

Министерство образования и науки Украины

ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ АВТОМОБИЛЬНО-ДОРОЖНЫЙ
УНИВЕРСИТЕТ

ОПОРНЫЙ КОНСПЕКТ ЛЕКЦИЙ

по дисциплине
«Статистика»

для иностранных студентов всех форм обучения
отрасли знаний 0305 – «Экономика и предпринимательство»
направления подготовки 6.030503 – «Международная экономика»

Харьков ХНАДУ 20__

Составитель: Е.П. Болдовская

Кафедра международной экономики

«СТАТИСТИКА»

(краткий конспект лекций для иностранных студентов)

ТЕМА 1 ПОНЯТИЕ СТАТИСТИКИ

1.1 Статистика как общественная наука.

1.2 Предмет и метод статистики.

1.3 Основные понятия, категории и показатели статистики.

1.1 Статистика как общественная наука

Термин *статистика* может употребляться в разных значениях:

- 1) отрасль практической деятельности по сбору, обработке, анализу и публикации массовых цифровых данных о различных социально-экономических явлениях и процессах;
- 2) совокупность сведенных итоговых цифровых показателей, собранных для количественной характеристики к.-л. сферы общественных явлений или отдельного вопроса;
- 3) отрасль общественных наук, специальная научная дисциплина.

Основные разделы статистической науки:

- *теория статистики* – рассматривает общие принципы и методы изучения общественно-экономических явлений и процессов;

- *социально-экономическая статистика* – изучает методологию построения макроэкономических показателей и их анализ, а также социальных условий жизни и труда населения, потребления им материальных благ и услуг;

- демографическая, промышленная, сельскохозяйственная и иные виды *отраслевой статистики* – изучает отдельные сферы общественной жизни или виды экономической деятельности.

1.2 Предмет и метод статистики

Специфика предмета статистики заключается в том, что эта наука изучает размеры и количественные соотношения массовых общественно-экономических явлений и процессов в неразрывной связи с их содержанием; она количественно исследует закономерности развития общественных явлений при конкретных условиях места и времени.

Выявление и оценка размеров, изменений и количественных соотношений определенных массовых общественных явлений производится путем реализации трех основных стадий экономико-статистического исследования:

- *статистическое наблюдение* – получение первичных значений исследуемых признаков от каждой единицы статистической совокупности путем их регистрации (учета) на основе тщательно разработанной программы;

- *статистическая сводка и группировка* первичных данных – всесторонняя систематизация материала статистического наблюдения (дает возможность выделить в исследуемой совокупности социально-экономические типы явлений, охарактеризовать их структуру, выявить взаимосвязи и взаимозависимости между показателями);

- *анализ статистической информации* – проведение анализа данных на основе вычисления обобщающих показателей, напр., абсолютных, относительных и средних величин, индексов и показателей, характеризующих динамику явлений, их вариацию и тесноту связи между ними (дает возможность раскрыть причинные связи исследуемых явлений, определить влияние и взаимодействие разных факторов, оценить эффективность принятых УР и возможные социально-экономические последствия сложившейся ситуации).

1.3 Основные понятия, категории и показатели статистики

Статистика оперирует категориями, отражающими существенные, наиболее общие свойства действительности.

Статистическая совокупность – множество единиц (объектов или явлений), объединенных к.-л. общими свойствами, обобщенных единой закономерностью и варьирующих (изменяющихся) в пределах общего качества.

При организации статистического наблюдения различают:

- *генеральную совокупность (N)* – вся совокупность реально существующих статистических объектов, т.е. все возможные реализации интересующего нас признака;

Генеральная совокупность может быть конечной или бесконечной.

- *выборочную совокупность (n)* – совокупность единиц, отобранных из генеральной совокупности по определенным правилам, признакам.

Статистическая совокупность может быть *однородной* (ее элементы схожи между собой по к.-л. существенным признакам) и *неоднородной* (элементы, ее составляющие, относятся к различным типам изучаемого явления).

Отдельные объекты, явления и процессы, образующие статистическую совокупность, называются **единицами совокупности**.

Единицы статистической совокупности характеризуются **признаками**, т.е. отличительными чертами, свойствами, качествами, присущими как всей совокупности, так и отдельной единице совокупности.

Отдельные значения признака называются **вариантами**.

