

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
Белорусский национальный технический университет

---

Кафедра «Инженерная экономика»

Л. В. Бутор  
Т. А. Сахнович

# СТАТИСТИКА

Учебно-методическое пособие  
для направления специальности 1-27 01 01-01  
«Экономика и организация производства (машиностроение)»  
в рамках специальности 1-27 01 01 «Экономика  
и организация производства (по направлениям)»

*Рекомендовано учебно-методическим объединением по образованию  
в области экономики и организации производства*

Минск  
БНТУ  
2022

УДК 311:338.45  
ББК 65.051я7  
Б93

**Р е ц е н з е н т ы:**  
кафедра «Банковская экономика» БГУ  
(зав. каф., канд. экон. наук, доцент *А. И. Короткевич*);  
доцент кафедры «Менеджмент» БГУИР, канд. экон. наук,  
доцент *И. В. Насонова*

**Бутор, Л. В.**  
Б93      Статистика : учебно-методическое пособие для направления специальности 1-27 01 01-01 «Экономика и организация производства (машиностроение)» в рамках специальности 1-27 01 01 «Экономика и организация производства (по направлениям)» / Л. В. Бутор, Т. А. Сахнович. – Минск : БНТУ, 2022. – 52 с.  
ISBN 978-985-583-472-5.

Учебно-методическое пособие состоит из 14 тем, соответствующих темам курса общей статистики. Пособие предназначено для закрепления теоретических знаний по курсу «Статистика», а также для оценки знаний, полученных в ходе изучения предмета.

**УДК 311:338.45**  
**ББК 65.051я7**

**ISBN 978-985-583-472-5**

© Бутор Л. В., Сахнович Т. А., 2022  
© Белорусский национальный  
технический университет, 2022

## СОДЕРЖАНИЕ

1. Введение в статистику. Определение, термины и понятия статистики как науки.....	4
2. Разработка и реализация статистического наблюдения.....	6
3. Сводка и группировка статистических данных .....	9
4. Статистические таблицы и графики.....	13
5. Абсолютные и относительные величины .....	16
6. Ряды распределения и их показатели.....	19
7. Характеристика формы распределения .....	25
8. Изучение взаимосвязи между явлениями.....	29
9. Стохастическое моделирование .....	32
10. Основы кластерного анализа .....	38
11. Ряды динамики .....	43
12. Выборочное наблюдение.....	46
13. Индексы .....	48
14. Работа в статистических пакетах.....	50
Список литературы .....	51

# 1. ВВЕДЕНИЕ В СТАТИСТИКУ. ОПРЕДЕЛЕНИЕ, ТЕРМИНЫ И ПОНЯТИЯ СТАТИСТИКИ КАК НАУКИ

## Теоретические положения

*Статистика* – отрасль общественной науки, которая изучает количественную сторону качественно определенных массовых социально-экономических явлений и процессов, их структуру и распределение, размещение в пространстве, движение во времени, выявляя действующие количественные зависимости, тенденции и закономерности, причем в конкретных условиях места и времени.

Статистика – это многогранный инструмент. Статистика в бизнесе помогает экономить деньги и открывать новые рынки.

*Предмет познания* статистики – совокупность определенных явлений, предметов, вещей и их характерных свойств.

*Предметом изучения* статистики является количественная определенность, присущая общественным явлениям.

При изучении массовых общественных явлений используют следующие статистические термины:

1. *Признак* – свойство, характерная особенность единицы объекта, который можно наблюдать или измерить. Признаки делятся на качественные и количественные.

2. *Вариация* – многообразие изменений значений признаков.

3. *Статистическая совокупность* – это множество объектов или явлений, изучаемых статистикой, которые имеют один или несколько общих признаков и различаются по другим признакам.

4. *Показатель* – это обобщенная количественная характеристика явления. Величина показателей устанавливается в результате расчетов.

5. *Система показателей* – это группа показателей для характеристики одного явления или процесса.

## **Контрольные вопросы**

1. Назовите предмет и методы статистики.
2. Назовите основные черты статистики.
3. Назовите основные статистические термины.

## 2. РАЗРАБОТКА И РЕАЛИЗАЦИЯ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

### Теоретические положения

Для исследования социально-экономических явлений и процессов общественной жизни следует собрать о них необходимые статистические данные.

*Статистические данные* – совокупность количественных характеристик социально-экономических явлений и процессов, полученных в результате статистического наблюдения, их обработки и соответствующих расчетов.

*Статистическое наблюдение* – это массовое планомерное, научно организованное наблюдение за явлениями социальной жизни, которое заключается в регистрации отобранных признаков у каждой единицы совокупности.

*Цель наблюдения* – получение достоверной информации для выявления закономерностей развития явлений и процессов. При подготовке наблюдения следует четко определить, что именно подлежит исследованию, т. е. установить объект наблюдения.

Под *объектом наблюдения* понимается некоторая статистическая совокупность, в которой протекают исследуемые социально-экономические явления и процессы.

*Программа наблюдения* – перечень признаков, подлежащих регистрации в процессе наблюдения.

*Объективным временем (критический момент)* называется дата, в которую регистрируются сведения.

*Субъективным временем* называется период времени, в течении которого проводится наблюдение.

К *организационным формам* статистического наблюдения относятся статическая отчетность, специализированное наблюдение, регистры.

*Статистическая отчетность* – основная форма статистического наблюдения, с помощью которой статистические ор-

ганы получают от предприятий необходимые данные в виде установленных в законном порядке отчетных документов, скрепляемых подписями лиц, ответственных за предоставление сведений.

По срокам представления отчетность делят на ежедневную, декадную, месячную, квартальную, годовую.

Наблюдение может быть *специально организованным* (перепись), и *регистрационным*.

*Регистровое наблюдение* – это форма непрерывного статистического наблюдения за долговременными процессами, имеющего фиксированное начало, стадию развития и фиксированный конец.

*Непосредственным* называется такое наблюдение, при котором сами регистраторы путем замера, подсчета, взвешивания устанавливают факт подлежащий регистрации.

*Документальный способ* наблюдения основан на использовании данных статистической отчетности для получения необходимых сведений.

*Опрос* – это способ наблюдения, при котором необходимые сведения получают со слов респондента.

Задачей *сплошного наблюдения* является получение информации о всех единицах совокупности (трудоемко и дорого).

*Не сплошное наблюдение* изначально предполагает, что обследованию подлежит лишь часть единиц изучаемой совокупности. При его проведении следует заранее определить, какая часть единиц изучаемой совокупности подлежит обследованию.

*Точностью статистического наблюдения* называют степень соответствия величины какого-либо показателя (значение какого-либо признака), определенной по материалам статистического наблюдения, действительной его величине.

Расхождение между расчетным и действительным значением изучаемых величин называется *ошибкой наблюдения*.

*Ошибки регистрации* – это отклонения между значением показателя, полученного в ходе статистического наблюдения, и фактическим, действительным, его значением.

Отклонение значения показателя обследованной совокупности от его величины по исходной совокупности называется *ошибкой репрезентативности*.

### **Контрольные вопросы**

1. Что понимается под статистическим наблюдением?
2. Назовите виды статистического наблюдения.
3. Перечислите ошибки статистического наблюдения.



### 3. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

#### Теоретические положения

На основании информации, собранной в результате наблюдения, нельзя непосредственно охарактеризовать закономерности в развитии явления. Так как информация собирается по каждой единице объекта, то полученные данные не являются обобщающими.

Поэтому целью следующего этапа исследования является систематизация данных, собранных при наблюдении.

