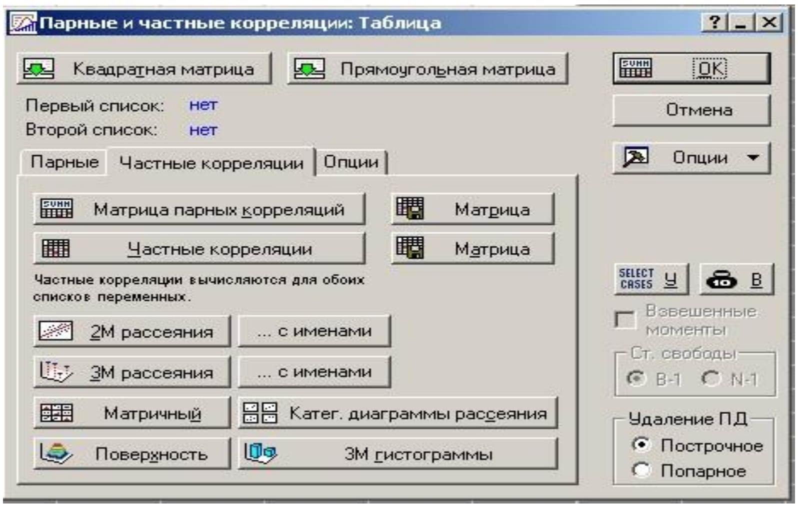


Рис. 6.2. Диалоговое окно «Парные и частные корреляции»

Полученные для каждой пары $Y = f(X_i)$ графики, уравнения регрессии и значения парных коэффициентов корреляции, выпишите в тетрадь и проанализируйте результаты.

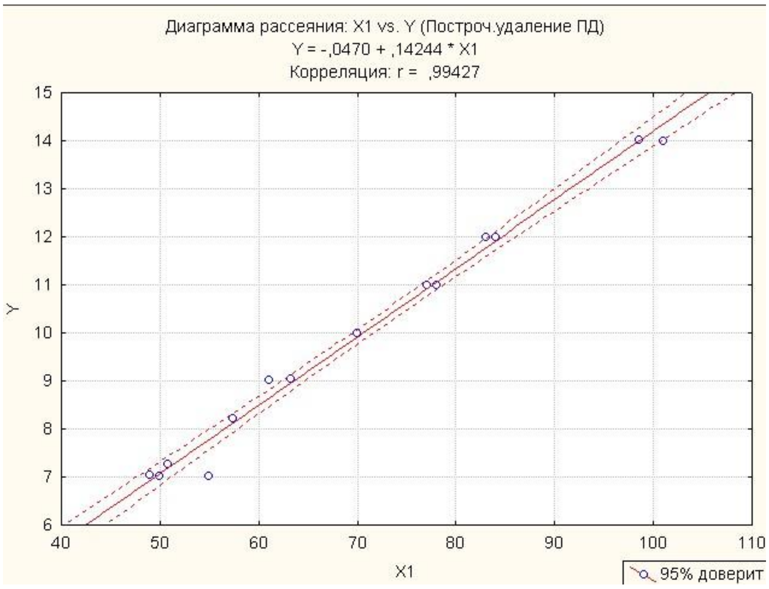


Рис. 6.3. Диаграмма зависимости $Y = f\,(X1)$

2. Получение матрицы парных корреляций

Выберите в главном меню «*Анализ*» → «*Основные статистики и таблицы*» → «*Парные и частные корреляции*» → на вкладке «*Частные корреляции*» откройте список «*Прямоугольная матрица*». В появившемся окне в первом и во втором списках выберите все переменные своего варианта – *Y*, *X*1, *X*2, *X*4, *X*5 → «ОК». Далее выберите опцию «*Матрица парных корреляций*» → «ОК» и получите таблицу парных корреляций

(рис. 6.4).

Отмеченные на экране красным цветом коэффициенты корреляций значимы на уровне $p < 0{,}05$.

Корреляции (Таблица)
Отмеченные корреляции значимы на уровне p <,05
N=15 (Построчное удаление ПД)

Переменная	Y	X1	X2	X4	X5
Y	1,00	0,99	0,96	0,68	0,99
X1	0,99	1,00	0,95	0,71	0,99
X2	0,96	0,95	1,00	0,68	0,95
X4	0,68	0,71	0,68	1,00	0,72
X5	0,99	0,99	0,95	0,72	1,00

Рис. 6.4. Матрица парных корреляций

Проанализируйте полученные результаты (рис. 6.4), сравните с данными полученными на диаграммах рассеяния и сделайте вывод.

3. Вычисление частных коэффициентов корреляции

Проведите самостоятельные исследования при вычислении частных коэффициентов корреляции, выберите в окне «*Парные и частные корреляции*» → «Частные корреляции». Открывается окно для выбора двух списков переменных.

	Частные корреляции (Таблица) Отмеченные корреляции значи N=15 (Построчное удаление П				
Переменная	Y	X1	X2	X4	X5
Y	1,00			-0,60	0,00
X1		1,00			
X2			1,00		
X4	-0,60			1,00	-0,00
X5	0,00			-0,00	1,00

$Y = f(X4\ )$

	Частные корреляции (Табл Отмеченные корреляции зı N=15 (Построчное удалени				
Переменная	Y	X1	X2	X4	X5
Y	1,00				0,64
X1		1,00			
X2			1,00		
X4				1,00	
X5	0,64				1,00

$Y = f(X5)$

Рис. 6.5. Результаты частных корреляций

В первый список введите все анализируемые переменные, а во втором укажите те, которые хотите исключить из рассмотрения. Перебирая разные варианты исключения, сопоставьте результаты.

Сделайте вывод по полученным результатам.

Частные корреляции значимы на уровне $p < 0{,}05$. Красным цветом на экране выделены значимые значения коэффициентов частных корреляций (рис. 6.5).

4. Построение матричной диаграммы рассеяния

Выберите в главном меню «Анализ» → «*Основные статистики и таблицы*» → «*Парные и частные корреляции*» → на вкладке «*Частные корреляции»* выберите «*Матричный*» → «*ОК*». В появившемся окне «*Переменные для матричного графика*» в обоих списках выберите все переменные (*Y*, *X1*, *X2*, *X4*, *X5)* → «ОК». В результате получите матричную диаграмму рассеяния. Проанализируйте и сделайте вывод.

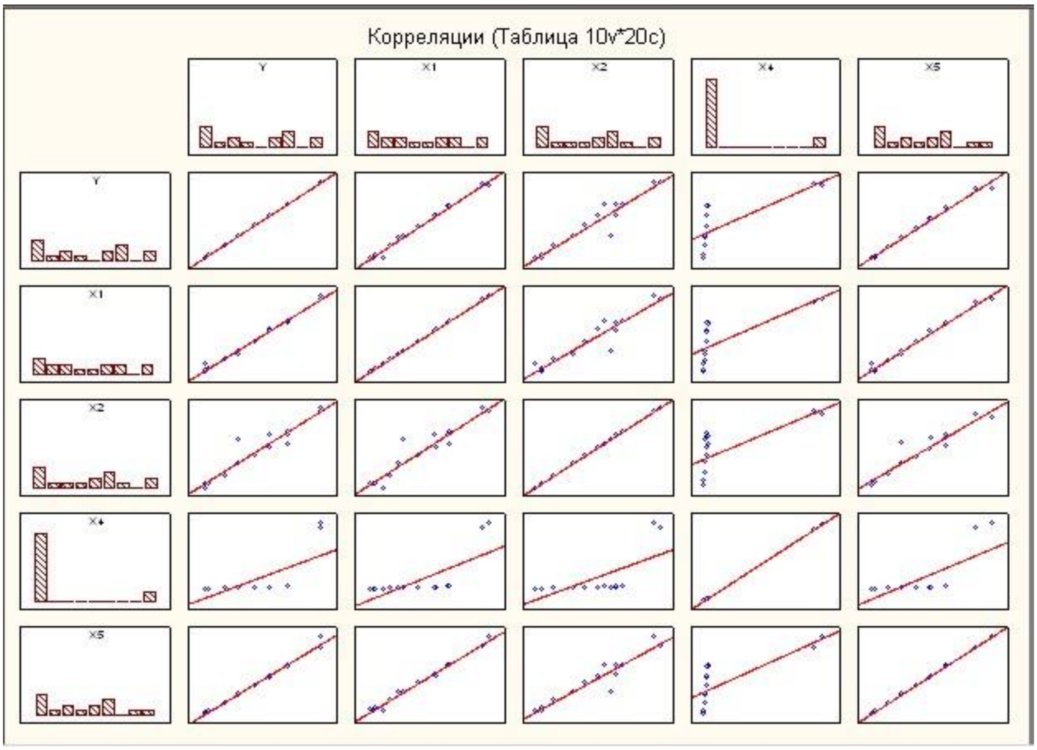


Рис. 6.6. Матрица диаграмм рассеяния

Интерпретация приведенного рисунка (рис. 6.6) такова: чем ближе к теоретической линии регрессии сгруппированы точки, тем теснее связь между изучаемыми показателями.

5. Расчет коэффициентов ранговой корреляции

В меню «*Анализ*» → «*Непараметрическая статистика*» → «*Корреляции Спирмена, тау Кендала, гамма*» → «*ОК*» (рис. 6.7).

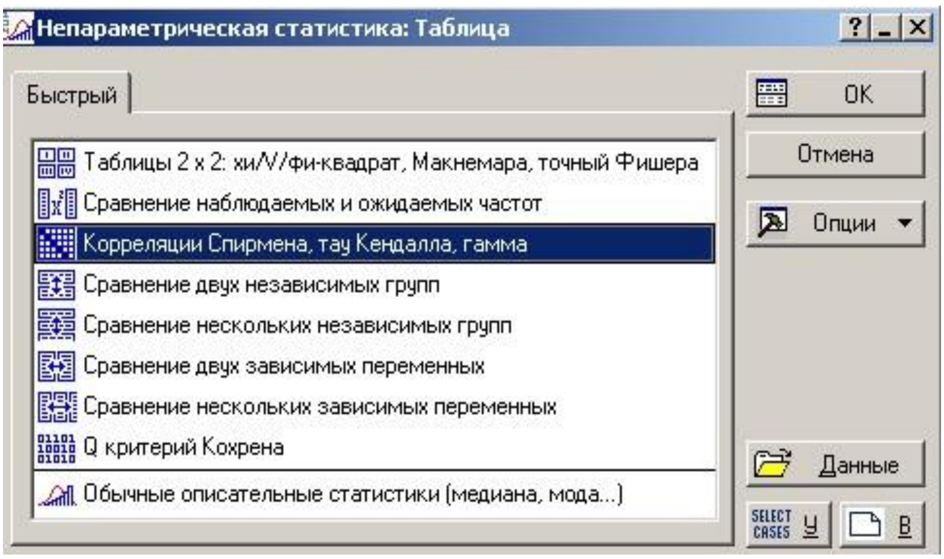


Рис. 6.7. Непараметрическая статистика

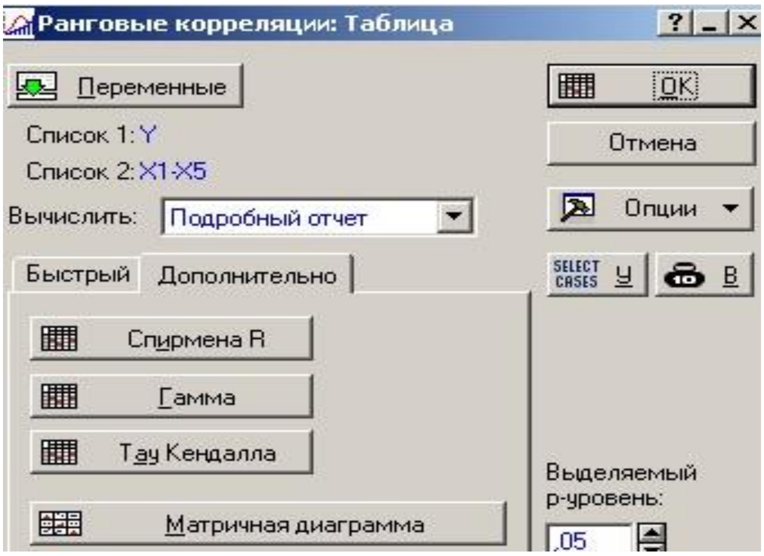


Рис. 6.8. Окно настройки параметров ранговой корреляции

В окне настройки параметров ранговой корреляции (рис. 6.8) в списке «*Вычислить*» установите «*Подробный отчет*». В первый список переменных введите Y, а во второй – все X.

Далее в опции «*Дополнительно*» выберите название вычисляемого коэффициента → «*Спирмена*».

Ранговые корреляции Спирмена (Таблица)
ПД попарно удалены
Отмеченные корреляции значимы на уровне p <,0500

Пара перем.	Число набл.	Спирмена R	t(N-2)	p-уров.
Y & X1	15	0,971289	14,72038	0,000000
Y & X2	15	0,945898	10,51108	0,000000
Y & X4	15	0,977548	16,72705	0,000000
Y & X5	15	0,992812	29,90935	0,000000

Рис. 6.9. Результаты отчета

Результаты отчета показывают значения коэффициентов корреляции, а также проверку из значимости по t-критерию Стьюдента на уровне $p < 0,05$.

Если в опции «*Вычислить*» установите «*Квадратная матрица*», то все переменные выбираются в одном списке. В результате получите таблицу парных коэффициентов корреляции.

Перем.	Ранговые корреляции Спирмена (Таблица) ПД попарно удалены Отмеченные корреляции значимы на уровне p <,050 Y	X1	X2	X4	X5
Y	1,000000	0,971289	0,945898	0,977548	0,992812
X1	0,971289	1,000000	0,904045	0,974084	0,974957
X2	0,945898	0,904045	1,000000	0,914728	0,933516
X4	0,977548	0,974084	0,914728	1,000000	0,977619
X5	0,992812	0,974957	0,933516	0,977619	1,000000

Рис. 6.10. Ранговые коэффициенты корреляции Спирмена

Аналогичным образом получите ранговые коэффициенты корреляции Кендалла, по которым также наблюдается прямая сильная корреляционная взаимосвязь между *Y* и *X* (рис. 6.11).

Пара перем.	Тау корреляции Кендалла (Таблица) ПД попарно удалены Отмеченные корреляции значимы на ур **Число набл.**	Кендалла тау	Z	p-уров.
Y & X1	15	0,897828	4,665251	0,000003
Y & X2	15	0,875633	4,549921	0,000005
Y & X4	15	0,902134	4,687627	0,000003
Y & X5	15	0,965623	5,017523	0,000001

Рис. 6.11. Ранговые коэффициенты корреляции Кендалла

Регрессионный анализ

Выберите в главном меню пункт «*Анализ*» → «*Множественная регрессия*». В этом окне введите значения переменных, выбранные для исследования (рис. 6.12): зависимые переменные – *Y*, независимые переменные *X1*, *X*2 и остальные переменные, далее «*OK*».