Вариация – различия в значениях того или иного признака у отдельных единиц статистической совокупности; она складывается в результате того, что индивидуальное значение признака зависит от разных условий.

В статистике признаки условно делят:

1) *количественные* (имеют числовое выражение) и *качественные*, или *атрибутивные* (признаки, варианты которых, характеризуя особенности отдельных единиц, не имеют количественного выражения. Если качественные признаки могут принимать только одно из двух противоположных значений, их наз. *альтернативными*;

2) *вариационные* (имеют разные значения у отдельных единиц исследуемого явления) и *постоянные* (имеют неизменные значения у всех единиц совокупности).

Статистический показатель – обобщенная количественная характеристика социально-экономических явлений и процессов в их качественной определенности относительно конкретных условий места и времени.

Совокупность показателей, всесторонне характеризующая развитие общества, называется **системой показателей**.

Сложный комплекс экономических явлений описывают сведенные экономические показатели, которые иногда называют *синтетическими* (напр., национальное богатство, ВВП).

Значение показателя является результатом измерения объектов и изменяется в зависимости от методологических особенностей его (показателя) построения, обусловленных, в свою очередь, степенью охвата исследуемых процессов.

Показатели бывают *натуральными*, если они выражены в физических единицах (штуках, метрах, тоннах и т.п.) и *стоимостными*, если они соответствуют денежной оценке экономических объектов.

Часто статистические показатели условно делят на *объемные (первичные)*, при помощи которых измеряют объем совокупности объектов (напр., произведенная продукция, численность работников), и *качественные (вторичные)*, которые характеризуют уровень развития явления (напр., себестоимость единицы продукции, уровень рентабельности, производительности труда).

ТЕМА 2 СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1 Сущность статистического наблюдения. Статистические данные.

2.1 Сущность статистического наблюдения. Статистические данные

Статистическое исследование основывается на массовых первичных данных, полученных путем сбора и учета отдельных фактов об исследуемых явлениях и процессах в соответствии с целью и задачами. Метод, используемый на этом этапе исследования, называется статистическим наблюдением.

Статистическое наблюдение (СН) – первая стадия статистического исследования, представляющая собой научно-организованный сбор данных (информации) о массовых явлениях и процессах общественной жизни.

Задачей СН является получение достоверной исходной информации, которая объективно отражает фактическое состояние явления. Для этого учет фактов и сбор первичных материалов должны быть тщательно спланированы и организованы на научной основе.

Не все первичные данные можно положить в основу выводов и обобщений. Поэтому *статистические данные*, пригодные для научно обоснованных обобщений, должны быть:

- полными, а не случайными, отрывистыми;
- достоверными и точными;
- однотипными и сравнимыми (сопоставимыми во времени и пространстве).

Статистические данные – это данные, собранные на реальных экономических объектах, представляющие собой набор определенных характеристик объекта, отмеченных за определенный период времени.

При измерении количественных признаков могут быть получены два типа статистических данных:

- динамические (временные) ряды;
- вариационные (перекрестные) ряды.

Временные ряды – это последовательность наблюдений за одним и тем же процессом или явлением в различные (как правило, равные) промежутки времени.

Отдельные значения признака, относящиеся к определенным периодам времени, называют уровнями динамического ряда (x_t).

При формировании динамических рядов могут возникать трудности, связанные с недостатком необходимых данных. Один из наиболее распространенных способов его преодоления – выявление закономерностей, которым подчиняется динамический ряд (такие закономерности описывают при помощи математических уравнений) с последующей интерполяцией или экстраполяцией недостающих его уровней.

Интерполяция – нахождение неизвестных показателей (уровней ряда) в рамках имеющегося ряда динамики.

Экстраполяция – нахождение неизвестных показателей за рамками ряда динамики, т.е. в конце или в начале.

Вариационные ряды – последовательность наблюдений по какому-либо экономическому показателю для разных однотипных процессов или объектов. Все замеры (наблюдения) производятся в одно и то же время. Значения вариационного ряда располагают в порядке возрастания. Отдельные значения признака, относящиеся к определенным объектам наблюдения, называют вариантами (x_i).

ТЕМА 3

СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

3.1 Сущность, организация и техника статистической сводки.

3.2 Основные виды и задачи статистической группировки.

3.3 Способы статистической группировки.

3.4 Статистические ряды распределения.