*Сводка* – второй этап статистического исследования. Она включает в себя действия по упорядочению и обработке первичного материала с целью выявления типичных черт и закономерностей изучаемых явлений и процессов.

*Простая сводка* – это операции по подсчету общих итогов.

*Сложная сводка* – это комплекс операций по группировке, подсчету итогов по каждой группе и по всему явлению с представлением результатов в виде таблицы.

*Группировка* представляет собой распределение результатов сводки на однородные группы по существенным признакам и определенным правилам. Если группировка ведется на основании одного признака, она называется *простой*. Если на основании нескольких признаков – *комбинационной* (возраст, стаж, профессия).

Важнейшими *видами* статистических группировок являются следующие: типологическая, структурная, аналитическая.

*Типологическая* группировка – это распределение качественно разнородной совокупности на социально-экономические группы, классы (группировка предприятий по формам собственности).

*Структурной* называется группировка, в которой происходит распределение однородной совокупности на группы, характеризующие ее структуру (население – городское, сельское).

*Аналитическая* группировка выявляет связь между явлениями и их признаками (производительность труда – стаж, квалификация). В основу группировки положен факторный признак, а каждая группа характеризуется средним значением результативного признака.

*Факторные признаки* – это признаки, под действием которых изменяются другие признаки – *результативные*.

Группировки могут быть построены по *одному* или *двум* признакам. Признаки делятся на качественные или количественные.

Группировки по *качественным* признакам содержат столько групп, сколько разновидностей группировочного признака существует.

*Группировочным* называется признак положенный в основу группировки.

Число повторений признака в группе называется *частотой*. Кроме частот применяются *частоты*, которые определяются отношением каждой частоты к сумме частот.

Группировки по *количественным* признакам подразделяются на дискретные, с равными интервалами, с интервалами разной величины.

В *дискретных* группировках число разновидностей группировочного признака не должно быть более 10. В этих группировках группой является каждое значение.

Для группировок с *равновеликими интервалами* величина интервала определяется по формуле (3.1):

$$h = \frac{x_{\text{макс}} - x_{\text{мин}}}{K}, \quad (3.1)$$

где  $h$  – величина интервала;

$x_{\text{макс}}$  – максимальное значение признака;

$x_{\text{мин}}$  – минимальное значение признака;

$K$  – количество групп, которое можно определить по формулам (3.2), (3.3):

$$K = 2 \ln n, \quad (3.2)$$

$$K = 1 + 3,322 \lg n, \quad (3.3)$$

где  $n$  – число анализируемых признаков (объем совокупности).

Количество групп можно определить из показателя среднего квадратического отклонения  $\sigma$  (3.4). Если величина интервала равна  $0,5\sigma$ , то совокупность разбивается на 12 групп, если  $1\sigma$ , то на 6 групп, если  $2/3\sigma$ , то на 9 групп.

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}, \quad (3.4)$$

$$\bar{x} = \frac{\sum x_i}{n}, \quad (3.5)$$

где  $x$  – значение каждого признака.

Если число признаков не более 200, то количество интервалов менее 10 и количество групп должно быть не менее трех.

Полученную величину интервала необходимо округлить до ближайшего удобного числа (заканчивающегося на цифру 5 или 10).

Нижнюю границу первого интервала определяют как разницу между минимальным значением признака и половиной рассчитанного интервала. Полученное значение округляется до ближайшего удобного числа.

Нижней границей интервала называется наименьшее значение признака в интервале ( $x_{\text{ниж}}$ ).

Верхняя граница интервала определяется прибавлением к нижней границе величины интервала:

$$x_{\text{верхн}} = x_{\text{нижн}} + h. \quad (6)$$

Отсюда величина интервала определяется как разность между верхней и нижней границей.

Группировка считается построенной правильно, если в ней нет пустых и малочисленных групп.

### **Контрольные вопросы**

1. Что понимается под статистической группировкой?
2. Что понимается под статистической сводкой?
3. Каким образом определяется величина интервала при группировке признака и как она округляется?
4. Какие существуют методы образования группировок?

## 4. СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ И ГРАФИКИ

### Теоретические положения

*Статистическая таблица* представляет собой форму наиболее рационального, наглядного и систематизированного изложения числовых результатов сводки и группировки.

Статистическую таблицу от других табличных форм отличает:

- 1) наличие результатов подсчета эмпирических данных;
- 2) должна быть итогом сводки и группировки.

Таким образом, *статистической* называется таблица, которая содержит сводную числовую характеристику исследуемой совокупности по одному или нескольким существенным взаимосвязанным признакам.

К элементами статистической таблицы относятся:

- 1) статистическое подлежащее;
- 2) статистическое сказуемое.

*Статистическое подлежащее* таблицы – это та статистическая совокупность, которая характеризуется числовыми показателями (обычно содержание строк).

*Статистическое сказуемое* таблицы – это показатели, которые характеризуют подлежащее (цифровая определенность граф).

В зависимости от строения подлежащего статистические таблицы делятся на *простые, групповые, комбинационные*.

При построении таблиц пользуются следующими правилами:

- 1) заглавие таблицы, строк и граф должно формулироваться кратко и полностью характеризовать содержание таблицы;
- 2) таблица должна быть небольшой, компактной;
- 3) в случае отсутствия факта в таблице ставится прочерк;
- 4) таблица, как правило, должна иметь итоговые результаты.

Анализ статистической таблицы включает:

- а) структурный анализ, который предполагает анализ строения таблицы; характеристику представленных в таблице совокупностей;

б) содержательный анализ предполагает изучение внутреннего содержания таблицы; групп подлежащего и элементов сказуемого;

в) логическая проверка заключается в сопоставлении признаков;

г) счетная проверка предполагает выборочный расчет отдельных показателей.

*Графиками* в статистике называют условные изображения числовых величин их соотношений в виде различных геометрических знаков. Главное достоинство графиков – наглядность.

К основным элементам графиков относят графический образ, поле графика, пространственные ориентиры, масштабные ориентиры, экспликация.

Статистические диаграммы подразделяются на *сравнения*, *структурные* и *динамики*.

Диаграммы сравнения (сопоставления) дают наглядную сравнительную характеристику статистических совокупностей по одному варьирующему признаку.

Структурные диаграммы представляют статистическую совокупность соотношением ее частей в абсолютном и относительном выражении.

Диаграммы динамики строятся для анализа развития явления во времени. Для графического изображения рядов динамики используют столбиковые, ленточные, круговые, линейные, радиальные.

*Линейные диаграммы* строятся на системе прямоугольных координат. По оси абсцисс откладывают время, по оси ординат – размеры явлений (размеры признаков).

*Радиальные диаграммы* строятся в полярных координатах. Они разделяются на *замкнутые* и *спиральные* и характеризуют циклически развивающиеся явления. В качестве центра отсчета выбирается центр круга или окружность.

*Статистические карты* представляют собой вид графических изображении статистических данных на схеме географиче-

ческой карты. Они характеризуют степень распространения явления на определенной территории.

### **Контрольные вопросы**

1. Назовите общие понятия и элементы статистических таблиц.
2. Назовите виды статистических таблиц.
3. Сущность, значение и основные виды графиков.
4. Что представляют собой статистические карты?

## 5. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

### Теоретические положения

*Конкретный статистический показатель* характеризует величину признака в конкретное время и данном месте.

*Абсолютные показатели* являются первичной формой статистических показателей. Они подразделяются на индивидуальные и общие.

*Индивидуальные абсолютные* показатели получают в результате статистического наблюдения путем подсчета, измерения.

*Общие (сводные) объемные* показатели получают в результате сводки и группировки статистических данных, которые представляют собой общую величину всего объекта или его части.