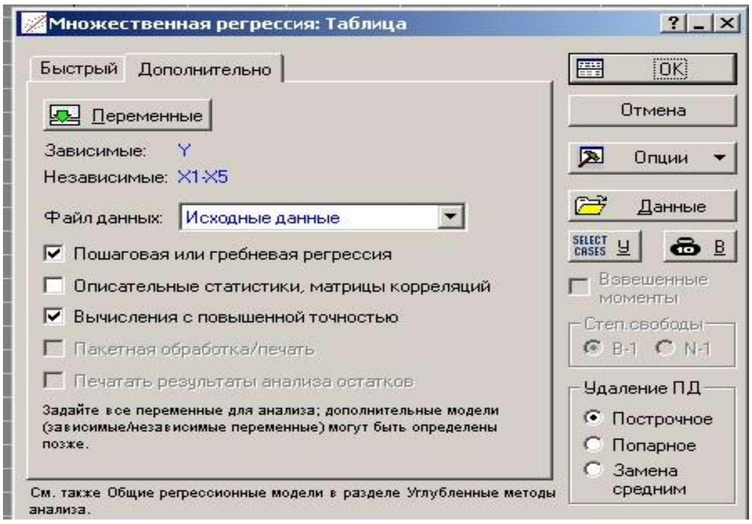


Рис. 6.12. Диалоговое окно «*Множественная регрессия*»

Открывается следующее окно (рис. 6.13), в котором выберите модель и укажите процедуру поиска: стандартную, пошаговую вперед или пошаговую назад. Выберите «Стандартную» процедуру, затем «*OK*» и получите первый результат множественной регрессии (рис. 6.14).

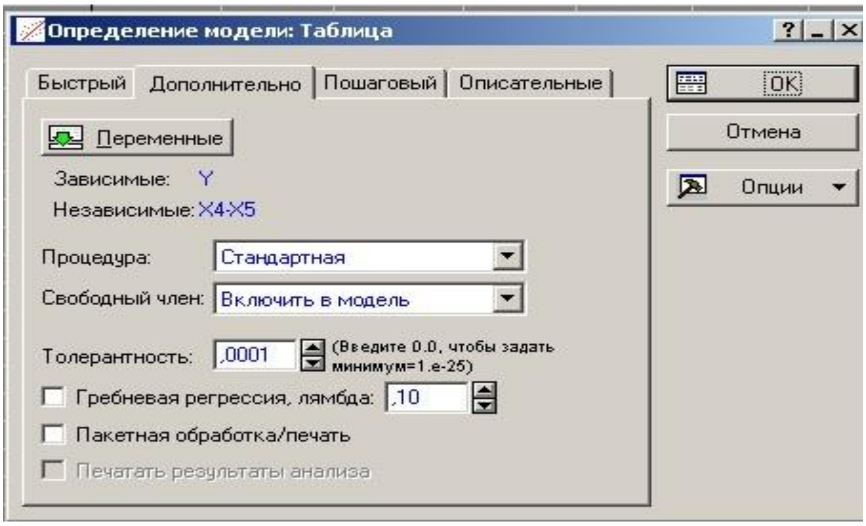


Рис. 6.13. Окно выбора модели

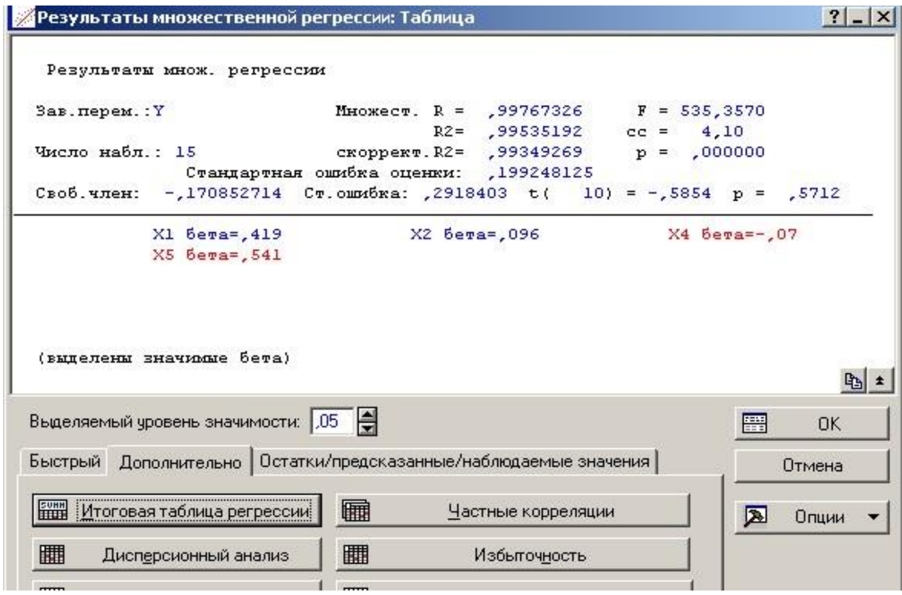


Рис. 6.14. Результаты множественной регрессии

В верхней части окна (рис. 6.14) представлены все итоговые статистики, а также коэффициенты уравнения регрессии.

Далее установите «*Итоговая таблица регрессии*», где будут представлены две таблицы: оцененные параметры модели (рис. 6.15) и основные показатели адекватности множественного уравнения регрессии – итоговые статистики (табл. 6.4).

Таблица 6.4

Итоговые статистики

Статистика	Значение
Множественный коэффициент корреляции, *R*	0,997
Множественный коэффициент детерминации, R^2	0,995
Скорректированный коэффициент детерминации, *R*2	0,993
Критерий Фишера расчетный – *F*(4,10)	535,36
P	0,000
Стандартная ошибка оценки	0,1992

Каждый показатель данной таблицы проанализируйте и сделайте соответствующий вывод.

Итоги регрессии для зависимой переменной: Y (Таблица)
R= ,99767326 R2= ,99535192 Скорректир. R2= ,99349269
F(4,10)=535,36 p<,00000 Станд. ошибка оценки: ,19925

N=15	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(10)	p-уров.
Св.член			-0,170853	0,291840	-0,58543	0,571231
X1	0,418578	0,192558	0,059965	0,027586	2,17378	0,054819
X2	0,096084	0,071237	0,035942	0,026648	1,34879	0,207154
X4	-0,074558	0,031585	-0,000036	0,000015	-2,36055	0,039916
X5	0,540838	0,205794	3,857704	1,467892	2,62806	0,025246

Рис. 6.15. Результаты оценивания уравнения регрессии

Рассмотрите результаты оценки параметров уравнения регрессии по столбцам (рис. 6.15). В первом столбце перечислены члены регрессионного уравнения, в том числе и свободный член уравнения.

Во втором столбце содержатся β-коэффициенты, которые являются отвлеченными (абстрактными) величинами и указывают на сколько среднеквадратических отклонений увеличится зависимая переменная при изменении соответствующего независимой переменной на одно среднеквадратическое отклонение. На практике данный показатель используется для выявления фактора оказывающего наибольшее влияние на зависимую переменную. В рассматриваемом примере, наибольшее (положительное) влияние оказывает показатель *X*5 ($\beta_5 = 0,54$).

В четвертом столбце содержатся значения параметров a_j оцененного уравнения, т.е. в данном случае получили следующую регрессионную модель:

$$Y(X) = -0{,}17 + 0{,}059 \cdot X_1 + 0{,}036 \cdot X_2 - 0{,}000036 \cdot X_4 + 3{,}857 \cdot X_5.$$

В пятом столбце указаны стандартные ошибки коэффициентов уравнения. Стандартные ошибки показывают статистическую надежность коэффициента. Если стандартные ошибки имеют нормальное распределение, то примерно в 2 случаях из 3 истинный коэффициент регрессии находится в пределах одной стандартной ошибки соответствующего коэффициента, и примерно в 95 случаях из 100 в пределах двух стандартных ошибок. Значение стандартных ошибок используют для построения доверительных интервалов.

Шестой столбец выводит расчетное значение *t*-статистики Стьюдента. Ее значение используется для проверки значимости соответствующего коэффициента.

Анализируемый коэффициент a_0, a_1, $a_2, \ldots, a_m$ считается значимым, если рассчитанное системой *Statistica* для него значение *t*-критерия по абсолютной величине превышает $t_{табл}$, определяемым с использованием специальных таблиц по заданным уровню значимости (например, $\alpha = 0{,}05$) и числу степеней свободы ($cc = n - m - 1$).

Седьмой столбец (*p*-уровень) – показывает вероятность принять или отвергнуть гипотезу о равенстве нулю соответствующего коэффициента. Значения вероятности, указанные в таблице известны в статистике как уровни значимости α. Коэффициент регрессии a_0, a_1, $a_2, \ldots, a_m$ признается значимым, если рассчитанное системой *Statistica* для него значение уровня значимости *p* меньше (или равно) 0,05 (для 95 %-ной доверительной вероятности). Если значение вероятности ниже уровня значимости α, то гипотеза H_0 отвергается и соответствующий коэффициент не равен нулю.

В рассматриваемом примере параметры a_4 и a_5 при переменных X_4 и X_5 значимы при уровне значимости α меньше, чем 0,05. Остальные коэффициенты получены не значимыми при уровне α = 0,05, так как значение вероятности для них больше 0,05.

Проверить статистическую значимость параметров можно так же используя табличное значение *t*-критерия Стьюдента. В рассматриваемом примере для $\alpha = 0{,}05$ и $\nu = cc = 15 - 4 - 1 = 10$, $t_{10,\,0{,}05} = 2{,}23$. Сравните расчетное значение *t*-критерия Стьюдента (рис. 6.15) с табличным значением для:

a0 – $t_{рас}$. = |–0,599| < 2,23 → параметр статистически не значим; $a4$ – $t_{рас}$. = |–2,360| > 2,23 → параметр статистически значим.

Так как оцененная множественная регрессионная модель незначима по трем параметрам, то необходимо исключить их из рассмотрения.

Для получения новой модели повторите вышеописанную процедуру. При этом в модель не включайте незначимый свободный член. В список переменных введите только значимые параметры.

Выберите в меню «Анализ» → «Множественная регрессия» → «Переменные» → в появившемся окне «*Список переменных*» слева выберите Y, справа выберите *X*4, *X*5 → «ОК». Получите новый результат.

N=15	Итоги регрессии для зависимой переменной: Y (Таблица) R= ,99979057 R2= ,99958118 Скорректир. R2= ,99951674 F(2,13)=15513, p<0,0000 Станд. ошибка оценки: ,22687					
	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(13)	p-уров.
X4	-0,023811	0,007322	-0,000044	0,000013	-3,2518	0,006306
X5	1,014663	0,007322	7,574851	0,054664	138,5717	0,000000

Рис. 6.16. Результаты оценивания уравнения регрессии

Оценив вторую модель, можно утверждать, что она пригодна для практического использования, так как параметры модели статистически значимы по t-критерию Стьюдента, а уравнение в целом проходит тест по F-критерию Фишера. Запишите новое уравнение регрессии:

$$Y(X) = -\,0{,}000044\,X4 + 7{,}574\,X5.$$

В зависимости от целей исследования можете выбирать разные методы поиска модели. Можете выбрать не стандартную процедуру, а процедуру пошагового оценивания, указав – отображать результаты на каждом шаге или только итоги.

Для выполнения прогнозов по полученному уравнению необходимо показать, что регрессионная модель адекватна результатам наблюдений. С этой целью воспользуйтесь критерием Дарбина-Уотсона, согласно которого, рассчитанный системой *Statistica* коэффициент $d_{расч}$ необходимо сравнить с табличным значением $d_{табл}$ (для совокупности объемом n = 15, уровня значимости □ = 0,05 и двух оцениваемых параметров регрессии, значение *dтабл* =1,54). Если $d_{расч} > d_{табл}$, то полученная модель адекватна и пригодна для прогнозирования.

Для определения критерия Дарбина-Уотсона ($d_{расч}$) в окне «*Результаты множественной регрессии*» на вкладке «*Остатки*» выберите опцию «*Анализ остатков*» → «*Дополнительно*» →
«*Статистика Дарбина-Уотсона*» (рис. 6.17).

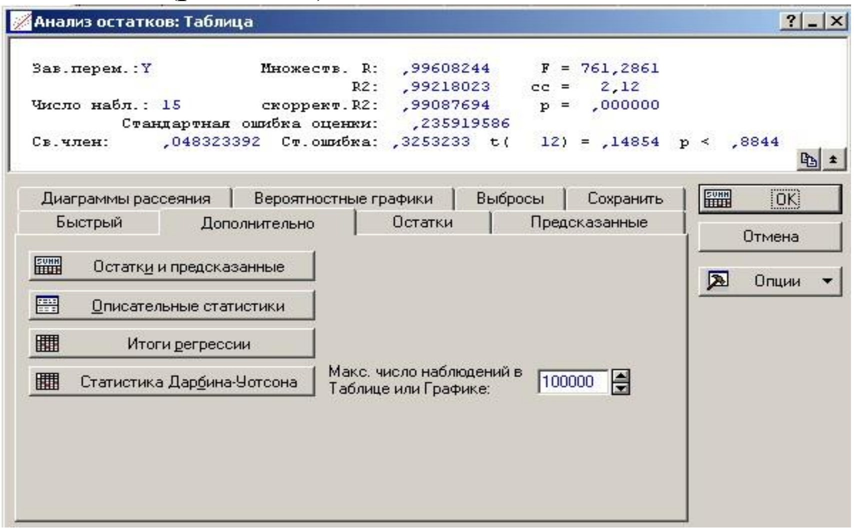


Рис. 6.17. Анализ остатков по критерию Дарбина-Уотсона

	Дарбина-Уотсона d (Таблица) и сериальная корреляция остатков	
	Дарбина-Уотсон.d	Сериал. Корр.
Оценка	2,737621	-0,584588

Рис. 6.18. Статистика Дарбина-Уотсона

В рассматриваемом примере $d_{расч} = 2,74 > 1,54$, следовательно, модель можно использовать для прогнозирования.

В случае, когда модель адекватна результатам наблюдения для выполнения прогноза в окне «*Результаты множественной регрессии»* на вкладке «О*статки/ предсказанные / наблюдаемые значения»* выберите опцию «*Предсказать зависимую переменную*».