3.1 Сущность, организация и техника статистической сводки

Статистическая сводка и группировка является вторым этапом статистического наблюдения.

Статистическая сводка – систематизация и подытоживание данных (единичных фактов) с целью получения обобщающей (по определенным существенным признакам) характеристики исследуемого явления для его анализа и прогнозирования.

Этапы статистической сводки:

- 1) группировка статистических данных;
- 2) подсчет итоговых групповых данных, показателей (т.е. в разрезе групп);
- 3) представление полученных результатов в виде статистических рядов или таблиц.

3.2 Основные виды и задачи статистической группировки.

Поскольку все социально-экономические явления, изучаемые статистикой, отличаются многообразием форм и стадий развития, состоят из различных частей, имеющих специфические черты, то основой статистической сводки является деление исследуемой совокупности единиц на различные между собой, но внутренне однородные части путем их объединения в типовые по определенному признаку группы.

Группировка – процесс образования однородных групп на основе разделения всей совокупности исследуемого явления на отдельные группы (части) по наиболее существенным признакам.

Признаки, по которым производится разделение единиц определенной совокупности на группы, называются **признаками группировки**, или **основой группировки**.

Особенным видом группировки является классификация, представляющая собой устойчивую номенклатуру классов и групп и при помощи которой вариация признаков социально-экономических явлений фиксируется в определенном системном виде.

В зависимости от задач, решаемых при помощи статистической группировки, выделяют такие виды группировок:

- типологическая – решает задачу выделения социально-экономических типов явлений, т.е. разделения всей совокупности на качественно однородные группы, определения существенных отличий между ними и признаков, являющихся для них общими (напр., деление предприятий на мелкие, средние и крупные);

- структурная – решает задачу характеристики распределения качественно однородной совокупности на группы по определенному признаку для выявления закономерностей распределения единиц совокупности по варьирующим значениям исследуемого признака, для изучения состава совокупности и структурных сдвигов.

- аналитическая – решает задачу выявления и изучения взаимосвязи и зависимости между признаками явлений.

В предпринимательской деятельности возможны такие взаимосвязи между признаками явлений:

- фактор (причина) – количественный признак, а следствие (результат) – качественный (напр., стаж работы и квалификация работника);

- фактор – качественный признак, а следствие – количественный (напр., стаж работы и квалификация работника и производительность его труда);

- фактор и результат – качественные признаки (напр., категории работников и их образование);

- фактор и результат – количественные признаки (напр., производительность труда и средняя зарплата).

Характерной особенностью аналитических группировок является то, что каждая группа факторного признака характеризуется средними значениями результативного признака.

Группировка может быть *простой*, если группы образуются по одному признаку, и *сложной*, если группы образуются по нескольким признакам.

Структурные группировки, как и типологические, могут производиться по атрибутивным и количественным признакам.

В случае группировок на основе количественных признаков необходимо определить оптимальное количество образуемых групп и размер интервала и проследить, чтобы не исчезли особенности исследуемого явления. Эти два момента взаимосвязаны: чем меньше интервал, тем больше число групп и наоборот. Важным требованием при решении этого вопроса является выбор такого количества групп и такого интервала, которые бы давали возможность более-менее равномерно распределить все единицы совокупности в разрезе групп, обеспечили их представительность и качественную однородность.

3.3 Основные способы группировки

Способ 1. Группировка с использованием равных интервалов

Количество групп (интервалов), K , тесно связано с объемом совокупности и определяется по формуле Стерджесса:

$$K = 1 + 3,322 \cdot \lg(n), \quad (2.1)$$

где n – количество наблюдений;

$\lg(n)$ – десятичный логарифм числа n .

Полученное значение K обычно округляется до целого в большую сторону.

Интервал – количественное значение, определяющее и отделяющее одну группу от другой.

Далее рассчитывается ширина интервала группировки:

$$\Delta X = \frac{x^{\max} - x^{\min}}{K}, \quad (2.2)$$

где x_i^{\max} – максимальное значение изучаемого показателя;

x_i^{\min} – минимальное значение изучаемого показателя.

Значение ΔX округляется до целого числа.