Абсолютные показатели являются именованными величинами. Они измеряются в следующих единицах измерения:

1. *Натуральные единицы измерения*, которые выражают признак в физических мерах (кг, тонны, метры).

2. *Условно-натуральные единицы* применяются в том случае, если необходимо свести в единый итог разнородную продукцию, имеющую одинаковые потребительные свойства.

3. *Стоимостные абсолютные показатели* дают денежную оценку явлению.

4. *Трудовые единицы измерения абсолютных величин* позволяют учитывать трудоемкость как отдельных видов продукции, так и объема всего выпуска продукции.

5. *Комбинированные абсолютные показатели* содержат в единицах измерения несколько наименований абсолютных величин (кВт-час, человеко-час).

*Относительные величины* определяются соотношением двух абсолютных величин.

Измеряются относительные величины в локальных единицах (без размерных), в процентах, промилле, пропромилле и т. д.



Статистика выделяет следующие виды относительных величин:

- планового задания (ОВПЗ);
- выполнения плана (ОВВП);
- динамики (ОВД);
- структуры (ОВСтр);
- координации (ОВК);
- сравнения (ОВСр);
- интенсивности развития (ОВИР).

*Относительная величина планового задания* (ОВПЗ) определяется отношением плановой величины признака в отчетном периоде к его фактической величине в базисном периоде.

*Относительная величина выполнения плана* (ОВВП) определяется отношением фактической величины признака к его плановой величине в анализируемом периоде.

*Относительная величина динамики* (ОВД) определяется отношением фактического (планового) уровня данного периода к фактическому (плановому) уровню базисного периода.

Между этими тремя показателями существует взаимосвязь (5.1):

$$\text{ОВД} = \text{ОВПЗ} \cdot \text{ОВВП}. \quad (7)$$

*Относительная величина структуры* (ОВСтр) определяется отношением частей целого к общему итогу.

*Относительная величина координации* (ОВК) определяется соотношением между частями единого целого.

*Относительная величина сравнения* (ОВСр) определяется соотношением двух одинаковых признаков, принадлежащих к разным объектам в одном периоде времени.

*Относительная величина интенсивности развития* (ОВИР) определяется соотношением различных признаков, принадлежащих одному объекту в одном периоде времени. Она показывает степень распространения одного признака в другом. Единицы измерения ОВИР состоят из единиц измерения тех

абсолютных величин, которые используются в определении относительной величины.

### **Контрольные вопросы**

1. Назовите виды абсолютных и относительных величин.
2. В каких единицах измерения представляются абсолютные величины и относительные величины всех видов?

## 6. РЯДЫ РАСПРЕДЕЛЕНИЯ И ИХ ПОКАЗАТЕЛИ

### Теоретические положения

К рядам распределения относятся ряды, у которых признак выражен числом. Ряды распределения подразделяются на дискретные и интервальные.

*Дискретные ряды распределения* могут изображаться при помощи перпендикуляров, кумуляты, полигона частот.

*Интервальные ряды динамики* – при помощи гистограммы, кумуляты, полигонов частот.

*Плотность распределения* определяется для интервальных рядов распределения и подразделяется на абсолютную и относительную.

*Абсолютная плотность распределения* определяется отношением частоты признака к величине интервала.

*Относительная плотность распределения* определяется отношением частоты к величине интервала.

Методы, наиболее подходящие для анализа совокупности данных в статистике, включают определение следующих групп показателей: простые средние и степенные средние.

К *простым средним* относятся мода, медиана и средняя хронологическая величина.

1. *Мода* – это варианта с наибольшей частотой. Если в ряду несколько признаков имеют одинаковые наибольшие частоты, то ряд называется *бимодальным*.

В *интервальных* рядах распределения мода определяется по формуле (6.1):

$$M_o = x_{m_o} + \frac{f_{m_o} - f_{(m_o-1)}}{(f_{m_o} - f_{(m_o-1)}) + (f_{m_o} - f_{(m_o+1)})} h_{m_o}, \quad (6.1)$$

где  $f_{m_o}$  – частота модального интервала, т. е. интервала содержащего наибольшее число вариантов (наибольшую частоту);

$f_{(m_0-1)}$  – частота интервала, предшествующего модальному;

$f_{(m_0+1)}$  – частота интервала, следующего за модальным;

$h_{m_0}$  – длина модального интервала;

$x_{m_0}$  – нижняя граница модального интервала.

2. *Медиана* соответствует варианту, стоящей в середине ранжированного (упорядоченного) ряда.

Для *дискретных* рядов положение медианы определяется ее номером (6.2):

$$N_{Me} = \frac{n+1}{2}, \quad (6.2)$$

где  $n$  – объем ряда.

Для определения *медианы* в *интервальном* ряду необходимо пользоваться формулой

$$Me = x_{me} + \frac{\frac{\sum f_i}{2} - S_{(me-1)}}{f_{me}} h_{me},$$

где  $S_{(me-1)}$  – кумулятивная (накопленная) частота интервала, предшествующего медианному;

$x_{me}$  – нижняя граница медианного интервала;

$f_{me}$  – частота медианного интервала;

$h_{me}$  – длина медианного интервала.

*Медианный интервал* – первый интервал, накопленная частота которого превышает половину суммы частот.

3. *Средняя хронологическая* величина определяется по формуле (6.3):

$$\bar{x}_{\text{хрон}} = \frac{\frac{1}{2}x_1 + x_2 + \dots + \frac{1}{2}x_n}{n-1}. \quad (6.3)$$

Средняя хронологическая применяется в том случае, если признаки отстоят друг от друга по времени на равные даты.

*Степенные средние* определяются по формуле (6.4):

$$\bar{x} = \sqrt[m]{\frac{\sum x^m f}{\sum f}}, \quad (6.4)$$

где  $m$  – показатель степени.

В зависимости от показателя степени  $m$  образуются различные виды степенных средних:

– при  $m = 1$ , получаем среднюю *арифметическую* (6.5):

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}; \quad (6.5)$$

– при  $m = 2$ , получаем среднюю *квадратическую* (6.6):

$$\bar{x}_{\text{кв.}} = \sqrt[2]{\frac{\sum x^2 f}{\sum f}}; \quad (6.6)$$

– при  $m = 3$ , получаем среднюю *кубическую* (6.7):

$$\bar{x}_{\text{куб.}} = \sqrt[3]{\frac{\sum x^3 f}{\sum f}}; \quad (6.7)$$

– при  $m = 0$ , получаем среднюю *геометрическую* (6.8):

$$\bar{x}_{\text{геом}} = \sqrt[n-1]{\prod x^f}; \quad (6.8)$$

– при  $m = -1$ , получаем среднюю *хронологическую* (6.9):

$$\bar{x} = \frac{\sum f_i}{\sum \frac{1}{x_i} f_i}. \quad (6.9)$$

Наиболее широко из средних величин в экономике используется средняя арифметическая величина.

Определение средних величин для *дискретных* и *интервальных* рядов имеет особенность. В дискретных рядах признак берется сам по себе, а в интервальных – заменяется серединой интервала.

Мода, медиана и средние значения называются *мерами центральной тенденции*.

Для характеристики надежности средней арифметической определяются показатели вариации.

К *абсолютным показателям вариации* относят размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение.

*Размах вариации* признака определяется по формуле (6.10):

$$R = x_{\text{макс}} - x_{\text{мин}}, \quad (6.10)$$

где  $x_{\text{макс}}$  – максимальное значение признака;

$x_{\text{мин}}$  – минимальное значение признака.