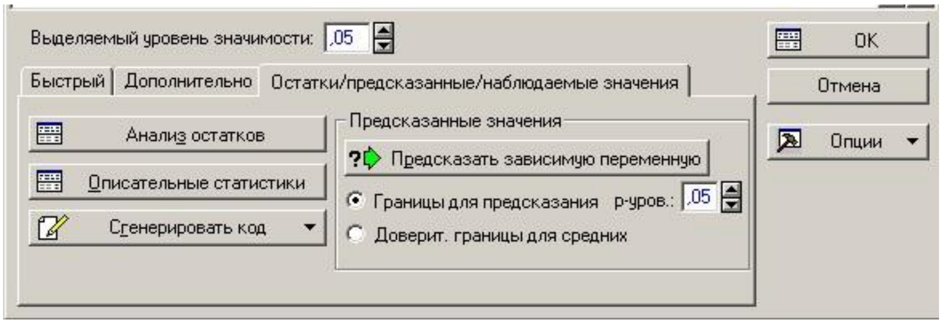


Рис. 6.19. Окно установки прогноза (показана нижняя часть окна)

Для того, чтобы найти прогнозное значение зависимой переменной в пространственной модели задайте (введите в соответствующие поля окна) значение независимой переменной в соответствии с содержанием задачи на условия прогноза.

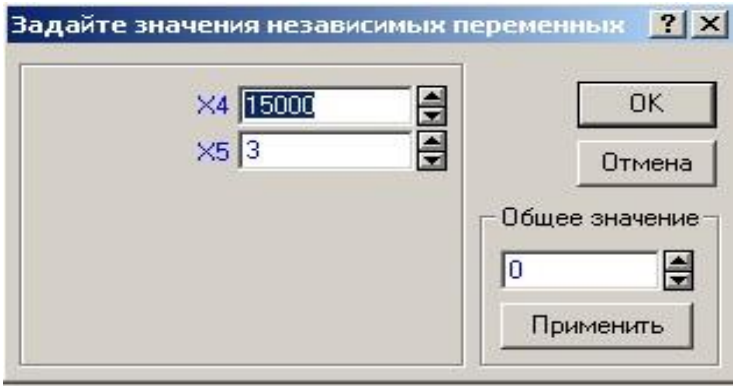


Рис. 6.20. Установка прогнозируемых значений переменной

Для выполнения прогноза: «Как изменится производительность труда (Y) на московском предприятии, если среднегодовую численность рабочих ($X4$) сократить на 780 человек, а коэффициент сменности оборудования ($X5$) повысить до 3?» необходимо в окне (рис. 6.20) задать соответствующие значения в $X4$ и $X5$. Результаты полученного прогноза представлены на рис. 6.21.

	Предск.значения для (Таблица) перемен.: Y		
Переменная	**B-Веса**	Значение	B-Веса * знач.
X4	-0,000042	15000,00	-0,62817
X5	7,536567	3,00	22,60970
Св.член			0,04832
Предсказ.			22,02986
-95,0%ДП			21,32611
+95,0%ДП			22,73360

Рис. 6.21. Прогнозные значения

В первом столбце (рис. 6.21) содержатся наименования расчетных и исходных показателей. Во втором столбце приведено значение параметров a_4 и a_5. В третьем – значение независимых переменных используемых для расчета прогноза. В четвертом – значение независимой переменной (с доверительным интервалом) рассчитанное в результате оценивания прогноза.

По рассмотренному алгоритму можно выполнять прогнозирование для каждого из значимых факторов в разных вариантах.

Порядок выполнения работы

1. Ознакомьтесь с методикой выполнения корреляционнорегрессионного анализа в пакете *Statistica*.
2. У преподавателя получите вариант индивидуального задания.
3. Определите парные, частные и множественные коэффициенты корреляций и на их основе выявите факторы, оказывающие наибольшее влияние на результативный показатель.
4. Получите регрессионную модель и проведите ее полный статистический анализ.
5. Выполните прогнозирование по заданию своего варианта.
6. Сделайте выводы по работе и оформить отчет.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) результаты вычислений индивидуальных заданий;
3) анализ полученных результатов в пакете *Statistica*.

Варианты индивидуальных заданий по регрессионному анализу

Рассматриваются следующие показатели для 50 предприятий:

Y_1 – производительность труда;

Y_2 – индекс снижения себестоимости продукции;

Y_3 – рентабельность;

X_4 – трудоемкость единицы продукции;

X_5 – удельный вес рабочих;

$X6$ – удельный вес покупных изделий;

$X7$ – коэффициент сменности оборудования;

$X8$ – премии и вознаграждения на одного работника;

$X9$ – удельный вес потерь от брака;

$X10$ – фондоотдача;

$X11$ – среднегодовая численность работников;

$X12$ – среднегодовая стоимость основных производственных фондов;

$X13$ – среднегодовой фонд заработной платы работников;

$X14$ – фондовооруженность труда;

$X15$ – непроизводственные расходы.

Таблица 6.5 Варианты заданий 1–10

№ варианта	Результативный признак Y_j	Номер факторных признаков X_i
1	1	6, 8, 11, 12, 15
2	1	8, 11, 12, 13, 15
3	1	8, 9, 13, 14, 15
4	3	8, 9, 10, 11, 15
5	3	8, 9, 10, 12, 15
6	2	4, 5, 6, 8, 9
7	2	4, 5, 6, 7, 9
8	2	4, 5, 6, 8, 9
9	2	4, 5, 8, 9, 15
10	2	4, 5, 7, 9, 15

Таблица 6.6

Таблица исходных данных

№ предприятия	Y_1	Y_2	Y_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	9,26	204,2	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45
2	9,38	209,6	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30
3	12,11	222,6	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37
4	10,81	236,7	12,85	0,17	0,70	0,50	1,42	0,43	0,18	1,65
5	9,35	62,0	10,63	0,23	0,62	0,40	1,35	0,88	0,15	1,91
6	9,87	53,1	9,12	0,43	0,76	0,19	1,39	0,57	0,34	1,68
7	8,17	172,1	25,83	0,31	0,73	0,25	1,16	1,72	0,38	1,94
8	9,12	56,5	23,39	0,26	0,71	0,44	1,27	1,70	0,09	1,89

9	5,88	52,6	14,68	0,49	0,69	0,17	1,16	0,84	0,14	1,94
10	6,30	46,6	10,05	0,36	0,73	0,39	1,25	0,60	0,21	2,06
11	6,22	53,2	13,99	0,37	0,68	0,33	1,13	0,82	0,42	1,96
12	5,49	30,1	9,68	0,43	0,74	0,25	1,10	0,84	0,05	1,02
13	6,50	146,4	10,03	0,35	0,66	0,32	1,15	0,67	0,29	1,85
14	6,61	18,1	9,13	0,38	0,72	0,02	1,23	1,04	0,48	0,88
15	4,32	13,6	5,37	0,42	0,68	0,06	1,39	0,66	0,41	0,62
16	7,37	89,8	9,86	0,30	0,77	0,15	1,38	0,86	0,62	1,09
17	7,02	62,5	12,62	0,32	0,78	0,08	1,35	0,79	0,56	1,60
18	8,25	46,3	5,02	0,25	0,78	0,20	1,42	0,34	1,76	1,53
19	8,15	103,5	21,18	0,31	0,81	0,20	1,37	1,60	1,31	1,40
20	8,72	73,3	25,17	0,26	0,79	0,30	1,41	1,46	0,45	2,22
21	6,64	76,6	19,40	0,37	0,77	0,24	1,35	1,27	0,50	1,32
22	8,10	73,01	21,0	0,29	0,78	0,10	1,48	1,58	0,77	1,48
23	5,52	32,3	6,57	0,34	0,72	0,11	1,24	0,68	1,20	0,68
24	9,37	199,6	14,19	0,23	0,79	0,47	1,40	0,86	0,21	2,30
25	13,17	598,1	15,81	0,17	0,77	0,53	1,45	1,98	0,25	1,37
26	6,67	71,2	5,23	0,29	0,80	0,34	1,40	0,33	0,15	1,51
27	5,68	90,8	7,99	0,41	0,71	0,20	1,28	0,45	0,66	1,43

Продолжение табл. 6.6

№ предприятия	Y_1	Y_2	Y_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
28	5,22	82,1	17,50	0,41	0,79	0,24	1,33	0,74	0,74	1,82
29	10,02	76,2	17,16	0,22	0,76	0,54	1,22	0,03	0,32	2,62
30	8,16	119,5	14,54	0,29	0,78	0,40	1,28	0,99	0,89	1,75
31	3,78	21,9	6,24	0,51	0,62	0,20	1,47	0,24	0,23	1,54
32	6,48	48,4	12,68	0,36	0,75	0,64	1,27	0,57	0,32	2,25
33	10,44	173,5	19,49	0,23	0,71	0,42	1,51	1,22	0,54	1,07
34	7,65	74,1	9,28	0,26	0,74	0,27	1,46	0,68	0,75	1,44
35	8,77	68,6	11,42	0,27	0,65	0,37	1,27	1,0	0,16	1,40
36	7,00	60,8	10,31	0,29	0,66	0,38	1,43	0,81	0,24	1,31
37	11,06	355,6	8,65	0,01	0,84	0,35	1,50	1,27	0,59	1,12
38	9,02	264,8	10,94	0,02	0,74	0,42	1,35	1,14	0,56	1,16
39	13,28	526,6	9,87	0,18	0,75	0,32	1,41	1,89	0,63	0,88
40	9,27	118,6	6,14	0,25	0,75	0,33	1,47	0,67	1,10	1,07
41	6,70	37,1	12,93	0,31	0,79	0,29	1,35	0,96	0,39	1,24
42	6,69	57,7	9,78	0,38	0,72	0,30	1,40	0,67	0,73	1,49
43	9,42	51,6	13,22	0,24	0,70	0,56	1,2Ь	0,98	0,28	2,03
44	7,24	64,7	17,29	0,31	0,66	0,42	1,15	1,16	0,10	1,84
45	5,39	48,3	7,11	0,42	0,69	0,26	1,09	0,54	0,68	1,22
46	5,61	15,0	22,49	0,51	0,71	0,16	1,26	1,23	0,87	1,72
47	5,59	87,5	12,14	0,31	0,73	0,45	1,36	0,78	0,49	1,75
48	6,57	108,4	15,25	0,37	0,65	0.31	1,15	1,16	0,16	1,46
49	6,54	267,3	31,34	0,16	0,82	0,08	1,87	4,44	0,85	1,60
50	4,23	34,2	11,56	0,18	0,80	0,68	1,17	1,06	0,13	1,47

Продолжение табл. 6.6

№ предприятия	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
1	26006	167,69	47750	6,40	17,72
2	23935	186,10	50391	7,80	18,39
3	22589	220,45	43149	9,76	26,46
4	21220	169,30	41089	7,90	22,37
5	7394	39,53	14257	5,35	28,13
6	11586	40,41	22661	9,90	17,55
7	26609	102,96	52509	4,50	21,92
8	7801	37,02	14903	4,88	19,52
9	11587	45,74	25587	3,46	23,99
10	9475	40,07	1661	3,60	21,76
11	10811	45,44	19459	3 56	25,68
12	6371	41,08	12973	5,65	18,13
13	26761	136,14	50907	4,28	25,74
14	4210	42,39	6920	8,85	21,21
15	3557	37,39	5736	8,52	22,97
16	14148	101,78	26705	7,19	16,38
17	9872	47,55	20068	4,82	13,21
18	5975	32,61	11487	5,46	14,48
19	16662	103,25	32029	6,20	13,38

Окончание табл. 6.6

№ предприятия	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
20	9166	38,95	18946	4,25	13,69
21	15118	81,32	28025	5,38	16,66
22	11429	67,26	20968	5,88	15,06
23	6462	59,92	11049	9,27	20,09
24	24628	107,34	45893	4,36	15,98
25	49727	512 60	99400	10,31	18,27
26	11470	53,8,1	20719	4,69	14,42
27	19448	80,83	36813	4,16	22,76
28	18963	59,42	33956	3,13	15,41
29	9185	36,96	17016	4,02	19,35
30	17478	91,43	34873	5,23	16,83
31	6265	17,16	11237	2,74	30,53
32	8810	27,29	17306	3,10	17,98
33	17659	184,33	39250	10,44	22,09
34	10342	58,42	19074	5,65	18,29
35	8901	59,40	18452	6,67	26,05
36	8402	49,63	17500	5,91	26,20
37	32625	391,27	7888	11,99	17,26
38	31160	258,62	58947	8,30	18,83
39	46461	75,66	94697	1,63	19,70

40	13833	123,68	29626	8,94	16,87
41	6391	37,21	11688	5,82	14,63
42	11115	53,37	21955	4,80	22,17
43	6555	32,87	12243	5,01	22,62
44	11085	45,63	20193	4,12	26,44
45	9484	48,41	20122	5,10	22,26
46	3967	13,58	7612	3,49	19.13
47	15283	63,99	27404	4,19	18,28
48	20874	104,55	39648	5,01	28,23
49	19418	222,11	43799	11,44	12,39
50	3351	25,76	6235	7,67	11,64

Контрольные вопросы

1. Перечислите виды статистического анализа данных.
2. Для чего используются коэффициенты парной корреляции?
3. В каких случаях применяются коэффициенты множественной корреляции?
4. Как определить характер связи по значению коэффициента корреляции?
5. Какие задачи решаются методом корреляции?
6. Когда применяются коэффициенты ранговой корреляции?
7. С помощью какого критерии проверяется значимость уравнения регрессии?
8. Как оценивается статистическая значимость параметров уравнения регрессии?
9. Для каких целей используется β-коэффициент?
10. Что показывает коэффициент эластичности?

Практическая работа № 7

Выявление и характеристика основной тенденции развития в рядах динамики (4 часа)

Цель работы

1. Изучить методы укрупнения интервалов, скользящей средней и аналитического выравнивания для выявления тренда в рядах динамики.
2. Изучить методику расчета параметров трена в рядах динамики.
3. Сформировать практические навыки расчета параметров уравнения тренда динамического ряда.
4. Построить уравнение тренда динамического ряда и провести статистический анализ значимости его параметров.

Краткая теория

Ряд динамики – числовые значения статистического показателя, представленные во временной последовательности. Он состоит из двух граф: 1) в первой указывают периоды (даты); 2) во второй указывают показатели, характеризующие данный объект за эти периоды (даты). Показатели второй графы называются уровнями ряда. Первый показатель называется начальным уровнем, последний – конечным.

Уровни ряда могут быть выражены абсолютными, средними или относительными величинами. Ряды динамики относительных и средних величин строятся на основе рядов абсолютных величин.

По времени ряды динамики разделяются на *моментальные* и *интервальные.*

Моментальный ряд динамики – ряд, уровни которого характеризуют состояние явления на определенные моменты времени. Примером такого ряда могут служить данные о численности населения РФ (млн. чел.) по состоянию на 1 января в разные года. *Моментальные ряды нельзя суммировать*, так как в каждом последующем уровне полностью или частично содержится значение предыдущего уровня.