Границы интервалов определяются таким образом:

- нижняя граница первого интервала группирования:

$$a_1 = x^{\min} - 0,5 \cdot \Delta X . \quad (2.3)$$

- верхняя граница первого интервала группирования:

$$b_1 = a_1 + \Delta X \quad (2.4)$$

- границы дальнейших интервалов устанавливают по правилу: нижняя граница очередного интервала (a_k) равняется верхней границе предыдущего интервала (b_{k-1}), а верхняя граница (b_k) – равняется нижней плюс ширина интервала группировки:

$$a_2 = b_1, \quad b_2 = a_2 + \Delta X ; \quad (2.5)$$

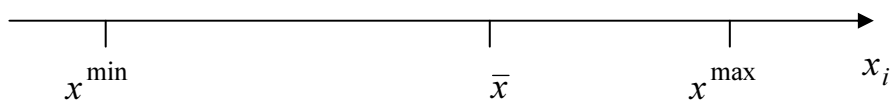
$$a_k = b_{k-1}, \quad b_k = a_k + \Delta X \quad (2.6)$$

Интервалы могут быть открытыми и закрытыми.

Способ 2. Группировка с использованием неравных интервалов

Группировку по неравным интервалам используют для описания статистических данных, имеющих явную асимметрию распределения частот и частостей. Ширину и границы этих интервалов устанавливают на основе логического анализа предварительных сведений о качественных и количественных характеристиках изучаемого явления.

1) Для выделения групп объектов изучаемого признака по неравным интервалам ранжируют его значения. Далее весь интервал возможных значений признака $[x^{\min}; x^{\max}]$ делят на два интервала, отделяемые друг от друга средним значением признака \bar{x} .



Среднее значение рассчитывается по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i .$$

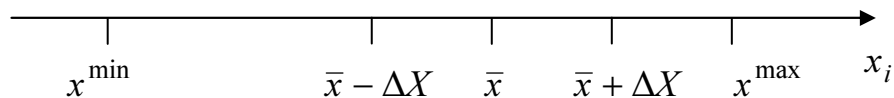
На первом интервале $[x^{\min}; \bar{x}]$ будут размещены значения изучаемого признака, которые меньше среднего значения \bar{x} . На втором интервале $[\bar{x}; x^{\max}]$ будут размещены значения изучаемого признака, которые больше среднего значения \bar{x} .

В случае асимметричного распределения точка, соответствующая среднему значению признака \bar{x} , не будет делить интервал $[x^{\min}; x^{\max}]$ на равные части, а будет смещена к какому-либо из концов интервала.

2) Из двух интервалов, разделенных средним значением \bar{x} , выбирают интервал меньшей длины, для чего сравнивают по модулю величины $[x^{\min} - \bar{x}]$ и $[\bar{x} - x^{\max}]$.

3) Длину наименьшего из двух сравниваемых интервалов делят на две равные части, и полученное значение ΔX прибавляют к среднему \bar{x} и отнимают от него.

Полученные координаты двух точек $(\bar{x} - \Delta X)$ и $(\bar{x} + \Delta X)$ наносят на числовую ось вариационного ряда слева и справа от среднего значения.



4) В результате числовая ось, которая соответствует ранжируемому вариационному ряду значений изучаемого признака, делится на три интервала: $[x^{\min}; \bar{x} - \Delta X]$, $[\bar{x} - \Delta X; \bar{x} + \Delta X]$ и $[\bar{x} + \Delta X; x^{\max}]$.

При достаточно большой величине размаха вариации исследуемого признака процедура дробления всей числовой оси может быть повторена.

Способ 3. Группировка с использованием среднеквадратического отклонения.

1. По несгруппированным данным рассчитывают среднее значение (\bar{x}) и среднеквадратическое отклонение (σ_x).

Среднеквадратическое (стандартное) отклонение рассчитывается по формуле:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}.$$

2. От среднего значения \bar{x} , двигаясь вначале влево, а затем вправо, откладываем интервалы группировки. Границы интервалов для распределения объектов совокупности по 6 группам будут такими:

$$1) \bar{x} - 3\sigma_x < X \leq \bar{x} - 2\sigma_x;$$

$$2) \bar{x} - 2\sigma_x < X \leq \bar{x} - \sigma_x;$$

$$3) \bar{x} - \sigma_x < X \leq \bar{x};$$

$$4) \bar{x} < X \leq \bar{x} + \sigma_x;$$

$$5) \bar{x} + \sigma_x < X \leq \bar{x} + 2\sigma_x;$$

$$6) \bar{x} + 2\sigma_x < X \leq \bar{x} + 3\sigma_x;$$

3.4 Статистические ряды распределения

В результате статистического наблюдения получают отдельные значения признаков единиц исследуемой совокупности

В результате сводки и группировки материалов статистического наблюдения получают **статистические ряды распределения**, представляющие собой упорядочение единиц изучаемой совокупности по группам по определенному признаку.