*Среднее линейное отклонение* признака от средней арифметической определяется по формуле (6.11):

$$d = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}. \quad (6.11)$$

*Дисперсия количественного* признака определяется по формуле (6.12) или (6.13):

$$\sigma_{(x)}^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}; \quad (6.12)$$

$$\sigma_{(x)}^2 = \frac{\sum x^2 f_i}{\sum f_i} - (\bar{x})^2. \quad (6.13)$$

Дисперсия качественного признака находится по формуле (6.14):

$$\sigma_{(x)}^2 = w(w-1), \quad (6.14)$$

где  $w$  – удельный вес признака.

Среднеквадратическое отклонение признака определяется по формуле (6.15):

$$\sigma_{(x)} = \sqrt{\sigma_{(x)}^2}. \quad (6.15)$$

К относительным показателям вариации относят коэффициент осцилляции, относительное линейное отклонение и коэффициент вариации. Все они вычисляются как отношение абсолютных показателей вариации к средней арифметической и используются для сравнения различных признаков одной и той же совокупности или колеблемости одного и того же признака в различных совокупностях.

Коэффициент осцилляции отражает относительную колеблемость крайних значений признака вокруг средней и определяется по формуле (6.16):

$$V_{(R)} = \frac{R_{(x)}}{\bar{x}} 100 \%. \quad (6.16)$$

Коэффициент вариации по линейному отклонению (относительное линейное отклонение) определяется по формуле (6.17):

$$V_{(d)} = \frac{d_{(x)}}{\bar{x}} 100 \%. \quad (6.17)$$

Коэффициент вариации по среднему квадратическому отклонению определяется по формуле (6.18):

$$V_{(\sigma)} = \frac{\sigma_{(x)}}{\bar{x}} 100 \%. \quad (6.18)$$

Если значение этого показателя меньше 34 %, то ряд распределения равномерен.

Дисперсия определяется по формуле (6.19) либо (6.20):

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}; \quad (26)$$

$$\sigma^2 = \frac{\sum x^2 f}{\sum f} - (\bar{x})^2. \quad (6.19)$$

Среднее квадратическое отклонение – это корень квадратный из дисперсии (6.21):

$$\sigma = \sqrt{\sigma^2}. \quad (6.21)$$

На практике, обычно, отклонения не превышают  $3\sigma$ . Это положение называют *правилом трех сигм*.

### **Контрольные вопросы**

1. Назовите виды средних величин.
2. Какие показатели вариации средней арифметической можно рассчитать?
3. Методы определения степенной средней.



## 7. ХАРАКТЕРИСТИКА ФОРМЫ РАСПРЕДЕЛЕНИЯ

### Теоретические положения

Средняя арифметическая и дисперсия являются частными случаями более широкого понятия о моментах распределения. Различают начальные и центральные эмпирические моменты распределения.

*Начальным эмпирическим моментом  $q$  порядка* называется средняя арифметическая взвешенная  $q$  степеней признака (7.1):

$$\bar{V}_q = \frac{\sum x_i^q f_i}{\sum f_i}. \quad (7.1)$$

Начальные эмпирические моменты могут определяться в условных значениях, но в действительные они не переводятся. Исключение составляет начальный момент первого порядка.

*Центральным эмпирическим моментом  $q$  порядка* называется средняя арифметическая взвешенная  $q$  степеней отклонений вариант от средней арифметической:

$$\bar{\mu}_q = \frac{\sum (x_{(i)} - \bar{x})^q f_{(i)}}{\sum f_{(i)}}. \quad (7.2)$$

Центральные эмпирические моменты определяются в условных значениях и переводятся в действительные по формуле (7.3):

$$\bar{\mu}_q = \bar{\mu}'_q C^q. \quad (7.3)$$

Наиболее распространенными для анализа кривой нормального распределения являются *коэффициенты асимметрии* и *эксцесса*.

*Коэффициент асимметрии* характеризует симметричность кривой нормального распределения относительно средней арифметической и рассчитывается по формуле

$$A_s = \frac{\mu_3}{\sigma_{(x)}^3}, \quad (7.4)$$

где  $\mu_3$  – центральный момент третьего порядка, который определяется по формуле

$$\mu_3 = \frac{\sum (x - \bar{x})^3 f_i}{\sum f_i}. \quad (7.5)$$

Если коэффициент асимметрии:

- $A_s < 0$  – асимметрия левосторонняя;
- $A_s = 0$  – асимметрия отсутствует;
- $A_s > 0$  – асимметрия правосторонняя.

При сравнении нескольких рядов по их симметричности можно воспользоваться формулой (7.6) либо (7.7):

$$A_s = \frac{\bar{x} - Mo}{\sigma_{(x)}}, \quad (7.6)$$

$$A_s = \frac{\bar{x} - Mo}{\sigma_{(x)}^2}. \quad (7.7)$$

Оценка существенности асимметрии проводится с помощью среднеквадратической ошибки (7.8):

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}, \quad (7.8)$$

где  $n$  – число наблюдений.

В случае  $\frac{|A_s|}{\sigma_{A_s}} \geq 3$ , асимметрия существенна.

*Коэффициент эксцесса* отражает форму вершины кривой нормального распределения и рассчитывается по формуле (7.9):

$$E_s = \frac{\mu_4}{\sigma_{(x)}^4} - 3, \quad (7.9)$$

где  $\mu_4$  – центральный момент четвертого порядка, который определяется по формуле (7.10):

$$\mu_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}. \quad (7.10)$$

Если коэффициент эксцесса:

- $E_s < 0$  – кривая нормального распределения имеет плоскую вершину;
- $E_s = 0$  – кривая нормального распределения идеальная;
- $E_s > 0$  – кривая нормального распределения имеет острую вершину.

В случае  $\frac{|E_s|}{\sigma_{E_s}} \geq 3$ , эксцесс существенен.

Средняя квадратическая ошибка эксцесса определяется по формуле (7.11):

$$\sigma_{E_s} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}. \quad (7.11)$$

Для проверки гипотезы о соответствии данной кривой типу кривых нормального распределения необходимо определить

*теоретические частоты* по формуле плотности нормального распределения (7.12), (7.13):

$$F = \frac{1}{\sigma_{(x)}\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x})^2}{2\sigma_{(x)}^2}}. \quad (7.12)$$

Для удобства расчета теоретических частот обозначим:

– через коэффициент доверия  $t = \frac{(x_i - \bar{x})}{\sigma_{(x)}}$ ;

– через функцию  $\varphi = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ .

$$F = \frac{1}{\sigma_{(x)}} \varphi. \quad (7.13)$$

Если суммы теоретических и эмпирических частот равны, то расчеты верны.

### **Контрольные вопросы**

1. Какие виды форм распределения существуют?
2. По каким показателям оценивается симметричное и асимметричное распределение?
3. Какой показатель характеризует форму вершины кривой нормального распределения?

## 8. ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ

### *Теоретические положения*

Связи между признаками делятся на функциональные и стохастические.

*Функциональные* предполагают соответствие определенному значению факторного признака – соответствующее значение результативного.

*Стохастические* предполагают соответствие определенному значению факторного признака – среднее значение результативного (корреляция).

Соответствие данного распределения нормальному проверяем по критерию Хи-квадрат (8.1):

$$X^2 = \sum \frac{(f_{\text{теор}} - f_i)^2}{f_{\text{теор}}}. \quad (8.1)$$

Квадрат в числителе этой формулы убирает знак, а знаменатель приводит Хи-квадрат в нужную размерность. Если теоретическая частота равна эмпирической, то, применив эту формулу, мы получим 0.