Интервальный ряд динамики – ряд, уровни которого характеризуют размер явления за конкретный период времени. Примером такого ряда могут служить данные о динамике добычи нефти в РФ (млн. тонн) за определенный период. *Интервальные ряды можно суммировать*, так как значение предыдущего уровня не содержится в последующем.

Это позволяет получать статистические данные за более длительный период времени.

Одна из задач статистики – определение в рядах динамики *общей тенденции развития явления.* На развитие явления во времени оказывают влияние факторы, различные по характеру и силе воздействия. Одни из них оказывают постоянное воздействие и формируют в рядах динамики определенную *тенденцию* развития. Воздействие других факторов может носить *случайный* характер. Поэтому при анализе динамики исследоваться должна *основная тенденция*, устойчивая на протяжении изучаемого этапа развития.

Основная тенденция развития (тренд) – плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний.

С целью выявления тренда ряды динамики исследуются *методами укрупнения интервалов, скользящей средней* и *аналитического выравнивания.*

Метод укрупнения интервалов основан на укрупнении периодов времени, к которым относятся уровни ряда динамики. Например, ряд ежесуточного выпуска продукции заменяется рядом месячного выпуска продукции, который, в свою очередь, может быть заменен рядом квартального выпуска продукции. Для каждого укрупненного интервала вычисляются средние уровни ряда $\bar{y}_i$ по формуле простой арифметической средней. Например, если укрупненный интервал образован объединением трех периодов, эти величины равны:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3}; \qquad \bar{y}_2 = \frac{y_4 + y_5 + y_6}{3} \quad \text{и т.д.,}$$

где $y_1, y_2, \ldots, y_6$ – уровни исходного ряда динамики.

Сравнение рассчитанных таким образом средних позволяет выявить направление и характер (ускорение или замедление роста) основной тенденции развития, но не математическую модель тренда.

Метод скользящей средней основан на расчете средней величины, которая вычисляется по ряду при последовательном передвижении на один интервал, т.е. сначала вычисляют средний уровень из определенного числа первых по порядку уровней ряда, затем – средний уровень из такого же числа членов, начиная со второго и т.д. Например, скользящие средние с продолжительностью периода, равной 3, вычисляются как:

y □ y □ y y2□y3 □ y4 y y1 □ 3 2 ; y2 □ 3 ; y3 □ 3 □ y34 □ y5 и т.д.
1 3

Таким образом, средняя как бы «скользит» по ряду динамики. Если в ряду динамики имеются периодические колебания, то период скользящей средней совпадает с периодом колебания или будет кратным ему. Метод позволяет выявить направление и характер основной тенденции развития, но не математическую модель тренда.

Метод аналитического выравнивания – наиболее эффективный способ выявления основной тенденции развития. Он позволяет определить аналитическое выражение, отражающее закономерность изменения явления как функцию времени $\hat{y}_t$ □ f (t), где t – условное обозначение времени. Например,

год 2000 2001 2002 2003 2004 2005 t 1 2 3 4 5 6

Метод аналитического выравнивания основан на замене фактических значений уровней y_i плавно изменяющимися величинами $\hat{y}_t$.

Аналитическое выравнивание может быть осуществлено по любому рациональному многочлену (см. табл. 7.3). Выбор функции производится, во-первых, на основе *анализа характера закономерностей динамики* данного явления.

1. При *равномерном развитии* явления во времени используется полином 1-й степени (прямая): $\hat{y}_t$ □ a_0 □ a_1□t . В этом случае абсолютные приросты □y (первые разности) практически постоянны. При a_1 □ 0 уровни динамики равномерно возрастают, при a_1 □ 0 – равномерно снижаются. Линейный тренд хорошо отражает тенденцию изменений при действии множества факторов, изменяющихся по разным закономерностям, что приводит к взаимопогашению особенностей отдельных факторов. Примером могут служить тенденции динамики урожайности для масштаба области, республики, страны в целом.
2. При *равноускоренном* (*равнозамедленном*) *развитии* явления во времени применяется полином 2–й степени (парабола): $\hat{y}_t$ □ a_0 □

$a1□t$ □ $a2$ □t^2. В этом случае постоянными являются вторые разности: □2y □

□y □ □y . Такой характер развития возникает при i i i □1

наличии важных факторов прогрессивного развития (поступление нового высокопроизводительного оборудования, снятие ограничений в распределении дохода и пр.). Параметр a выражает начальную скорость 1 роста, коэффициент a – постоянную скорость изменения прироста 2

(ускорение). Параболическая форма тренда с отрицательным ускорением (a □ 0) приводит со временем не только к приостановке роста уровня, 2 но и к его снижению со всё большей скоростью. Такой характер развития свойственен производству устаревшей продукции.

3. *Развитие* явления *с переменным ускорением* (*замедлением*) описывается полином 3–й степени: $y\hat{}t$ □ $a0$ □$a1□t$ □$a2$ □t^2 □$a3□t^3$.

В данном случае постоянны третьи разности: □^3y □ □^2y □ □^2y , i i i □1

то есть абсолютные приросты ускоренно возрастают ($^a3$ □ 0) или замедляются ($^a3$ □ 0).

4. *Развитие* явления, характеризующееся *постоянным темпом роста* (*снижения*), описывается показательной функцией: $\hat{y}\,t$ □ $^a0$ □$^a1^t$ (в частном

случае – экспоненциальной функцией: $y\hat{}t$ □ $^a0$ □ea^1t).

—

1 a^1) характеризует среднегодовой коэффициент роста $Кр$. Величина a (e

Если $^a1$ □1, экспоненциальный (показательный) тренд выражает тенденцию ускоренного и все более ускоряющегося возрастания уровней динамического ряда. Такой характер свойственен размножению организмов при отсутствии ограничения со стороны среды: сорняков, вирусных заболеваний. При росте по экспоненте абсолютный прирост пропорционален достигнутому уровню. Однако, такой рост может продолжаться только небольшой исторический период, так как любой процесс развития всегда встретит ограничения.

При $^a1$ □1, экспоненциальный (показательный) тренд означает тенденцию все

более замедляющегося снижения уровней ряда. Такая тенденция присуща динамике трудоёмкости продукции, удельных затрат топлива или металла на единицу полезного эффекта (на 1 $кВт \cdot ч$, на 1 $м^2$ жилой площади и т.д.) при технологическом процессе.

5. При *развитии* явления *с замедлением роста в конце периода* применяется логарифмическая функция: $\hat{y}_t = a_0 + a_1 \cdot \lg(t)$. Логарифмический тренд используется для отображения тенденции замедляющегося роста уровней при отсутствии предельного значения. Замедление роста становится все меньше и меньше, и при достаточно большом значении t логарифмическая кривая становится мало отличимой от прямой линии. Такая тенденция присуща росту спортивных достижений (чем они выше, тем труднее их улучшить), росту производительности аппарата по мере его совершенствования без качественных преобразований.

6. Гиперболическая форма тренда имеет вид: $\hat{y}_t = a_0 + a_1 \frac{1}{t}$. Если $a_1 > 0$, то гиперболический тренд соответствует тенденции замедляющегося снижения уровня, стремящегося к пределу a_0. Если $a_1 < 0$, тренд выражает тенденцию замедляющегося роста уровней, стремящихся в пределе к a_0. Следовательно, данная форма тренда используется для отображения процессов, *ограниченных предельным значением уровня* (предельным коэффициентом полезного действия двигателя, пределом 100 %-ной грамотности населения и т.п.).

Во-вторых, выбор типа модели должен быть основан на *анализе графического изображения* уровней динамического ряда (линейной диаграммы).

Методика измерения параметров тренда

Когда тип тренда установлен, вычисляют оптимальные значения параметров тренда, исходя из фактических уровней. Для этого обычно используют *метод наименьших квадратов* (МНК). В этом методе минимизируется сумма квадратов отклонений фактических уровней динамического ряда y_i от выровненных уровней $\hat{y}_t$ (от тренда).

Для каждого типа тренда МНК дает *систему нормальных уравнений*, решая которую, рассчитывают параметры тренда. Однако вычислительный процесс определения параметров тренда при сохранении полной идентичности конечных результатов может быть упрощен, если ввести обозначения дат (периодов) таким образом, чтобы $\sum t_i = 0$.

Если количество уровней в ряду динамики нечетное, то временные даты (*t*) обозначаются следующим образом (табл. 7.1):

Таблица 7.1

Условное обозначение времени при нечетном числе уровней

Временные даты (периоды)	2009	2010	2011	2012	2013
Уровни ряда динамики, y_i	y_1	y_2	y_3	y_4	y_5
Обозначения временных дат, *t*	–2	–1	0	1	2

Если количество уровней в ряду динамики четное, то счет ведется полугодиями и обозначения временных дат (*t*) принимают следующий вид (табл. 7.2):

Таблица 7.2

Условное обозначение времени при четном числе уровней

Временные даты (периоды)	2009	2010	2011	2012	2013	2014
Уровни ряда динамики, y_i	y_1	y_2	y_3	y_4	y_5	y_6
Обозначения временных дат, *t*	-5	-3	-1	1	3	5

В таблице 7.3 приводятся различные виды трендовых моделей и системы нормальных уравнений для определения параметров тренда.

Прогнозирование развития социально-экономических явлений на основе тренда

Аналитическое выравнивание позволяет не только определить общую тенденцию изменения явления на рассматриваемом отрезке времени, но и выполнить расчеты для периодов, для которых нет исходных данных.

Интерполяции – определение недостающих значений признака внутри рассматриваемого периода**.**

Экстраполяция – определение недостающих значений признака за пределами рассматриваемого периода.

Применение экстраполяции для прогнозирования основывается на предположении, что найденная закономерность развития внутри динамического ряда сохраняется и вне этого ряда. Это справедливо, если исследуемое явление развивается в достаточно стабильных условиях. Так как анализируемые ряды динамики обычно относительно короткие, то и период экстраполяции не может быть бесконечным. Поэтому срок прогноза – *период упреждения* (период от конца базы расчета до прогнозируемого периода) не должен превышать $1/3$ длительности базы расчета тренда. На основе динамических рядов получают надежные прогнозы, если уровни ряда сопоставимы и получены по единой методологии.

Таблица 7.3

Виды трендовых моделей

Наименование функции	Вид функции	Система нормальных уравнений для определения параметров тренда
Линейная	$\hat{y}_t = a_0 + a_1 \cdot t$	$a_0 \cdot n = \sum y_i$ $a_1 \cdot \sum t_i^2 = \sum y_i \cdot t_i$
Полином 2-й степени (парабола)	$\hat{y}_t = a_0 + a_1 \cdot t + a_2 \cdot t^2$	$a_0 \cdot n + a_2 \cdot \sum t_i^2 = \sum y_i$ $a_1 \cdot \sum t_i^2 = \sum y_i \cdot t_i$ $a_0 \cdot \sum t_i^2 + a_3 \cdot \sum t_i^4 = \sum y_i \cdot t_i^2$
Полином 3-й степени	$\hat{y}_t = a_0 + a_1 \cdot t + a_2 \cdot t^2 + a_3 \cdot t^3$	$a_0 \cdot n + a_2 \cdot \sum t_i^2 = \sum y_i$ $a_1 \cdot \sum t_i^2 + a_3 \cdot \sum t_i^4 = \sum y_i \cdot t_i$ $a \cdot \sum t^2 + a \cdot \sum t^4 = \sum y \cdot t^2$

		$a_0 \sum t_i^3 + a_1 \sum t_i^4 + a_2 \sum t_i^5 + a_3 \sum t_i^6 = \sum y_i t_i^3$
Показательная	$\hat{y}_t = a_0 \cdot a_1^t$	$n \cdot \lg a_0 + \lg a_1 \sum t_i = \sum \lg y_i$ $\lg a_0 \sum t_i + \lg a_1 \sum t_i^2 = \sum t_i \lg y_i$
Гиперболическая	$\hat{y}_t = a_0 + a_1 \frac{1}{t}$	$a_0 n + a_1 \sum \frac{1}{t_i} = \sum y_i$ $a_0 \sum \frac{1}{t_i} + a_1 \sum \frac{1}{t_i^2} = \sum \frac{y_i}{t_i}$

В отличие от прогноза на основе регрессионного уравнения прогноз по тренду *учитывает факторы* развития только *в неявном виде*, что не позволяет «проигрывать» разные варианты прогнозов при разных возможных значениях факторов, влияющих на изучаемый признак. Однако прогноз по тренду *охватывает все факторы*, в то время как в регрессионную модель невозможно включить в явном виде более 10–20 факторов.

При составлении прогнозов уровней социально-экономических явлений рассчитывают *доверительные интервалы* прогноза, используя *интервальную* оценку. Границы интервалов определяют по формуле:

$$\hat{y}_t \pm t_{\alpha} \cdot S_{\hat{y}}, \qquad (7.1)$$

где $\hat{y}_t$– точечный прогноз, рассчитанный по модели тренда на заданную дату;

$t_{\alpha,\nu}$ – коэффициент доверия по распределению Стьюдента;

$$S_{\hat{y}} = \sqrt{\frac{\sum (y_i - \hat{y}_{t,i})^2}{\nu}}$$ – среднее квадратическое отклонение от тренда, скорректированное по числу степеней свободы $\nu = n - m$; y_i, $\hat{y}_{t,i}$ – соответственно фактические и расчётные значения

уровней динамического ряда; n – число уровней ряда динамики.

m – число параметров адекватной модели тренда (для линейной функции m = 2, для параболы m = 3 и т.д.).

Величины $t_{\alpha,\nu}$ при различных значениях ν приведены в табл. 5.3 (Практическая работа № 5).

Вероятностные границы интервала прогнозируемого явления есть:

$$\hat{y}_t - t_{\alpha,\nu} \cdot S_{\hat{y}} \le y_{пр} \le \hat{y}_t + t_{\alpha,\nu} \cdot S_{\hat{y}} . \qquad (7.2)$$

Методические указания по выявлению основной тенденции развития в рядах динамики

Задание

Динамика производства промышленной продукции в одном из регионов за 2005–2011 гг. (по условным данным, млн. руб.) приведена в таблице 7.4 (столбцы А и 1).

Используя метод аналитического выравнивания, построить модель тренда, отражающего закономерность развития явления.

Составить интервальный прогноз ожидаемого объема производства продукции в регионе на 2014 г., гарантируя результат с вероятностью 0,95.

Порядок выполнения задания

1) В качестве уравнения тренда выбираем линейную функцию:

$\hat{y}_t = a_0 + a_1 \cdot t.$

Так как количество уровней в ряду динамики нечетное, то временные даты (t) обозначим следующим образом (столбец 2 в табл. 7.4).