Ряд распределения характеризуют:

- **варианты** (x_i) – отдельные числовые значения количественного признака в вариационном ряду;

- **частоты** (f_i) – 1) число, характеризующие количество повторений той или иной варианты (т.е. показывающие, сколько раз варианта встречается в данной совокупности) или 2) число, характеризующие количество попаданий наблюдений (значений признака) на тот или иной интервал. Сумма частот равна объему ряда распределения, т.е. объему изучаемой совокупности;

- **частоты** (ω_i) – отношение частоты каждой группы к числу единиц совокупности. Сумма частот ряда распределения равна 100 % или 1.

$$\omega_i = \frac{f_i}{n}.$$

- **накопленные (кумулятивные) частоты и частоты** ($\sum f_i, \sum \omega_i$).

$$\sum_{i=1}^k f_k = \sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k, \quad \sum_{i=1}^k \omega_k = \sum_{i=1}^k \omega_i = \omega_1 + \omega_2 + \dots + \omega_k.$$

Ряд распределения можно представить графически в виде:

- **гистограммы** – строятся прямоугольной в системе координат: по оси абсцисс откладываются варианты, по оси ординат – частоты (или частоты);

- **полигона** – многоугольник, обрамленный ломаной, соединяющей середины столбцов гистограммы

- **кумуляты** – строятся по накопленным частотам.

ТЕМА 4 ОБОБЩАЮЩИЕ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

4.1 Абсолютные и относительные величины.

4.2 Средние величины.

4.1 Абсолютные и относительные величины

Абсолютные величины – показатели, выражающие размеры социально-экономических явлений и процессов в конкретных условиях места и времени.

По способу выражения размеров исследуемых явлений делятся абсолютные величины на:

- *индивидуальные* – характеризуют количественные признаки у отдельных единиц;
- *итоговые* – характеризуют размер признака совокупности, полученного от сложения значений признаков отдельных единиц совокупности.

В зависимости от причин и заданий исследования абсолютные величины делятся на:

- *натуральные*;
- *условно-натуральные* (чел.-ч., т.-км.);
- *стоимостные*.

Относительные величины – показатели, выражающие количественные соотношения между явлениями социально-экономической жизни, т.е. это сопоставление величин тех или иных показателей друг с другом. Их получают как частное от деления двух абсолютных величин.

При вычислении относительных величин следует иметь в виду, что числитель – это изучаемый показатель, его называют *отчетной величиной*. Величину, с которой сопоставляют другие величины (знаменатель), называют *основой*, или *базой сравнения*, *базисной величиной*.

Различают следующие относительные величины:

а) отношение между одноименными показателями:

1) *относительные величины выполнения договорных обязательств* (в плановой экономике – выполнения плана) – показатели, определяемые делением объема

выполненных обязательств на весь объем обязательств, предусмотренных договором (или иначе, факт / план);

2) *относительные величины динамики* – показатели, выражающие степень изменения явлений во времени; характеризуют направление и скорость изменения явлений во времени, темпы их развития.

В зависимости от характера базы сравнения различают два вида относительных величин динамики:

- с переменной базой сравнения (цепные), когда за базу сравнения принимается значение показателя в предшествующем году: $Tr_t = \Pi_t / \Pi_{t-1}$;

- с постоянной базой сравнения (базисные), когда за базу сравнения принимается значение показателя в к.-л. одном году, принятом за основу: $Tr_t = \Pi_t / \Pi_0 = \Pi_t / \Pi_{баз}$.

3) *относительные величины структуры* – характеризуют состав исследуемой совокупности. Их рассчитывают как отношение абсолютной величины каждого составляющего элемента к абсолютной величине всей совокупности; выражают обычно в процентах (база сравнения = 100) или долях единицы, в т.н. коэффициентной форме (база сравнения = 1).