Расчетное значение критерия сравнивается с табличным, если оно меньше или равно табличному значению, значит, гипотеза о принадлежности кривой нормальному распределению верна.

При отсутствии табличного значения можно оценить гипотезу по критерию Романовского (8.2):

$$C = \frac{X^2 - m}{\sqrt{2m}}, \quad (8.2)$$

где  $m$  – число групп.

Если  $C < 3$ , то гипотеза о принадлежности кривой нормальному распределению верна.

Примером стохастической связи является корреляционная (регрессионная) связь.

Регрессионный анализ позволяет осуществлять прогнозирование будущих результатов и применяется в том случае, если признаки количественные. Он включает в себя определение аналитического выражения связи и измерение тесноты связи.

Для измерения тесноты связи при прямолинейной зависимости используется *коэффициент корреляции*

$$r_{(xy)} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_{(y)}\sigma_{(x)}}. \quad (8.3)$$

Если один признак количественный, а второй качественный, то для оценки тесноты связи между ними используется критерий Фишера:

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \quad (8.4)$$

где  $\sigma_1^2$  – факторная дисперсия на одну степень свободы:

$$\sigma_1^2 = \frac{\sum (\bar{x}_i - \bar{x}_0)^2 n_i}{K_1}, \quad (8.5)$$

где  $K_1$  – число степеней свободы:

$$K_1 = m - 1, \quad (8.6)$$

где  $m$  – число групп;

$\sigma_2^2$  – случайная дисперсия на одну степень свободы:

$$\sigma_2^2 = \frac{\sum (x_i - \bar{x}_i)^2 f_i}{K_2}, \quad (8.7)$$

где  $K_2$  – число степеней свободы:

$$K_2 = n - m. \quad (8.9)$$

Рассчитанный критерий сравнивается с табличным значением.

Если расчетное значение больше табличного, то это означает, что расхождение между дисперсиями существенно и факторы влияют друг на друга.

### **Контрольные вопросы**

1. Какие виды корреляционных зависимостей используются в статистике?
2. Каким образом производится выбор вида зависимости?
3. Какие показатели тесноты связи применяются для количественных и качественных признаков?

## 9. СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

### Теоретические положения

Наиболее широкое применение в практике экономико-статистических исследований получило стохастическое моделирование, которое является дополнением и углублением детерминированного анализа. В анализе финансово-хозяйственной деятельности стохастические модели используются, когда необходимо:

- 1) оценить влияние факторов, по которым невозможно построить жестко детерминированную модель;
- 2) изучить и измерить влияние факторов, которые невозможно включить в одну и ту же детерминированную модель;
- 3) выделить и оценить влияние сложных факторов, которые не могут быть выражены одним определенным количественным показателем.

Изучение стохастической взаимосвязи осуществляется посредством проведения корреляционного и регрессионного анализа.

В современных пакетах для обработки данных (Statistica, Microsoft Excel и др.) представлено большое число статистических, финансовых и инженерных функций. Некоторые из них являются встроенными, другие доступны только после установки *пакета анализа* – набор средств анализа данных, предназначенный для решения сложных статистических задач. Для анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры; анализ будет выполнен с помощью подходящей статистической макрофункции, а результат будет помещен в выходной диапазон. Другие средства позволяют представить результаты анализа в графическом виде.

*Стохастический* (от греч. “stochastikos” – умеющий угадывать) *анализ* – это метод решения широкого класса задач статистического оценивания. Он предполагает изучение массо-



вых данных путем построения моделей изменения показателей за счет факторов, не находящихся в функциональной связи, в прямой или обратной взаимозависимости.

Стохастический анализ направлен на изучение косвенных связей в случае невозможности определения цепи прямой связи. Его использование обусловлено тем, что в экономических исследованиях чаще встречаются стохастические зависимости, которые отличаются приближенностью и неопределенностью.

В стохастической модели отражаются средние значения количественных характеристик связей экономических показателей. Стохастическая связь между показателями проявляется только в массе наблюдения, когда влияние других факторов на результивный показатель сглаживается.

Какое-либо явление природы и общества не может быть осознанным и понятным без обоснования его связей с другими явлениями. Чтобы познать сущность явлений, необходимо раскрыть их взаимоотношения, количественно определить влияние тех или иных объективных и субъективных факторов.

Влияние этих факторов на уровень экономических показателей в экономике до недавнего времени определялось в основном с помощью метода статистических группировок. Соотношения признаков, выявленных в результате статистических группировок, отличаются от соотношений, которые имеют место при функциональных связях, когда каждому значению аргумента соответствует определенное значение функции. Метод статистических группировок позволяет установить только наличие связи между явлениями, не определяя при этом его сравнительные количественные параметры. Поэтому наряду с методом группировок, которые играют важную роль в экономико-статистических исследованиях, для решения подобных вопросов необходимо применять и другие методы, в частности, метод корреляции.

Термин «корреляция» впервые применил Ж. Кювье в труде «Лекции по сравнительной анатомии» (1800–1805 гг.). Начальные математические построения метода корреляции были да-

ны О. Браве в 1846 г. «Корреляция» (от лат. “correlation” – отношения), означает соотношение, соответствие предметов или понятий).

Корреляция используется для выявления зависимости между изучаемыми явлениями. *Корреляционной связью* называется такая статистическая связь, при которой различным значениям одной переменной соответствуют разные средние значения другой. Различают парную и множественную корреляцию.

*Парная корреляция* – это связь между двумя показателями, один из которых является факторным, а другой – результативным.

*Множественная корреляция* возникает от взаимодействия нескольких факторов с результативным показателем.

*Корреляционный анализ* – это метод установления связи и измерения ее тесноты между наблюдениями, которые можно считать случайными и выбранными из совокупности, распределенной по многомерному нормальному закону.

Регрессия изучает форму связи. Регрессия позволяет выявить аналитическую зависимость между явлениями, т. е. представить ее в виде уравнения (функции).

*Регрессионный анализ* – это метод установления аналитического выражения стохастической зависимости между исследуемыми признаками.

Таким образом, применение корреляционно-регрессионного анализа позволяет решить две задачи:

1. Выявить и изучить тесноту связи между исследуемыми показателями, т. е. устанавливается относительная степень зависимости результативного показателя от каждого фактора. Степень тесноты связи оценивают коэффициентами корреляции ( $R$ ), изменяющимися от 0 до 1. Малое значение коэффициента свидетельствует о слабой связи, а значение, близкое по величине к 1, характеризует очень сильную связь и часто позволяет предположить о наличии функциональной причинно-следственной связи. Качественные оценки тесноты связи с помощью коэффициента корреляции приведены в табл. 9.1.

Характеристика тесноты связи  
при различных значениях коэффициента корреляции

Коэффициент корреляции	Теснота связи
0,91–0,99	очень сильная
0,81–0,90	весьма сильная
0,65–0,80	сильная
0,45–0,64	умеренная
0,25–0,44	слабая
до 0,25	очень слабая

2. Изучается характер связи, т. е. количественно измеряется действие одного или нескольких факторов на изменение результативного показателя. В результате по значению коэффициента регрессии ( $b$ ) устанавливают, насколько изменится величина результативного показателя при изменении факторного показателя на единицу.

Долю влияния учтенных факторов, включенных в многофакторную корреляционную модель, на результативный показатель можно установить с помощью коэффициента множественной детерминации ( $R_2$ ).

Важную роль в корреляционно-регрессионном анализе для отражения характера изучаемой связи между факторами и результативным показателем играет построение уравнения регрессии, которое отражает модель зависимости результативного показателя ( $y$ ) от изменения факторов ( $x$ ).