Таблица 7.4

Исходные и расчетные данные для определения параметров тренда

Год	Объём промышленной продукции y_i, млн руб.	t	t^2	$y \cdot t$	$\hat{y}_t$	$y_i - \hat{y}_t$	$(y_i - \hat{y}_t)^2$
А	1	2	3	4	5	6	7
2005	20,1	-3	9	-60,3	20,04	0,06	0,0036
2006	20,7	-2	4	-41,4	20,53	0,17	0,0289
2007	21,0	-1	1	-21,0	21,02	-0,02	0,0004
2008	21,2	0	0	0	21,51	-0,31	0,0961
2009	21,9	1	1	21,9	22,00	-0,1	0,0100
2010	22,6	2	4	45,2	22,49	0,11	0,0121
2011	23,1	3	9	69,3	22,98	0,12	0,0144
Итого	**150,6**	-	28	13,7	**150,6**	-	0,1655

2) Система нормальных уравнений для определения параметров a_0 и a_1 имеет вид (см. табл. 7.3):

$$a_0 \cdot n = \sum y_i$$

$$a_1 \cdot \sum t_i^2 = \sum y_i \cdot t_i ,$$

откуда: $a_0 = \frac{\sum y_i}{n}$ (представляет собой средний уровень ряда динамики

$\bar{y}$);

$$a_1 = \frac{\sum y_i t_i}{\sum t_i^2}.$$

Расчет необходимых величин для вычисления a_0 и a_1 дан в табл. 7.4 (столбцы 3, 4). По итоговым данным определите параметры уравнения:

$$a_0 = \frac{150{,}6}{7} = 21{,}51, \qquad a_1 = \frac{13{,}7}{28} = 0{,}49.$$

В результате получите следующее уравнение основной тенденции производства промышленной продукции в одном из регионов за 2005–2011 гг.:

$$\hat{y}_t = 21{,}51 + 0{,}49 \cdot t. \qquad (7.3)$$

Данное уравнение показывает, что в течение исследуемого периода выпуск промышленной продукции возрастал в среднем на 0,49 млн. руб. в год.

3) Вероятностные границы интервала прогноза объема выпуска продукции в 2014 г. есть:

$$\hat{y}_t - t_{\nu,\alpha} S_{\hat{y}} \le y_{пр} \le \hat{y}_t + t_{\nu,\alpha} S_{\hat{y}}.$$

Для 2014 г. показатель времени t = 6, используя уравнение (7.3), определите:

$$\hat{y}_{t2014} = 21{,}51 + 0{,}49 \cdot 6 = 24{,}45 \text{ млн. руб.}$$

При вероятности $\Phi(t) = 0{,}95$ значение уровня значимости $\alpha = 1 - \Phi(t) = 0{,}05$. Для линейной модели тренда число m = 2 и число степеней свободы $\nu = n - m = 7 - 2 = 5$. По табл. 5.3 лабораторной работы № 5, определите значение коэффициента $t_{5;0,05} = 2{,}571$.

Используя приведенное уравнение тренда (3.3), рассчитайте для каждого года теоретические (выровненные) значения $\hat{y}_{t,i}$: для 2005 г. $\hat{y}_{t=-3} = 21{,}51 + 0{,}49 \cdot (-3) = 20{,}04$;

для 2006 г. $\hat{y}_{t=-2} = 21{,}51 + 0{,}49 \cdot (-2) = 20{,}53$ и т.д. (см. столбец 5 в табл. 7.4). При правильном расчете выравненных уровней динамического ряда выполняется соотношение: $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_{t,i}$ (см. итоги столбцов 1 и 5).

Значение $S_{\hat{y}} = \sqrt{\dfrac{\sum (y_i - \hat{y}_{t,i})^2}{n}} = \sqrt{\dfrac{0{,}1655}{5}} = \sqrt{0{,}0331} = 0{,}18$ млн. руб.

Таким образом, объем выпуска промышленной продукции в регионе в 2014 г. составит:

$$24{,}45 - 2{,}571 \cdot 0{,}18 \le y_{пр} \le 24{,}45 + 2{,}571 \cdot 0{,}18;$$

$24{,}0$ млн. руб. $\le y_{пр} \le 24{,}9$ млн. руб.

Порядок выполнения работы

1. Ознакомьтесь с методикой выявления тренда в рядах динамики.
2. У преподавателя получите вариант индивидуального задания.
3. Используя метод аналитического выравнивания, постройте модель тренда, отражающего закономерность развития явления. 4. Сделайте выводы по работе и оформите отчет.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) результаты вычисления индивидуальных заданий;
3) анализ результатов в табличной форме.

Индивидуальные задания

Исследуйте основную тенденцию развития в рядах динамики по статистическим данным. Для этого:

1) используя метод аналитического выравнивания, постройте модель тренда, отражающего закономерность развития явления; исходные и расчетные данные для определения параметров тренда представьте в табличном виде;

2) изобразите графически фактические и выравненные уровни исследуемого динамического ряда;

3) составьте интервальный прогноз ожидаемого значения уровня ряда на год, указанный преподавателем, гарантируя результат с заданной вероятностью $\Phi(t)$ = 95 %.

1. Данные о численности населения РФ (на конец года) за 2003–2008 гг. приведены в табл. 7.5.

Таблица 7.5

Динамика численности населения РФ

Показатель	2003	2004	2005	2006	2007	2008
Численность населения, млн человек	144,2	143,5	142,8	142,2	142,0	141,9

Составьте интервальный прогноз ожидаемой численности населения РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

2. Данные о численности безработных РФ (на конец года) за 2003–2008 гг. приведены в табл. 7.6.

Таблица 7.6

Динамика численности безработных РФ

Показатель	2003	2004	2005	2006	2007	2008
Численность безработных, тыс. человек	5683	5775	5208	4999	4246	5289

Составьте интервальный прогноз ожидаемой численности безработных РФ на 2012 год, гарантируя результат с заданной вероятностью 99 %.

3. Данные о валовом накоплении РФ за 2003–2008 гг. приведены в табл. 7.7.

Таблица 7.7

Динамика валового накопления РФ

Показатель	2003	2004	2005	2006	2007	2008
Валовое накопление, млрд руб.	2755,1	3558,9	4338,7	5748,8	8031,7	10642,5

Составьте интервальный прогноз ожидаемого валового накопления РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

4. Данные о динамике производительности труда в строительстве по РФ (в процентах к предыдущему году) за 2003–2007 гг. приведены в табл. 7.8.

Таблица 7.8

Динамика производительности труда в строительстве по РФ

Показатель	2003	2004	2005	2006	2007
Производительность труда в строительстве	105,3	106,9	105,9	115,8	112,8

Составьте интервальный прогноз ожидаемой производительности труда в строительстве на 2012 год, гарантируя результат с заданной вероятностью 99 %.

5. Данные о динамике производительности труда при добыче полезных ископаемых в РФ (в процентах к предыдущему году) за 2003–2007 гг. приведены в табл. 7.9.

Таблица 7.9

Динамика производительности труда в РФ

Показатель	2003	2004	2005	2006	2007
Производительность труда при добыче полезных ископаемых	109,2	107,3	106,3	102,5	102,3

Составьте интервальный прогноз ожидаемой производительности труда при добыче полезных ископаемых на 2012 год, гарантируя результат с заданной вероятностью 95 %.

6. Данные о среднемесячной номинальной начисленной заработной плате работников организаций РФ за 2003–2008 гг. приведены в табл. 7.10.

Таблица 7.10

Динамика среднемесячной номинальной заработной платы работников организаций РФ

Показатель	2003	2004	2005	2006	2007	2008
Среднемесячная номинальная заработная плата, руб.	5498,51	6739,5	8554,9	10633,9	13593,4	17226,3

Составьте интервальный прогноз ожидаемой среднемесячной номинальной начисленной заработной платы работников организаций РФ на 2012 год, гарантируя результат с заданной вероятностью 99 %.

6. Данные о численности пенсионеров РФ (на конец г ода) за 2003–2008 гг. приведены в табл. 7.11.

Составьте интервальный прогноз ожидаемой численности пенсионеров РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

Таблица 7.11 Динамика численности пенсионеров РФ

Показатель	2003	2004	2005	2006	2007	2008
Численность пенсионеров, тыс. человек	38164	38184	38313	38325	38467	38598

8. Данные об обороте розничной торговли РФ за 2003–2008 гг. приведены в табл. 7.12.

Таблица 7.12 Динамика оборота розничной торговли РФ

Показатель	2003	2004	2005	2006	2007	2008
Оборот розничной торговли, млрд руб.	4530	5642	7041	8712	10869	13915

Составьте интервальный прогноз об обороте розничной торговли РФ на 2012 год, гарантируя результат с заданной вероятностью 99 %.

9. Данные о вводе в действие общей площади жилых домов в РФ за 2003–2008 гг. приведены в табл. 7.13.

Таблица 7.13

Динамика общей площади жилых домов в РФ

Показатель	2003	2004	2005	2006	2007	2008
Общая площадь жилых домов, млн $м^2$	36,4	41,0	43,6	50,6	61,2	64,1

Составьте интервальный прогноз о вводе в действие общей площади жилых домов РФ на 2013 г., гарантируя результат с заданной вероятностью 95 %.

1. Данные о динамике производительности труда на транспорте и связи в РФ (в процентах к предыдущему году) за 2003–2007 гг. приведены в табл. 7.14.

Таблица 7.14

Динамика производительности труда на транспорте и связи в РФ

Показатель	2003	2004	2005	2006	2007
Производительность труда на транспорте и связи	107,5	108,7	102,1	110,7	107,6

Составьте интервальный прогноз ожидаемой производительности труда на транспорте и связи на 2012 год, гарантируя результат с заданной вероятностью 99 %.

11. Данные о численности женского населения РФ (на 1 января) за 2002–2008 гг. приведены в табл. 7.15.

Таблица 7.15 Динамика численности женского населения РФ

Показатель	2002	2003	2004	2005	2006	2007	2008
Численность женского населения, тыс. человек	77562	77473	77144	76871	76590	76372	76292

Составьте интервальный прогноз ожидаемой численности женского населения РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

12. Данные о грузообороте транспорта РФ за 2003–2008 гг. приведены в табл. 7.16.

Таблица 7.16

Динамика грузооборота транспорта РФ

Показатель	2003	2004	2005	2006	2007	2008
Грузооборот транспорта, трлн т·км	4,3	4,6	4,7	4,8	4,9	4,9

Составьте интервальный прогноз о грузообороте транспорта РФ на 2012 год, гарантируя результат с заданной вероятностью 99 %.

13. Данные о численности мужского населения РФ (на 1 января) за 2002–2008 гг. приведены в табл. 7.17.

Таблица 7.17

Динамика численности мужского населения РФ

Показатель	2002	2003	2004	2005	2006	2007	2008
Численность мужского населения, тыс. человек	67605	67491	67024	66603	66164	65849	65717

Составьте интервальный прогноз ожидаемой численности мужского населения РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

14. Данные о валовом внутреннем продукте РФ за 2003–2008 гг. приведены в табл. 7.18.

Таблица 7.18

Динамика валового внутреннего продукта РФ

Показатель	2003	2004	2005	2006	2007	2008
Валовый внутренний продукт, млрд руб.						

	13243	17048	21625	26904	33111	41668

Составьте интервальный прогноз о валовом внутреннем продукте РФ на 2012 год, гарантируя результат с заданной вероятностью 99 %.

15. Данные о численности экономически активного городского населения в возрасте от 15 до 72 лет в РФ за 2003–2009 гг. приведены в табл. 19.

Таблица 7.19

Динамика численности экономически активного городского населения в РФ

Показатель	2003	2004	2005	2006	2007	2008	2009
Численность экономически активного городского населения, тыс. человек	55176	55727	55715	55907	56578	56999	56580

Составьте интервальный прогноз ожидаемой численности экономически активного городского населения РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

16. Данные о динамике производительности труда при операциях с недвижимым имуществом, арендой и предоставлением услуг в РФ (в процентах к предыдущему году) за 2003–2007 гг. приведены в табл. 7.20.

Таблица 7.20

Динамика производительности труда при операциях с недвижимым имуществом, арендой и предоставлением услуг в РФ

Показатель	2003	2004	2005	2006	2007
Производительность труда	102,5	101,3	112,4	106,2	115,6

Составьте интервальный прогноз ожидаемой производительности труда при операциях с недвижимым имуществом, арендой и предоставлением услуг на 2013 год, гарантируя результат с заданной вероятностью 99 %.

17. Данные о платных услугах населению РФ за 2003–2008 гг. приведены в табл. 7.21.

Таблица 7.21

Динамика платных услуг населению РФ

Показатель	2003	2004	2005	2006	2007	2008
Платные услуги населению, млрд руб.	1431	1790	2272	2799	3435	4081

Составьте интервальный прогноз о платных услугах населению РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

18. Данные о динамике производительности труда в обрабатывающих производствах в РФ (в процентах к предыдущему году) за 2003–2007 гг. приведены в табл. 7.22.

Таблица 7.22

Динамика производительности труда в обрабатывающих производствах в РФ

Показатель	2003	2004	2005	2006	2007
Производительность труда в обрабатывающих производствах	108,8	106,3	107,1	108,1	106,5

Составьте интервальный прогноз ожидаемой производительности труда в обрабатывающих производствах на 2013 год, гарантируя результат с заданной вероятностью 99 %.

19. Данные о численности экономически активного городского населения в возрасте от 15 до 72 лет, занятого в экономике РФ, за 2003–2009 гг. приведены в табл. 7.23.

Таблица 7.23

Динамика численности экономически активного городского населения РФ

Показатель	2003	2004	2005	2006	2007	2008	2009
Численность экономически активного городского населения, занятого в экономике, тыс. человек	51363	51828	52464	53006	54079	53692	52413

Составьте интервальный прогноз ожидаемой численности экономически активного городского населения, занятого в экономике, на 2014 год, гарантируя результат с заданной вероятностью 95 %.

20. Данные о продукции сельского хозяйства в РФ за 2003–2008 гг. приведены в табл. 7.24.

Таблица 7.24

Динамика продукции сельского хозяйства в РФ

Показатель	2003	2004	2005	2006	2007	2008
Продукция сельского хозяйства, млрд руб.	1154,9	1345,2	1494,6	1711,3	1940,5	2602,7

Составьте интервальный прогноз о продукции сельского хозяйства в РФ на 2013 год, гарантируя результат с заданной вероятностью 99 %.

21. Данные о динамике производительности труда при производстве и распределении электроэнергии, газа и воды в РФ (в процентах к предыдущему году) за 2003–2007 гг. приведены в табл. 7.25.