В случае если связь всех факторных показателей с результативным показателем носит прямолинейный характер, то для записи этих зависимостей можно использовать линейную функцию.

Наиболее простым уравнением, которое устанавливает линейную зависимость между двумя показателями, является уравнение прямой:

$$y_x = a + bx,$$

где  $y_x$  – результативный показатель;

$x$  – факторный показатель;

$a$  – свободный член уравнения, который не зависит от фактора  $x$ ;

$b$  – коэффициент регрессии, который показывает влияние на результативный показатель изменения величины фактора на единицу его измерения.

Это уравнение описывает такую связь между двумя показателями, при которой с изменением факторного показателя на определенную величину наблюдается равномерное возрастание или убывание значений результативного показателя.

Уравнение множественной регрессии при линейной форме связи имеет вид:

$$y_x = a + b_1x_1 + b_2x_2 + \dots + b_nx_n,$$

где  $y_x$  – результативный показатель;

$a$  – свободный член уравнения, который не зависит от факторов  $x$ ;

$b_1, b_2, \dots, b_n$  – коэффициенты регрессии, которые показывают количественное влияние на результативный показатель изменения величины фактора на единицу его измерения;

$x_1, x_2, \dots, x_n$  – факторные показатели.

В многофакторном корреляционно-регрессионном анализе важным является отбор наиболее существенных факторов, оказывающих влияние на изучаемый результативный показатель.

При отборе факторов следует соблюдать следующие правила:

1) не включать в корреляционную модель факторы, связь которых с результативным показателем носит функциональный характер;

2) отбирать самые значимые факторы, которые оказывают решающее воздействие на результативный показатель;

3) для составления уравнения множественной регрессии необходимо, чтобы число наблюдений было в 5–6 раз больше числа отобранных факторов;

4) факторы, включаемые в корреляционно-регрессионную модель, должны быть количественно измеримы, т. е. иметь единицу измерения и отражаться в учете или в отчетности.

### **Контрольные вопросы**

1. Назовите и охарактеризуйте виды корреляции.
2. Какие задачи позволяет решить корреляционно-регрессионный анализ?
3. Что следует учитывать при отборе факторов для анализа?

## 10. ОСНОВЫ КЛАСТЕРНОГО АНАЛИЗА

### Теоретические положения

Термин *кластерный анализ* включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры, т. е. как развернуть таксономию.

Фактически, кластерный анализ является не столько обычным статистическим методом, сколько «набором» различных алгоритмов «распределения объектов по кластерам». Существует точка зрения, что в отличие от многих других статистических процедур, методы кластерного анализа используются в большинстве случаев тогда, когда вы не имеете каких-либо априорных гипотез относительно классов, но все еще находитесь в описательной стадии исследования. Следует понимать, что кластерный анализ определяет «наиболее возможно значимое решение». Поэтому проверка статистической значимости здесь неприменима.

Техника кластеризации применяется в самых разнообразных областях. Известны широкие применения кластерного анализа в маркетинговых исследованиях. Всякий раз, когда необходимо классифицировать «горы» информации к пригодным для дальнейшей обработки группам, кластерный анализ оказывается весьма полезным и эффективным.

Объединение, или метод древовидной кластеризации, используется при формировании кластеров несходства или расстояния между объектами. Эти расстояния могут определяться в одномерном или многомерном пространстве. Наиболее прямой путь вычисления расстояний между объектами в многомерном пространстве состоит в вычислении евклидовых расстояний. Если имеется двух- или трехмерное пространство, то эта мера является реальным геометрическим расстоянием между объектами в пространстве (будто расстояния между

объектами измерены рулеткой). Однако алгоритму объединения не важно, являются ли предоставленные для этого расстояния настоящими или некоторыми другими производными мерами расстояния, что более значимо для исследователя; задачей исследователей является подбор правильный метод для специфических применений.

*Евклидово расстояние* – это наиболее общий тип расстояния. Оно является геометрическим расстоянием в многомерном пространстве. Заметим, что евклидово расстояние (и его квадрат) вычисляется по исходным, а не по стандартизованным данным. Это обычный способ его вычисления, который имеет определенные преимущества (например, расстояние между двумя объектами не изменяется при введении в анализ нового объекта, который может оказаться выбросом). Тем не менее, на расстояние могут сильно влиять различия между осями, по координатам которых вычисляются эти расстояния. К примеру, если одна из осей измерена в сантиметрах, а вы потом переведете ее в миллиметры, умножая значения на 10, окончательное евклидово расстояние (или квадрат евклидова расстояния), вычисляемое по координатам, сильно изменится и, как следствие, результаты кластерного анализа могут сильно отличаться от предыдущих.

*Квадрат евклидова расстояния.* Это расстояние является просто средним разностей по координатам. В большинстве случаев эта мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида. Однако отметим, что для этой меры влияние отдельных больших разностей (выбросов) уменьшается, т. е. они не возводятся в квадрат).

*Расстояние Чебышева.* Это расстояние может оказаться полезным, когда желают определить два объекта как «различные», если они различаются по какой-либо одной координате (каким-либо одним измерением).

*Степенное расстояние.* Иногда желают прогрессивно увеличить или уменьшить вес, относящийся к размерности, для которой соответствующие объекты сильно отличаются. Это может быть достигнуто с использованием степенного расстояния.

*Процент несогласия.* Эта мера используется в тех случаях, когда данные являются категориальными.

На первом шаге, когда каждый объект представляет собой отдельный кластер, расстояния между этими объектами определяются выбранной мерой. Однако когда связываются вместе несколько объектов, возникает вопрос, как следует определить расстояния между кластерами? Другими словами, необходимо *правило объединения или связи для двух кластеров*. Здесь имеются различные возможности. Например, вы можете связать два кластера вместе, когда любые два объекта в двух кластерах ближе друг к другу, чем соответствующее расстояние связи. Таким образом, следует использовать «правило ближайшего соседа» для определения расстояния между кластерами; этот метод называется *методом одиночной связи*. Это правило строит «волоконистые» кластеры, т. е. кластеры, «сцепленные вместе» только отдельными элементами, случайно оказавшимися ближе остальных друг к другу. Как альтернативу можно использовать соседей в кластерах, которые находятся дальше всех остальных пар объектов друг от друга. Этот метод называется *методом полной связи*.

*Одиночная связь (метод ближайшего соседа).* Как было описано выше, в этом методе расстояние между двумя кластерами определяется расстоянием между двумя наиболее близкими объектами (ближайшими соседями) в различных кластерах. Это правило должно, в известном смысле, «нанализовать» объекты для формирования кластеров. Результирующие кластеры имеют тенденцию быть представленными длинными «цепочками».

*Полная связь (метод наиболее удаленных соседей).* В этом методе расстояния между кластерами определяются наибольшим расстоянием между любыми двумя объектами в различных кластерах, т. е. наиболее удаленными соседями. Если же кластеры имеют в некотором роде удлиненную форму или их естественный тип является «цепочечным», то этот метод непригоден.



*Невзвешенное попарное среднее.* В этом методе расстояние между двумя различными кластерами вычисляется как среднее расстояние между всеми парами объектов в них.

*Взвешенное попарное среднее.* Метод идентичен методу невзвешенного попарного среднего, за исключением того, что при вычислениях размер соответствующих кластеров, т. е. число объектов, содержащихся в них, используется в качестве весового коэффициента. Поэтому предлагаемый метод должен быть использован (чаще даже, чем предыдущий), когда предполагаются неравные размеры кластеров.

*Невзвешенный центроидный метод.* В этом методе расстояние между двумя кластерами определяется как расстояние между их центрами тяжести.

*Взвешенный центроидный метод (медиана).* Этот метод идентичен предыдущему, за исключением того, что при вычислениях используются веса для учета разницы между размерами кластеров, т. е. числами объектов в них. Поэтому, если имеются (или подозреваются) значительные отличия в размерах кластеров, этот метод оказывается предпочтительнее предыдущего.

*Метод Варда.* Этот метод отличается от всех других, поскольку он использует методы дисперсионного анализа для оценки расстояний между кластерами. Метод минимизирует сумму квадратов (SS) для любых двух (гипотетических) кластеров, которые могут быть сформированы на каждом шаге.

Кластеризация, как по наблюдениям, так и по переменным может привести к достаточно интересным результатам. Модуль «Кластерный анализ» в таких пакетах, как, например, Statistica, содержит эффективную *двухходовую процедуру объединения*, позволяющую проводить кластеризацию в обоих направлениях. Однако двухходовое объединение используется относительно редко в обстоятельствах, когда ожидается, что и наблюдения, и переменные одновременно вносят вклад в обнаружение осмысленных кластеров.

*Метод К средних.* Этот метод кластеризации существенно отличается от таких агломеративных методов, как объединение

(древовидная кластеризация) и двухходовое объединение. Предположим, уже имеются гипотезы относительно числа кластеров (по наблюдениям или по переменным). Можно указать системе образовать ровно три кластера так, чтобы они были настолько различны, насколько это возможно. Это именно тот тип задач, который решает алгоритм метода  $K$  средних. В общем случае метод  $K$  средних строит ровно  $K$  различных кластеров, расположенных на возможно больших расстояниях друг от друга.

С вычислительной точки зрения этот метод можно рассматривать, как дисперсионный анализ «наоборот». Программа начинает с  $K$  случайно выбранных кластеров, а затем изменяет принадлежность объектов к ним так, чтобы:

- 1) минимизировать изменчивость внутри кластеров;
- 2) максимизировать изменчивость между кластерами.

Данный способ аналогичен методу «дисперсионный анализ наоборот» в том смысле, что критерий значимости в дисперсионном анализе сравнивает межгрупповую изменчивость с внутригрупповой при проверке гипотезы о том, что средние в группах отличаются друг от друга.

### **Контрольные вопросы**

1. Что такое кластерный анализ?
2. Охарактеризуйте древовидную кластеризацию.
3. Как происходит проверка статистической значимости в кластерах?
4. Для чего применяется двухходовое объединение?

# 11. РЯДЫ ДИНАМИКИ

## Теоретические положения

Одной из важнейших задач статистики является изучение изменений анализируемых показателей во времени, т. е. их динамики. Эта задача решается при помощи анализа рядов динамики (или временных рядов).

*Ряд динамики* представляет собой ряд расположенных в хронологической последовательности числовых значений статистического показателя, характеризующих изменение общественных явлений во времени.

Построение и анализ рядов динамики позволяют выявить и измерить закономерности развития общественного явления во времени. Эти закономерности не проявляются четко на каждом конкретном уровне, а лишь в тенденции, в длительной динамике.

Для характеристики ряда динамики определяются следующие показатели:

1. *Абсолютный прирост* ( $\Delta y$ ) характеризует увеличение или уменьшение уровня ряда за определенный промежуток времени.

2. *Абсолютный прирост цепной* рассчитывается, как разница между последующими и предыдущими значениями.

3. *Абсолютный прирост базисный* рассчитывается, как разница между текущим значением ряда и базисным значением.

4. *Темп роста цепной* рассчитывается, как отношение последующего значения к предыдущему.

5. *Темп роста базисный* рассчитывается, как отношение каждого значения ряда к базисному.

6. *Темп прироста цепной* определяется, как отношение абсолютного прироста цепного к уровню предшествующего периода.

7. *Темп прироста базисный* определяется, как отношение абсолютного базисного прироста к уровню ряда, принятого за базу сравнения для данного ряда.

8. *Абсолютное значение одного процента прироста* рассчитывается, как отношение абсолютного прироста цепного к темпу прироста цепному, выраженному в процентах.

Анализ рядов динамики, в зависимости от целей исследования, проводится следующими методами:

1. *Сравнение рядов динамики* применяется при одновременном анализе двух и более рядов. Он показывает во сколько раз быстрее растут уровни одного ряда, по сравнению с другим. Для анализа определяются коэффициенты *опережения* и *ускорения*.

2. *Приведение рядов динамики к общему основанию* применяется в том случае, если сравниваются только относительные показатели. Для этого определяются базисные показатели, приводимые к единому году сравнения.

3. *Смыкание рядов динамики* применяется в том случае, если уровни за одни годы не сопоставимы с уровнями за другие.

4. *Выявления общей тенденции* в стремлении ряда к росту, стабильности, снижению.

5. *Приемом укрупнения периодов* пользуются в том случае, если необходимо выявить общую тенденцию динамики, переходя от суточных уровней к декадным; от декадных к месячным; от месячных к квартальным и т. д.

6. Анализ рядов динамики при помощи *скользящей средней* применяется в том случае, если по исходным данным трудно предположить вид зависимости.

7. Аналитическое выравнивание ряда динамики.

Метод определения признака на перспективу или в прошлом называется *экстраполяцией*. Метод определения значения признака внутри ряда называется *интерполяцией*.

Наиболее сложным при прогнозировании является вопрос о заблаговременности прогноза. На основании опыта предполагают, что прогноз следует проводить не более чем на 5 лет. Так как за более длительный период могут изменяться условия развития явления.

## **Контрольные вопросы**

1. Назовите основные элементы ряда динамики.
2. Как определяются абсолютные и относительные показатели ряда динамики?
3. Назовите методы анализа рядов динамики.
4. Интерполяция и экстраполяция в рядах динамики.

## 12. ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

### Теоретические положения

Под *выборочным наблюдением* понимается такое не сплошное наблюдение, при котором обследованию подвергаются единицы совокупности, выбранные случайным образом. Выборочное наблюдение позволяет по выбранной совокупности дать характеристику показателям генеральной совокупности.

К *видам выборочного наблюдения* относятся индивидуальный, групповой и комбинированный. При *индивидуальном* отборе в выборочную совокупность отбираются из генеральной совокупности отдельные единицы; при *групповом* отборе – группы единиц; при *комбинированном* – предполагается сочетание двух предыдущих видов отбора.

Выделяют следующие *способы выборочного наблюдения*:

- а) случайная выборка;
- б) механическая выборка;
- в) типическая выборка;
- г) серийная выборка;
- д) комбинированная выборка;
- е) малая выборка;
- ж) метод моментных наблюдений.

Случайный отбор может быть повторным и бесповторным.

При *бесповторном отборе* отобранная единица регистрируется и в генеральную совокупность не возвращается.

При *повторном отборе* – возвращается и у нее появляется шанс быть отобранной еще.

*Механический отбор* проводится в выборе единиц совокупности механически: каждая десятая (при 10 % отборе), двадцатая (при 5 % отборе), пятая (при 20 % отборе), четвертая (при 25 % отборе) и т. д.

Механический отбор удобен при неограниченной генеральной совокупности.

Под *типическим отбором* понимается отбор, при котором генеральная совокупность делится на группы по типическому признаку, а затем из каждой группы отбирают единицы. Различают отбор пропорциональный и непропорциональный.

При *пропорциональном отборе* из каждой группы отбирают одинаковый процент признаков. При *непропорциональном отборе* из каждой группы одинаковое количество признаков.