Таблица 7.25

Динамика производительности труда при производстве и распределении электроэнергии, газа и воды в РФ

Показатель	2003	2004	2005	2006	2007
Производительность труда	103,7	100,4	103,7	103,2	100,3

Составьте интервальный прогноз ожидаемой производительности труда при производстве и распределении электроэнергии, газа и воды на 2012 год, гарантируя результат с заданной вероятностью 95 %.

22. Данные об индексе потребительских цен (декабрь к декабрю предыдущего года) в РФ за 2003–2008 гг. приведены в табл. 7.26.

Таблица 7.26

Динамика индекса потребительских цен в РФ

Показатель	2003	2004	2005	2006	2007	2008
Индекс потребительских цен, %	112,0	111,7	110,9	109,0	111,9	113,3

Составьте интервальный прогноз индекса потребительских цен РФ на 2012 год, гарантируя результат с заданной вероятностью 99 %.

23. Данные об объеме средств Стабилизационного фонда РФ (на начало года) за 2004–2008 гг. приведены в табл. 7.27.

Таблица 7.27

Динамика объема средств Стабилизационного фонда РФ

Показатель	2004	2005	2006	2007	2008
Объем средств Стабилизациионного фонда РФ, млрд руб.	106,0	522,3	1237,0	2346,9	3849,1

Составьте интервальный прогноз объема средств Стабилизационного фонда РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

24. Данные о динамики среднемесячной номинальной начисленной заработной платы работников, занятых в строительстве по республике Татарстан, за 2003–2008 гг. приведены в табл. 7.28.

Таблица 7.28

Динамика среднемесячной номинальной заработной платы работников по республике Татарстан

Показатель	2003	2004	2005	2006	2007	2008
Заработная плата, руб.	4427	5584	7878	9475	13225	19753

Составьте интервальный прогноз ожидаемой среднемесячной номинальной начисленной заработной платы работников, занятых в строительстве по республике Татарстан, на 2013 год, гарантируя результат с заданной вероятностью 99 %.

25. Данные о среднедушевых денежных расходах населения в РФ за 2003–2008 гг. приведены в табл. 7.29.

Таблица 7.29

Динамика среднедушевых денежных расходах населения в РФ

Показатель	2003	2004	2005	2006	2007	2008
Среднедушевые денежные расходы, руб. в месяц	5030	6296	7992	9852	12118	15111

Составьте интервальный прогноз о среднедушевых денежных расходах населения в РФ на 2012 год, гарантируя результат с заданной вероятностью 95 %.

Контрольные вопросы

1. Дайте определение ряда динамики. Из каких элементов он состоит?
2. Какие динамические ряды называются моментными, в чем заключается их особенность?
3. Какие динамические ряды называются интервальными, в чем заключается их особенность?
4. Каковы причины возникновения несопоставимости динамических рядов?
5. Дайте определение основной тенденции развития в рядах динамики.
6. Какие статистические методы используются для выявления тренда в динамических рядах?
7. В чем сущность метода укрупнения интервалов и для чего он применяется?
8. Как производится сглаживание рядов динамики методом скользящей средней?
9. В чем сущность метода аналитического выравнивания динамических рядов?
10. Как определяется тип уравнения тренда?

Практическая работа № 8

Использование индексов в экономико-статистических исследованиях (4 часа)

Цель работы

1. Изучить методику расчета индивидуальных, агрегатных, средних индексов, индексов постоянного и переменного состава.
2. Сформировать практические навыки расчета индивидуальных агрегатных, средних индексов, индексов постоянного и переменного состава.
3. Провести расчет индивидуальных, агрегатных, средних индексов, индексов постоянного и переменного состава.

Краткая теория

Экономический индекс – это относительная величина, которая характеризует изменение исследуемого явления во времени, в пространстве или по сравнению с некоторым эталоном (планируемым, нормативным уровнем и т.п.). Если в качестве базы сравнения используется уровень за какой либо предшествующий период, получают динамический индекс; если же базой является уровень того же явления по другой территории, то территориальный индекс.

Индивидуальные индексы и сводные индексы в агрегатной форме [6]

Простейшим показателем, используемым в индексном анализе, является индивидуальный индекс, который характеризует изменение во времени (или в пространстве) отдельных элементов той или иной совокупности. Так, *индивидуальный индекс цены* рассчитывается по формуле

$$i_p = \frac{p_1}{p_0}, \quad (8.1)$$

где p_1 − цена товара в текущем периоде; p_0 − цена товара в базисном периоде.

Оценить изменение объемов продажи товара в натуральных единицах измерения позволяет *индивидуальный индекс физического объема реализации:*

$$i_q = \frac{q_1}{q_0}, \qquad (8.2)$$

где q_1 – количество товара, реализованное в текущем периоде; q_0 – количество товара, реализованное в базисном периоде.

Изменение объема реализации товара в стоимостном выражении отражается *индивидуальный индекс товарооборота:*

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0}. \qquad (8.3)$$

Индивидуальные индексы, в сущности, представляют собой относительные показатели динамики или темпы роста и по данным за несколько периодов времени могут рассчитываться в цепной или базисной формах.

Сводный индекс – это сложный относительный показатель, который характеризует среднее изменение социально – экономического явления, состоящего из непосредственно несоизмеримых элементов. Исходной формой сводного индекса является агрегатная.

При расчете агрегатного индекса для разнородной совокупности находят такой общий показатель, в котором можно объединить все ее элементы. Рассмотрим например с розничным ценами. Цены различных товаров, реализуемых в розничной торговле, складывать неправомерно, однако с экономической точки зрения вполне допустимо суммировать товарооборот по этим товарам. Если мы сравним товарооборот в текущем периоде в его величиной в базисном периоде, то получим *сводный индекс товарооборота:*

$$i_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}. \qquad (8.4)$$

На величину данного индекса оказывает влияние изменение как цен на товары, так и объемов их реализации. Для того чтобы оценить изменение только цен (индексируемой величины) необходимо количество проданных товаров (веса индекса) зафиксировать на каком либо постоянном уровне. При исследовании динамики таких показателей, как цена, себестоимость, производительность труда, урожайность количественный показатель обычно фиксируют на уровне текущего периода. Таким способом получают *сводный индекс цен*:

индекс **Паше:** $$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1};\tag{8.5}$$

индекс **Ласпейреса:** $$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}.\tag{8.6}$$

Числитель общего индекса – фактический товарооборот текущего периода, знаменатель – условная величина, показывающая, каким был бы товарооборот в текущем периоде при условии сохранения цен на базисном уровне. Разность числителя и знаменателя показывает сумму **экономии** (если знак «–») или сумму **перерасхода** («+») покупателей при изменении цен на эти товары:

$$E = \sum p_1 q_1 - \sum p_0 q_1.\tag{8.7}$$

Индекс Пааше несколько занижает темпы инфляции, индекс Ласпейреса – завышает.

Индексы Пааше и Ласпейреса имеют **разное экономическое содержание**:

– индекс Пааше показывает, на сколько изменились цены в отчетном периоде по сравнению с базисным по товарам, реализованным в отчетном периоде, и *фактическую экономию* (*перерасход*) от изменения цен, то есть индекс цен Пааше показывает, на сколько товары в отчетном периоде стали дороже (дешевле), чем в базисном;

– индекс Ласпейреса показывает, на сколько изменились цены в отчетном периоде по сравнению с базисным по товарам, реализованным в базисном периоде, и *условную экономию* (*перерасход*), которую можно было бы получить от изменения цен, то есть индекс цен Лайспейреса показывает, во сколько раз товары базисного периода подорожали (подешевели) из-за изменения цен на них в отчетном периоде.

Сводный индекс физического объема реализации характеризует изменение количества проданных товаров не в денежных, а в физических единицах измерения:

$$i_q = \frac{\sum q_1 p_0}{}.\tag{8.8}$$

$$\sum q_0 p_0$$

Весами в данном индексе выступают цены, которые фиксируются на базисном уровне.

Между рассчитанными индексами существует следующая взаимосвязь:

$$I_p \cdot I_q = I_{pq}. \tag{8.9}$$

Пример 1. [6] Имеются следующие данные о реализации плодовоягодной продукции в области (табл. 8.1).

Таблица 8.1 **Реализация плодово-ягодной продукции в области**

Наименование товара	Июль		Август		Расчетные графы, руб.		
	цена за 1 кг, руб p_0	продано, т q_0	цена за 1 кг, руб p_1	продано, т q_1	p_0q_0	p_1q_1	p_0q_1
Абрикосы	67	15	58	23	1005	1334	1541
Груши	54	23	46	28	1242	1288	1512
Яблоки	32	25	29	32	800	928	1024
Итого	–	–	–	–	3047	3550	4077

Рассчитайте индекс товарооборота.

Решение:

$$i_{pq} = \frac{\sum p_1q_1}{\sum p_0q_0} = \frac{3550}{3047} = 1{,}165, \text{ или } 116{,}5\ \%.$$

Следовательно, товарооборот в целом по данной товарной группе в текущем периоде по сравнению с базисным увеличился на 16,5 %

(116,5−100,0).

Вычислим сводный индекс цен:

$$i_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{3550}{4077} = 0{,}871, \text{ или } 87{,}1\ \%.$$

Этот результат означает, что цены в августе по сравнению с июлем в среднем снизились на 12,9 %.

Числитель представляет собой сумму денег, фактически уплаченных покупателями за приобретенные в текущем периоде товары. Знаменатель же показывает, какую сумму покупатели заплатили бы за те же товары, если бы цены не изменились. Разность числителя и знаменателя будет отражать величину экономии (если знак «–») или перерасхода («+») покупателей от изменения цен:

$$E = \sum p_1 q_1 - \sum p_0 q_1 = 3550 - 4077 = -572 \text{ тыс. руб.}$$

Размер экономии покупателей от снижения цен на продукцию составил 572 тыс. руб.

Индекс физического объема реализации составит:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{4077}{3047} = 1{,}338, \text{ или } 133{,}8\ \%.$$

Физический объем реализации (товарооборота) плодово-ягодной продукции увеличился на 33,8 %.

Используя взаимосвязь индексов, проверим правильность вычислений:

$$I_{pq} = I_p \cdot I_q = 0{,}871 \cdot 1{,}338 = 1{,}165, \text{ или } 116{,}5\ \%.$$

Следовательно, в результате изменения цен на продукцию и изменения объемов продаж стоимость товарооборота выросла на 16,5 %.

Индивидуальный индекс себестоимости характеризует изменение себестоимости отдельного вида продукции в текущем периоде по сравнению с базисным:

$$i_z = \frac{z_1}{z_0}. \qquad (8.10)$$

Для определения общего изменения уровня себестоимости нескольких видов продукции, выпускаемых предприятием, рассчитывается сводный индекс себестоимости. При этом себестоимость взвешивается по объему производства отдельных видов продукции текущего периода:

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}. \qquad (8.11)$$

Числитель этого индекса отражает затраты на производство текущего периода, а знаменатель – условную величину затрат при сохранении себестоимости на базисном уровне. Разность числителя и знаменателя показывает сумму экономии предприятия от снижения себестоимости:

$$E = \sum z_1 q_1 - \sum z_0 q_1. \qquad (8.12)$$

Сводный индекс физического объема продукции, взвешенный по себестоимости, имеет следующий вид:

$$I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0}. \qquad (8.13)$$

Сводный индекс затрат на производство:

$$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0}. \qquad (8.14)$$

Все три индекса взаимосвязаны между собой:

$$I_z \cdot I_q = I_{zq}. \qquad (8.15)$$

При анализе изменений в производительности труда возможны два подхода к расчету индексов. Первый подход основан на учете количества продукции, вырабатываемого в единицу времени (w). При втором подходе производительность труда определяется затратами рабочего времени на единицу продукции (t).

Количество продукции, вырабатываемое в единицу времени (в натуральном выражении), и затраты времени на единицу продукции взаимосвязаны между собой:

$$w = \frac{1}{t}.$$

Индивидуальные индексы производительности труда, основанные на этих показателях, имеют следующий вид:

$$i_w = \frac{w_1}{w_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0}; \qquad (8.16)$$

$$i_w = \frac{t_0}{t_1} = \frac{T_0}{q_0} : \frac{T_1}{q_1}, \qquad (8.17)$$

где T – суммарные затраты времени на выпуск данной продукции в человеко-часах, человеко-днях или человеко-месяцах (в последнем случае соответствует общей численности работников).

Трудоемкость является обратным показателем, поэтому снижение трудоемкости в текущем периоде по сравнению с базисным свидетельствует о росте производительности труда. *Сводный индекс производительности труда (по трудоемкости):*

$$I_w = \frac{\sum t_0 q_1}{\sum t_1 q_1}. \qquad (8.18)$$

Знаменатель этого индекса отражает реально имевшие место общие затраты времени на выпуск всей продукции в текущем периоде (T_1). Числитель представляет собой условную величину, показывающую, какими были бы затраты времени на выпуск этой продукции, если бы трудоемкость не изменилась.

Пример 2. [6] По данным таблицы 8.2 надо рассчитать рост производительности труда на предприятии.

Таблица 8.2

Трудоемкость и выпуск продукции на предприятии

Вид изделия	Затраты времени на 1 изделие, чел.-ч		Произведено, шт.		Расчетные графы, чел.-ч	
	январь t_0	февраль t_1	январь q_0	Февраль q_1	$t_0 q_1$	$t_1 q_1$
А	1,5	1,2	363	403	604,5	483,6
Б	1,0	0,8	411	424	424,0	339,2
В	0,9	0,7632	838	903	812,7	632,1
Итого	–	–	–	–	1841,2	1454,9

Рассчитайте сводный индекс производительности труда по трудоемкости.

Решение:

$$I_w = \frac{\sum t_0 q_1}{\sum t_1 q_1} = \frac{1841,2}{1454,9} = 1,266, \text{ или } 126,6\%.$$

Следовательно, прирост производительности труда в целом по предприятию составил 12,6 %.

Индекс производительности труда по трудоемкости связан с индексом *затрат рабочего времени (труда) и с индексом физического объема продукции, взвешенным по трудоемкости:*

$$I_w \cdot I_T = I_q, \tag{8.19}$$

или

$$I_q = \frac{\sum t_0 q_1}{\sum t_1 q_1} \cdot \frac{\sum T_1}{\sum T_0} = \frac{\sum q_1 t_0}{\sum q_0 t_0}. \tag{8.20}$$

При расчете *сводного индекса производительности труда в стоимостном выражении (по выработке)* необходимо количество продукции, произведенной за каждый период, взвесить по каким-либо ценам, принятым за сопоставимые. В качестве сопоставимых могут выступать цены текущего, базисного или какого – либо другого периода или средние цены. Индекс производительности труда рассчитывается по формуле:

$$I_w = \frac{\sum q_1 p}{\sum T_1} : \frac{\sum q_0 p}{\sum T_0}. \qquad (8.21)$$

Первая часть этой формулы представляет собой среднюю выработку в отчетном периоде, вторая часть – в базисном.