При *серийном (гнездовом) отборе* выбирают не отдельные единицы, а целые партии, серии, гнезда (группы студентов, бригады рабочих, партии продукции).

*Комбинированный отбор* предполагает использование нескольких видов отбора (серийная – механическая, гнездовая – повторная).

К малым выборкам относится отбор с объемом менее 30 единиц.

*Метод моментных наблюдений* – это особый вид выборочного наблюдения, при котором в определенные моменты времени фиксируется наличие определенных элементов процесса. Определяется при изучении рабочего времени.

### **Контрольные вопросы**

1. Что такое выборочное наблюдение?
2. Назовите способы выборочного наблюдения.
3. Охарактеризуйте способы выборочного наблюдения.

## 13. ИНДЕКСЫ

### Теоретические положения

Под *индексом* понимается относительный показатель, уровни которого являются экономическими величинами, которые не имеет смысла складывать. Индексы измеряются в тех же единицах, что и относительные величины. При определении индексов применяются следующие обозначения.

В статистике принято обозначать индивидуальные индексы «*i*» и общие индексы «*I*». Знак внизу справа от индексируемой величины означает период: «0» – базисный, «1» – отчетный. Для обозначения индексируемых величин используются следующие признаки:

1. По *степени охвата* явления индексы подразделяются на индивидуальные, общие и субиндексы (индексы не всей совокупности, а ее группы).

2. По *базе сравнения* динамические и территориальные. Первая группа отражает изменение по периодам времени, вторая по территориям.

3. По *виду весов* индексы бывают с переменными и постоянными весами.

4. В зависимости от *формы построения* различают индексы агрегатные и средние.

5. По *характеру объекта исследования* индексы подразделяются на индексы количественных (объем продукции) и качественных (выработка) показателей.

6. По *объекту исследования*: себестоимости и производительности.

7. По *составу явления*: переменного, постоянного состава и структурных изменений.

8. По *периоду исчисления*: годовые, квартальные, месячные.

Индексы подразделяются на индивидуальные и общие.

*Индивидуальные индексы* определяют изменение однородного явления. Определяются отношением индексируемой ве-



личины в отчетном периоде к индексируемой величине в базисном периоде.

За базу сравнения можно принять не только уровень прошлого года, но и плановый уровень, нормативный уровень.

*Общие индексы* определяют изменение разнородного явления и могут быть образованы двумя методами.

1 метод. Образование на основе введения в индивидуальный индекс соизмерителя. Общие индексы бывают трех видов: переменного состава, постоянного состава и структурных изменений.

В индексе *переменного состава* от периода к периоду изменяется и индексируемая величина, и соизмеритель.

В индексе *постоянного состава* изменяется только индексируемая величина, а соизмеритель находится на уровне базисного периода.

В индексе *структурных изменений* изменяется только соизмеритель, а индексируемая величина фиксируется на уровне отчетного или базисного периода.

2 метод: Образование общего индекса на основе усреднения индексируемой величины по формуле средней арифметической.

### **Контрольные вопросы**

1. Что понимается под индексом?
2. Какие виды индексов существуют?
3. Какова взаимосвязь между индексами?

## 14. РАБОТА В СТАТИСТИЧЕСКИХ ПАКЕТАХ

### Теоретические положения

На сегодняшний день существует огромное количество программных продуктов, которые позволяют работать если не со всеми, то во всяком случае с большинством статистических методов. Их можно разделить на два класса:

1) те, в которых команды задаются с помощью текстового ввода (например, R и Python);

2) те, где конкретный метод выбирается с помощью меню.

IBM SPSS – мощный пакет, способный справиться с абсолютным большинством статистических задач.

StatSoft Statistica – главный конкурент SPSS на отечественном рынке. Привлекателен наличием нейронных сетей, базирующихся на многослойных персептронах. Позволяет на основе статистических данных прогнозировать изменения на фондовых рынках, управлять данными предприятий, прогнозировать экономические изменения.

R-commander – графический интерфейс для языка программирования R. Распространяется бесплатно.

PSPP – бесплатный аналог SPSS со схожим интерфейсом.

Microsoft Excel с надстройкой «Анализ данных».

### Контрольные вопросы

1. По какому принципу происходит разделение на классы программных продуктов для обработки статистических данных?

2. Охарактеризуйте программные продукты для обработки статистических данных, не требующие знаний программирования.

## СПИСОК ЛИТЕРАТУРЫ

1. Соловьев, В. И. Анализ данных в экономике. Теория вероятностей, прикладная статистика, обработка и визуализ. данных / В. И. Соловьев. – М. : Кнорус, 2019. – 498 с.
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум для бакалавриата и магистратуры : учебник для студентов высших учебных заведений, обучающихся по инженерно-техническим, естественно-научным и экономическим направлениям и специальностям / Б. Г. Миркин. – М. : Юрайт, 2015. – 173, [1] с.
3. Гришин, А. Статистика : учебное пособие / А. Гришин. – М. : Финансы и статистика, 2003. – 240 с.
4. Гусаров, В. М. Теория статистики / В. М. Гусаров. – М. : Аудит: ЮНИТИ, 1998. – 247 с.
5. Макарова, Н.В. Статистика в Excel : учебное пособие / Н. В. Макарова, В. Трофимец. – М. : Финансы и статистика, 2002. – 368 с.
6. Каплан, А. В. Математика, статистика, экономика на компьютере / А. В. Каплан [и др.]. – Москва – М. : ДМК-Пресс, 2006. – 594 с. : ил., табл.
7. Буре, В. М. Методы прикладной статистики в R и Excel : учебное пособие / В. М. Буре, А. А. Седаков, Е. М. Парилина. – М. : Лань, 2019 – 152 с.
8. Практикум по теории статистики / под ред. проф. Р. А. Шмойловой. – М. : Финансы и статистика, 1998. – 534 с.
9. Савельев, В. Статистика и котики / В. Савельев. – М. : АСТ, 2019. – 192 с.
10. Морозова, С. В. Статистика предприятий отрасли : учебно-методическое пособие по проведению практических занятий для студентов экономических специальностей : в 2 ч. / С. В. Морозова, Г. В. Ходанович. – Минск : БНТУ, 2012. – Ч. 1. – 130 с.
11. Статистика : учебное пособие / под ред. В. Ионина. – М. : ИНФРА-М, 2002. – 384 с.

12. Теория статистики / под ред. проф. Р. А. Шмойловой. – М. : Финансы и статистика, 2002. – 557 с.

13. Экономическая статистика: учебное пособие / под ред. Ю. Иванова. – М. : ИНФРА-М, 2002. – 480 с.

Учебное издание

**БУТОР** Любовь Васильевна  
**САХНОВИЧ** Татьяна Александровна

## **СТАТИСТИКА**

Учебно-методическое пособие  
для направления специальности 1-27 01 01-01  
«Экономика и организация производства (машиностроение)»  
в рамках специальности 1-27 01 01 «Экономика  
и организация производства (по направлениям)»

Редактор *Е. И. Бенищевич*  
Компьютерная верстка *Е. А. Беспанской*

Подписано в печать 18.07.2022. Формат 60×84 <sup>1</sup>/<sub>16</sub>. Бумага офсетная. Ризография.  
Усл. печ. л. 3,08. Уч.-изд. л. 2,41. Тираж 100. Заказ 766.

Издатель и полиграфическое исполнение: Белорусский национальный технический университет.  
Свидетельство о государственной регистрации издателя, изготовителя, распространителя  
печатных изданий № 1/173 от 12.02.2014. Пр. Независимости, 65. 220013, г. Минск.