Пример 3. [6] Имеются следующие данные о производстве продукции и отпускных ценах предприятия (табл. 8.3).

Таблица 8.3

Данные о производстве продукции

Вид продукции	Сентябрь		Октябрь		Отпускная цена, руб. p	Расчетные графы, руб.	
	произве-дено, шт. q_0	трудовые затраты, чел.-ч T_0	произве-дено, шт. q_1	трудовые затраты, чел.-ч T_1		q_0p	q_1p
А	256	854	343	947	560	143360	192080
Б	203	931	273	989	784	159152	214032
В	432	1087	485	1105	965	416880	468025
Итого	–	2872	–	3041	–	719392	874137

Вычислить индекс производительности труда.

Решение:

$$I_w = \frac{\sum q_1 p}{\sum T_1} : \frac{\sum q_0 p}{\sum T_0} = \frac{874137}{3041} : \frac{719392}{2872} = 287{,}5 : 250{,}5 = 1{,}148, \text{ или } 114{,}8\ \%.$$

Итак, в текущем периоде за 1 чел. – ч вырабатывалось 287,5 руб. продукции, а в базисном – 250,5 руб. Прирост производительности труда составил 14,8 %.

Умножение индекса производительности труда по выработке на индекс затрат рабочего времени приводит к *индексу физического объема продукции, взвешенному по цене:*

$$I_w \cdot I_T = I_q, \tag{8.22}$$

или

$$\left(\frac{\sum q_1 p}{\sum T_1} : \frac{\sum q_0 p}{\sum T_0}\right) \cdot \frac{\sum T_1}{\sum T_0} = \frac{\sum q_1 p}{\sum q_0 p}. \tag{8.23}$$

Сводные индексы в средней арифметической и средней гармонической формах [6]

Часто на практике вместо индексов в агрегатной форме удобнее использовать средние арифметические и средние гармонические индексы. Любой сводный индекс можно представить как среднюю взвешенную из индивидуальных индексов. Однако при этом форму средней нужно выбрать таким образом, чтобы полученный средний индекс был тождествен исходному агрегатному индексу.

Если имеются данные о стоимости проданной продукции в текущем периоде (p_1q_1) и индивидуальными индексами цен $\left(i_p = \frac{p_1}{p_0}\right)$, тогда i_p =

в знаменателе сводного индекса цен $\left(I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \right)$ можно использовать следующую замену:

$$p_0 = \frac{1}{i_p} p_1.$$

Таким образом, сводный индекс цен будет выражен в форме средней гармонической из индивидуальных индексов:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1}. \qquad (8.24)$$

При расчете сводного индекса физического объема товарооборота $\left(I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \right)$ можно использовать среднюю арифметическую форму.

При этом в числителе производится замена:

$$q_1 = i_q q_0.$$

Тогда индекс примет вид:

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}. \qquad (8.25)$$

Пример 4. [6] По данным табл. 8.4 получите сводную оценку изменения цен.

Таблица 8.4

Реализация овощной продукции

Товар	Реализация в текущем периоде, руб. p_1q_1	Изменение цен в текущем периоде по сравнению с базисным, % i_p 100% − 100%	Расчетные графы	
			i_p	$\frac{p_1q_1}{i_p}$
Перец	43000	+ 1,5	1,050	40952
Капуста	12500	+ 1,1	1,01	12376
Картофель	14000	– 0,9	0,991	14127
Итого	69500			67455

Решение:

Вычислим средний гармонический индекс:

$$I_p = \frac{\sum p_1q_1}{\sum \frac{1}{i_p} p_1q_1} = \frac{69500}{67455} = 1{,}030 \text{ или } 103{,}0\ \%.$$

Цены по данной товарной группе в текущем периоде по сравнению с базисным в среднем выросли на 3,0 %.

При расчете сводного индекса физического объема товарооборота

$\left(I_p = \frac{\sum q_1p_0}{\sum q_0p_0}\right)$ можно использовать среднюю арифметическую форму.

При этом в числителе производится замена:

$q_1 = i_q q_0$.

Тогда индекс примет вид:

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}. \qquad (8.25)$$

Пример 5. [6] Имеются следующие данные о реализации товаров (табл. 8.5).

Таблица 8.5

Реализация товаров в натуральном и стоимостном выражениях

Товар	Реализация в базисном периоде, руб. $q_0 p_0$	Изменение физического объема реализации в текущем периоде по сравнению с базисным, % $i_q \cdot 100\ \% - 100\ \%$	Расчетные графы	
			i_q	$i_q \cdot q_0 p_0$
А	26000	– 3,7	0,963	25038
Б	38000	– 7,6	0,924	35112
В	43000	+ 2,5	1,025	107500
Итого	107000			167650

Рассчитайте средний индекс физического объема. **Решение:**

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{1676500}{107000} = 1{,}567, \text{ или } 156{,}7\ \%.$$

Физический объем реализации данных товаров в среднем увеличился на 56,7 %.

В средней арифметической форме также может рассчитываться и индекс производительности труда по трудоемкости, известный как *индекс С. Г. Струмилина:*

$$i_{T_1} \quad \frac{T_0 : T_1}{} \quad T_1$$

$$I_w = \frac{}{T_1} \quad q_0 \; T q_1 \; 1 \; . \qquad (8.26)$$

Индексы переменного и постоянного состава, индекс структурных сдвигов

На формирование среднего уровня качественного показателя оказывают влияние два фактора: во-первых, изменение индивидуальных значений самой индексируемой величины в отчетном периоде по сравнению с базисным и, во-вторых, изменение структуры исследуемой совокупности (уменьшение или увеличение доли единиц с более низким или высоким уровнем значения показателя).

Относительное изменение среднего уровня качественного показателя за счет каждого из этих факторов оценивается с помощью системы индексов переменного, постоянного (фиксированного) состава и индекса структурных сдвигов.

Индексом переменного состава ($I_{\text{пер.сост}}$) называют отношение средних уровней определенного показателя за два периода. Индекс переменного состава показывает изменение среднего уровня качественного показателя *за счет двух факторов.* В общем виде он рассчитывается как отношение среднего уровня показателя в отчетный период к среднему уровню показателя в базисном периоде:

$$I_{\text{пер.сост}} = \bar{x}_1 : \bar{x}_2 = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}. \qquad (8.27)$$

При изучении *изменения цен* индекс переменного состава можно записать следующим образом:

$$I_{\text{пер.сост}} = \bar{p}_1 : \bar{p}_2 = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0}. \qquad (8.28)$$

Абсолютное изменение среднего уровня качественного показателя *за счет всех факторов* покажет разница между числителем и знаменателем рассматриваемого индекса, например для цен:

$$\Delta^{цен}_{общ} = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0}. \qquad (8.29)$$

Индекс постоянного (фиксированного) состава показывает изменение среднего уровня качественного показателя *за счет изменений индивидуальных значений самой индексируемой величины*. Веса при этом фиксируются на уровне отчетного периода (f_1), то есть:

$$I_{пост.сост} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} = \frac{\sum x_1 f_1}{\sum x_0 f_1}. \qquad (8.30)$$

При изучении изменения цен:

$$I_{пост.сост} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_1}{\sum q_1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}. \qquad (8.31)$$

Абсолютное изменение среднего уровня качественного показателя *за счет изменения индексируемой величины* покажет разница между числителем и знаменателем рассматриваемого индекса, например для цен:

$$\Delta^{цен}_{p} = \frac{\sum p_1 q_1}{\sum q_1} - \frac{\sum p_0 q_1}{\sum q_1}. \qquad (8.32)$$

Индекс структурных сдвигов позволяет оценить влияние на формирование среднего уровня качественного показателя изменений в структуре исследуемой совокупности:

$$I_{стр} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}. \qquad (8.33)$$

При изучении изменения цен:

$$I_{стр} = \frac{\sum p_0 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0}. \qquad (8.34)$$

Абсолютное изменение среднего уровня качественного показателя *за счет структурных сдвигов* покажет разница между числителем и знаменателем рассматриваемого индекса; например для цен:

$$\Delta^{цен}_{стр} = \frac{\sum p_0 q_1}{\sum q_1} - \frac{\sum p_0 q_0}{\sum q_0}. \qquad (8.35)$$

Поскольку индекс переменного состава показывает изменение исследуемого явления за счет всех факторов, между индексами существует следующая взаимосвязь:

$$I_{пер.сост} = I_{пост.сост} \times I_{стр}. \qquad (8.36)$$

Пример 6. Имеются данные о производстве и себестоимости двух однотипных изделий (табл. 8.6).

Таблица 8.6.

Данные о себестоимости и объемах производства

	Произведено, тыс. шт.	Себестоимость одного изделия, руб.

Изделие	Базисный период	Отчетный период	Базисный период	Отчетный период
№ 1 № 2	91 150	135 103	20 15	19 19
Итого	140	150		

Рассчитайте:

1) изменение себестоимости в целом по обоим предприятиям с помощью индексов переменного и фиксированного состава; 2) индекс структурных сдвигов.

Решение:

Введите следующие обозначения:

z_0 – себестоимость одного изделия данного вида в базисном периоде; z_1 – себестоимость одного изделия данного вида в отчетном периоде; q_0 – количество изделий данного вида, произведенных в базисном периоде; q_1 – количество изделий данного вида, произведенных в отчетном периоде.

1) Индекс себестоимости переменного состава рассчитайте по формуле (8.27):

$$I_{пер.сост}^{z} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{19 \cdot 135 + 19 \cdot 103}{238} : \frac{20 \cdot 91 + 15 \cdot 150}{241} =$$

$$= \frac{2565 + 1957}{238} : \frac{1820 + 2250}{241} = \frac{4522}{238} : \frac{4070}{241} = \frac{19}{16,89} = 1,125$$

или 112,5 %.

Под влиянием изменения индивидуальных себестоимостей и структурных сдвигов в производстве данных изделий средняя себестоимость увеличилась на 12,5 % (112,5 % – 100 % = 12,5 %).

Абсолютное изменение среднего уровня себестоимости двух однотипных изделий *за счет всех факторов*, т. е. индивидуальных себестоимостей изделий и структурных сдвигов в их производстве, покажет разница между числителем и знаменателем рассматриваемого индекса:

$$\Delta \bar{z}_{общ} = 19 - 16{,}89 = 2{,}11 \text{ руб.}$$

2) Индекс себестоимости постоянного (фиксированного) состава рассчитайте по формуле (8.30):

$$I_{пост.сост\,z} = \frac{\sum z_1 q_1}{\sum z_0 q_1} = \frac{4522}{20 \cdot 135 + 15 \cdot 103} = \frac{4522}{2700 + 1545} = \frac{4522}{4245} = 1{,}065$$

или 106,5 %.

Под влиянием изменения индивидуальных себестоимостей средняя себестоимость увеличилась на 6,5 % (106,5 % – 100 % = 6,5 %).

Абсолютное изменение средней себестоимости изделий за счет изменения индивидуальных себестоимостей равно:

$$\Delta \bar{z} = \frac{4522}{238} - \frac{4245}{238} = 19{,}0 - 17{,}84 = 1{,}16 \text{ руб.}$$

3) Индекс структурных сдвигов рассчитайте по формуле (8.33):

$$I_{стр\,z} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{4245}{238} : \frac{4070}{241} = \frac{17{,}836}{16{,}888} = 1{,}056$$

или 105,6 %.

Это означает, что вследствие изменения структуры произведенной продукции средняя себестоимость увеличилась на 5,6 % (105,6 % – 100 % = 5,6 %).

Абсолютное изменение средней себестоимости изделия за счет изменения структуры произведенной продукции равно:

$$\Delta \bar{z}_{стр} = 17,84 - 16,89 = 0,95 \text{ руб.}$$

Проверим взаимосвязь индексов по формуле:

$$I_{пер.сост} = I_{пост.сост} \cdot I_{стр} = 1,065 \cdot 1,056 = 1,125.$$

Проверьте баланс абсолютных приростов:

$$\Delta \bar{z}_{общ} = \Delta \bar{z} + \Delta \bar{z}_{стр} = 1,16 \text{ руб} + 0,95 \text{ руб} = 2,11 \text{ руб.}$$

Порядок выполнения работы

1. Ознакомьтесь с методикой исчисления экономических индексов.
2. У преподавателя получите вариант индивидуального задания.
3. Проведите вычисления экономических индексов. 4. Сделайте выводы по работе и оформить отчет.

Оформление отчета

Отчет о лабораторной работе должен содержать:

1) постановку задачи;
2) результаты вычисления индивидуальных заданий; 3) анализ результатов.

Индивидуальные задания

Задание 1. Имеются следующие данные за два периода о ценах и объемах реализации трех видов товаров по одному из торговых предприятий (табл. 8.7).

Таблица 8.7 Данные о реализации трех видов товаров

Вид товара	Базисный период	Текущий период

	Цена за единицу, руб.	Продано товаров, шт.	Цена за единицу, руб.	Продано товаров, шт.
А	136	3421	189	2560
Б	245	1871	263	2123
В	147	964	185	1435

Рассчитайте:

1) индивидуальные индексы цен (по каждому виду товаров);
2) индивидуальные индексы физического объема реализации товаров;
3) общий индекс цен: Ласпейреса и Пааше;
4) общий индекс физического объема реализации;
5) индекс товарооборота (стоимость товаров).

Задание 2. Имеются следующие данные о продаже товаров на рынке города (табл. 8.8).

Таблица 8.8 Реализация продукции на рынке города

Вид товара	Продано, тыс. кг		Цена за 1 кг, руб.	
	Июль	Август	Июль	Август
Яблоки	86	157	58	49
Лук	78	126	25	17

Рассчитайте:

1) индивидуальные индексы цен и объема проданного товара;
2) общий индекс цен: Ласпейреса и Пааше;
3) общий индекс физического объема реализации;
4) прирост товарооборота – всего, в том числе за счет изменения цен и объема продажи товаров.

Задание 3. Имеются следующие данные о производстве и себестоимости некоторого продукта на двух предприятиях за два периода
(табл. 8.9).

Таблица 8.9 Выпуск и себестоимость продукции на предприятиях

Предприятия	Произведено, тыс. ед.		Себестоимость единицы продукта, руб.	
	Базисный период	Отчетный период	Базисный период	Отчетный период

№ 1	78	95	250	235
№ 2	62	55	195	175
Итого	140	150	–	–

Рассчитайте:

1) изменение себестоимости продукта по каждому предприятию;

2) изменение себестоимости в целом по обоим предприятиям с помощью индексов переменного и фиксированного состава; 3) индекс структурных сдвигов.

Покажите связь между исчисленными индексами. Сделайте выводы.

Задание 4. Имеются следующие данные о реализации некоторых товаров за два периода (табл. 8.10).

Таблица 8.10 Реализация товаров за два периода

Наименование товаров	Продажа в фактических ценах, млрд руб.		Изменение цен в отчетном периоде по сравнению с базисным периодом, %
	Базисный период	Отчетный период	
№ 1	155	210	+60
№ 2	180	255	+45
№ 3	270	340	+15

Рассчитайте:

1) индивидуальные и общие индексы цен;

2) общий индекс товарооборота в фактических ценах;

3) сумму экономического эффекта, полученную в результате изменения цен.

Задание 5. По предприятию имеются следующие данные о реализации продукции (табл. 8.11).

Таблица 8.11 Данные о реализации двух видов товаров

Вид продукции	Реализовано, тыс. ед.		Общая стоимость реализованной продукции, тыс. руб.	
	Базисный период	Отчетный период	Базисный период	Отчетный период
№ 1	48	79	580	675
№ 2	35	67	445	565

Рассчитайте:

1) среднее изменение цен на реализованную продукцию;

2) абсолютное изменение стоимости реализованной продукции за счет изменения цен;

3) общее изменение физического объема реализованной продукции предприятия;

4) абсолютное изменение стоимости реализованной продукции за счет изменения ее физического объема.

Задание 6. Имеются следующие данные о продаже картофеля по двум рынкам города (табл. 8.12).

Таблица 8.12 Реализация картофеля на рынках города

Рынок	Цена за 1 кг, руб.		Продано картофеля, т	
	Базисный период	Отчетный период	Базисный период	Отчетный период
№ 1	20	24	200	245
№ 2	25	27	120	166
Итого	45	51	320	411

Рассчитайте:

1) индивидуальные и общие индексы цен;

2) индивидуальные и общие индексы физического объема;

3) индекс цен переменного состава; 4) индекс цен фиксированного состава;

5) индекс влияния структурных сдвигов. Показать связь между исчисленными индексами. Сделать выводы.

Контрольные вопросы

1. Что в статистике называется индексом?
2. Какие задачи решают при помощи индексного анализа?
3. Какие бывают формы индексов?
4. Что характеризуют индивидуальные индексы?
5. Что показывают общие (групповые) индексы?
6. Что характеризует индекс структурных сдвигов и как он исчисляется?

7. Как исчисляется агрегатный индекс стоимости продукции и что он характеризует?

8. Что показывает индекс цен (Паше и Ласпейреса)?

9. Когда возникает необходимость преобразования агрегатного индекса цен в средний гармонический и средний арифметический индекс?

10. Что показывает индекс физического объема?

Библиографический список

1. Шмойлова Р.А. Теория статистики: учеб. пособие для вузов / Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова, Е.Б. Шувалова. – М.: Финансы и статистика, 2005. – 656 с.

2. Елисеева И.И. Общая теория статистики: учеб. пособие для вузов / И.И. Елисеева, М.М. Юзбашев. – М.: Финансы и статистика, 2004. – 656 с.

3. Ефимова М.Р. Практикум по общей теории статистики / М.Р. Ефимова, О.И. Ганченко, Е.В. Петрова. – М.: Финансы и статистика, 2008. – 368 с.

4. Макарова Н.В. Статистика в Excel: учеб. пособие / Н.В. Макарова, В.Я. Трофимец. – М.: Финансы и статистика, 2003. – 386 с.

5. Боровиков В.П. Statistika. Искусство анализа данных на компьютере. Для профессионалов / В.П. Боровиков. – СПб: Питер, 2001. – 656 с.

6. Шмойлова Р.А. Практикум по теории статистики / Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова. – М.: Финансы и статистика, 2007. – 416 с.

7. Филимонова Т.К. Общая теория статистики: учеб. пособие / Т.К. Филимонова, Н.Н. Шиманская. – Казань: Казан. гос. энерг. ун-т, 2006. – 99 с. 8. Салин В.Н. Курс теории статистики / В.Н. Салин, Э.Ю. Чурилова. – М.: Финансы и статистика, 2007. – 480 с.

9. Салин В.Н. Социально-экономическая статистика: учебник / В.Н. Салин, Е.П. Шпаковская. – М.: Юристъ, 2004. – 461 с.

10. Шиманская Н.Н. Применение статистических методов для решения социально-экономических задач. Ч. 1: учеб. пособие / Н.Н. Шиманская. – Казань: Казан. гос. энерг. ун-т, 2008. – 91 с.

Приложение

Сводка и группировка статистических данных

По данным таблицы произведите группировку (в зависимости от вашего варианта):

1. 38 коммерческих банков по величине прибыли (в млрд руб.).
2. 38 коммерческих банков по величине кредитных вложений (в млрд руб.).
3. 38 коммерческих банков по величине объема вложений в ценные бумаги (в млрд руб.).
4. 38 коммерческих банков по величине вкладов граждан (в млрд руб.).
5. 38 коммерческих банков по величине капитала (в млрд руб.).
6. 38 коммерческих банков по величине чистых активов (в млрд руб.). 7. 38 коммерческих банков по величине средств на корсчете в ЦБ (в млрд руб.).
8. 37 предприятий по величине стоимости основных производственных фондов (в млрд руб.).
9. 37 предприятий по величине реализованной товарной продукции (в млрд руб.).
10. 37 предприятий по величине среднесписочной численности производственного персонала (чел.).
11. 37 предприятий по величине себестоимости продукции (в млрд руб.).
12. 37 предприятий по величине прибыли от реализации (в млрд руб.).
13. 36 коммерческих банков по величине прибыли (в млрд руб.).
14. 36 коммерческих банков по величине кредитных вложений (в млрд руб.).
15. 36 коммерческих банков по величине объема вложений в ценные бумаги (в млрд руб.).
16. 36 коммерческих банков по величине вкладов граждан (в млрд руб.).
17. 36 коммерческих банков по величине капитала (в млрд руб.).
18. 36 коммерческих банков по величине чистых активов (в млрд руб.). 19. 36 коммерческих банков по величине средств на корсчете в ЦБ (в млрд руб.).
20. 38 предприятий по величине стоимости основных производственных фондов (в млрд руб.).
21. 38 предприятий по величине реализованной товарной продукции (в млрд руб.).
22. 38 предприятий по величине среднесписочной численности производственного персонала (чел.).
23. 38 предприятий по величине себестоимости продукции (в млрд руб.).

24. 38 предприятий по величине прибыли от реализации (в млрд руб.).

25. 38 коммерческих банков по величине кредитных вложений (в млрд руб.).

Таблица исходных данных

№	Варианты									
	1	2	3	4	5	6	7	8	9	10
1	118	618	266	584	870	1300	709	12,3	11,5	428
2	35	1303	574	1030	1510	2609	377	17,8	16,3	935
3	112	589	160	339	340	1504	310	16,7	19,40	537
4	185	298	591	881	229	1807	511	21,2	12,5	390
5	27	2330	238	1032	658	1086	392	11,3	14,6	814
6	147	626	377	2285	399	789	208	13,7	4,7	373
7	119	932	686	643	348	905	511	11,0	13,4	230
8	11	718	539	386	127	507	399	8,4	9,4	678
9	137	819	288	484	247	710	198	15,3	4,1	636
10	128	519	105	213	799	1053	210	18,5	11,5	580
11	133	419	422	799	249	618	604	12,7	5,3	216
12	56	981	352	326	867	1123	307	21,8	6,7	912
13	46	560	376	165	495	975	411	18,5	9,8	683
14	92	440	646	146	150	504	308	14,1	12,5	342
15	120	762	144	227	970	1106	205	29,0	7,4	280
16	114	944	353	275	400	809	690	13,1	13,1	336
17	152	1594	536	320	310	605	308	14,8	17,7	634
18	63	452	461	216	228	519	109	7,8	11,4	499
19	174	367	398	799	1320	1606	406	16,9	12,2	125
20	132	345	236	507	256	407	211	14,9	6,7	438
21	77	446	398	1045	1300	1102	406	22,1	9,3	244
22	125	579	674	175	500	864	315	18,4	4,7	620
23	96	381	113	294	137	616	914	17,3	6,9	672
24	84	454	587	194	169	502	112	15,4	14,6	760
25	97	654	224	127	421	891	713	12,8	7,9	449
26	110	852	105	161	752	1103	344	15,3	12,5	762

27	92	757	269	836	949	1262	212	19,9	12,0	886
28	153	638	138	1570	1260	1984	414	4,6	13,1	722
29	134	448	113	955	408	609	835	21,6	2,4	477
30	119	341	476	176	450	863	805	6,3	8,2	279
31	124	404	280	449	650	1042	274	4,7	5,8	737
32	97	740	103	163	200	813	405	6,5	9,2	465
33	111	346	108	598	430	954	716	7,2	6,7	708
34	127	368	791	186	731	1054	217	4,6	9,3	477
35	86	540	744	149	272	414	438	7,1	13,5	624
36	63	443	624	102	430	848	179	8,5	13,9	345
37	91	1008	476	115	321	734	612	6,9	6,1	529
38	114	823	328	643	702	1010	302	–	–	–

Продолжение табл.

№	Варианты									
	11	12	13	14	15	16	17	18	19	20
1	12,3	6,4	160	3,8	26,7	16,0	194,8	291,0	132,3	32,6
2	25,4	3,7	259	3,7	17,9	25,8	78,3	346,5	131,2	22,4
3	10,6	6,5	156	9,3	19,6	10,8	158,8	274,4	222,0	13,2
4	13,5	2,4	420	2,4	14,5	12,4	68,9	175,2	232,1	24,5
5	11,0	4,2	218	5,0	21,3	11,0	229,0	359,5	152,7	34,7
6	23,4	3,3	516	8,9	14,4	15,0	218,5	252,9	243,5	15,6
7	25,1	1,6	397	2,8	16,0	11,7	88,4	246,5	173,7	31,2
8	10,0	5,4	219	2,8	13,6	11,7	139,9	383,0	231,8	20,9
9	38,4	7,9	129	9,5	15,8	24,4	69,0	191,8	119,8	13,9
10	25,7	3,3	293	4,7	25,3	25,8	249,4	191,7	212,2	21,2
11	23,3	1,4	459	2,8	17,6	26,0	198,8	383,2	312,3	34,0
12	29,4	3,7	610	5,0	17,8	25,7	188,2	244,9	331,9	23,6
13	25,6	1,1	222	3,4	15,8	17,7	148,3	187,6	322,2	42,5
14	47,0	2,6	125	5,9	35,5	14,8	228,7	178,7	242,4	21,2
15	13,7	3,7	330	8,7	24,9	18,3	158,9	281,8	162,1	80,5
16	28,2	4,4	190	3,4	10,7	15,6	238,5	378,9	142,8	42,0
17	33,4	1,6	357	3,0	21,3	12,5	178,6	179,0	272,1	52,1
18	26,0	2,3	447	6,4	26,3	16,4	169,2	410,5	252,7	31,0
19	19,4	4,4	256	2,3	23,0	13,9	248,7	381,3	242,5	62,3

20	13,2	3,6	242	3,4	28,6	18,3	178,5	180,5	152,6	71,5
21	43,0	2,8	640	8,1	24,0	21,0	198,4	276,7	212,2	53,0
22	35,8	1,9	610	5,0	25,4	19,2	88,1	175,2	222,1	43,7
23	24,9	1,4	330	8,8	39,9	16,4	138,8	277,7	172,1	65,0
24	13,0	3,7	201	2,1	30,2	17,7	128,5	275,5	162,1	90,1
25	16,4	1,5	120	3,0	8,0	18,4	218,4	176,6	262,1	13,0
26	25,6	1,4	170	2,7	28,2	21,6	228,0	174,4	133,7	58,1
27	32,7	1,3	503	7,7	30,8	26,0	259,0	259,5	122,8	45,5
28	23,6	4,6	201	5,4	43,5	24,0	148,5	260,6	212,9	36,5
29	12,1	3,1	480	6,3	31,4	14,7	239,3	270,0	213,0	66,9
30	26,3	1,4	352	2,8	24,6	15,4	128,6	171,2	152,9	74,2
31	15,8	2,8	565	7,4	32,6	23,4	248,4	252,9	273,0	13,5
32	14,9	3,9	260	8,3	30,2	24,4	138,8	346,5	162,3	47,3
33	22,7	1,2	688	9,5	31,4	26,4	258,7	160,0	232,9	36,4
34	17,5	1,3	387	6,3	29,4	18,5	339,6	260,0	222,7	30,2
35	18,2	1,1	574	4,8	23,4	11,4	349,2	250,9	212,0	43,9
36	23,2	4,2	734	8,5	25,3	23,7	255,3	401,4	113,6	63,0
37	12,7	5,1	–	–	–	–	–	–	–	83,9
38	–	–	–	–	–	–	–	–	–	59,3

Окончание табл.

№	Варианты				
	21	22	23	24	25
1	27,2	420	11,4	10,2	23,7
2	19,1	290	25,6	10,6	14,8
3	14,4	254	21,4	20,3	32,2
4	12,9	328	48,5	10,6	25,8
5	25,5	276	32,6	30,3	11,6
6	26,0	429	14,0	20,8	12,0
7	12,7	324	14,0	30,6	39,1
8	39,8	431	29,0	31,3	24,4
9	24,8	268	33,1	21,1	31,4
10	12,9	529	41,0	40,5	45,8
11	15,0	725	32,3	31,9	39,5
12	15,0	428	12,5	10,6	12,8
13	36,1	232	14,0	10,3	10,6
14	32,4	630	44,2	40,6	16,8

15	24,6	624	51,6	30,3	25,8
16	28,3	327	49,3	20,4	19,2
17	24,9	373	54,3	30,9	37,6
18	49,0	535	33,6	41,3	49,2
19	44,0	323	32,5	31,1	51,8
20	34,1	422	44,5	50,5	31,5
21	22,0	230	54,2	31,3	43,6
22	36,9	278	61,8	41,2	22,2
23	17,7	522	11,3	21,2	10,5
24	49,1	634	15,2	20,9	9,8
25	32,0	238	12,3	30,4	12,5
26	44,0	626	58,3	40,8	11,4
27	36,3	429	52,8	10,3	42,7
28	38,6	328	41,7	30,8	35,3
29	21,4	238	32,5	20,9	21,4
30	22,1	718	24,9	31,4	47,8
31	13,0	617	34,0	20,3	32,8
32	13,2	437	12,3	31,5	29,9
33	13,4	286	37,5	40,5	10,4
34	34,2	527	23,7	30,2	13,8
35	22,8	722	44,0	50,4	31,3
36	44,0	323	51,4	50,3	23,6
37	32,5	623	14,5	40,7	11,5
38	24,3	245	34,8	31,3	38,5