4) *относительные величины координации* – соотношения отдельных частей определенной совокупности (напр., в группе учатся 20 девушек и 10 юношей, тогда относительная величина координации $20 : 10 = 2$ показывает, что численность девушек в два раза превышает численность юношей, а величина $10 : 20 = 0,5$, показывает, что численность юношей в два раза меньше численности девушек, т.е. составляет ее половину).

5) *относительные величины сравнения в пространстве* – соотношение одноименных величин разных объектов (напр., соотношение ВВП двух стран)

б) отношение между разноименными показателями – *относительные величины интенсивности* – показатели, характеризующие степень распространения, развитие явления в определенной среде. Эти величины указывают на то, сколько единиц одной совокупности приходится на единицу другой совокупности (напр., количество автомобилей на 1000 человек городского населения, авт./чел.). В отличие от других видов, относительная величина интенсивности всегда имеет название (размерность).

4.2 Средние величины

Вариация количественных признаков может быть:

- 1) *дискретной* (величина количественного признака может принимать только определенные значения), напр., разряд рабочего;
- 2) *непрерывной* (величина количественного признака может принимать любые значения в определенном числовом промежутке).

Средняя величина – обобщающая мера варьирующего признака, которая характеризует его типовой уровень в расчете на единицу однородной совокупности.

Средние величины могут быть простые (рассчитываются по первичным, т.е. негруппированным, данным) и взвешенные (по вторичным, т.е. сгруппированным).

В статистике различают следующие виды средних величин:

1) *Средняя арифметическая*

- простая:

$$\bar{x}_{\text{арифм. простая}} = \frac{\sum_{i=1}^n x_i}{n}.$$

- взвешенная:

$$\bar{x}_{\text{арифм. взвешенная}} = \frac{\sum_{i=1}^n x_i \cdot f_i}{n} \quad (\text{для дискретного ряда})$$

$$\text{или } \bar{x}_{\text{арифм. взвешенная}} = \frac{\sum_{k=1}^K x_k \cdot f_k}{\sum_{k=1}^K f_k} \quad (\text{для непрерывного ряда}).$$

f_i – частота признака в совокупности;

x_k – середина интервала группировки, $x_k = \frac{a_k + b_k}{2}$;

f_k – частота на k -ом интервале.

2) ***Средняя геометрическая*** – определяется как корень n -ой степени из произведения относительных величин динамики и позволяет определить средний темп роста:

$$\bar{x}_{\text{геометр. простая}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

3) Средняя квадратическая:

$$\bar{x}_{\text{квдр. простая}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}, \quad \bar{x}_{\text{квдр. взвешенная}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 \cdot f_i}{n}}$$

ТЕМА 5

ОПИСАТЕЛЬНЫЕ СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ: СТРУКТУРНЫЕ (НЕПАРАМЕТРИЧЕСКИЕ) СРЕДНИЕ ВЕЛИЧИНЫ

Непараметрические средние – это мода и медиана.

Мода (M_o) – величина признака (варианта), которая чаще всего встречается в изучаемой совокупности.

В *дискретном вариационном ряду* модальной является варианта, имеющая наибольшую частоту.

Пример

Размер костюма, x_i	44	46	48	50	52	54	56
Кол-во проданных костюмов, f_i	2	8	20	91	44	19	5

Т.о., модальной является варианта № 4, а мода равна 50.

Иногда встречаются ряды распределения, в которых не одна, а две варианты одинаково модальны, т.е. имеют наибольшие равные частоты. Это означает, что есть две моды – распределение бимодально.

В *непрерывном (интервальном) вариационном ряду* мода определяется по формуле:

$$M_o = X_{M_o} + h_{M_o} \cdot \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}$$

где X_{M_o} – нижняя граница *модального интервала* (интервала, который имеет наибольшую частоту);

h_{M_o} – длина модального интервала;

f_{M_o} – частота модального интервала;

f_{M_o-1} , f_{M_o+1} – соответственно частота интервала, предшествующего модальному и последующего за ним.

Медиана (M_e) – варианта, которая является серединой упорядоченного вариационного ряда, т.е. делит его на две равные части. Медиана указывает на значение признака, которого достигла половина единиц совокупности.

Для определения медианы в *дискретном вариационном ряду* сумму всех частот ряда делят на два и к полученному результату прибавляют 0,5.

В примере с костюмами:

$$M_e = \frac{\sum F}{2} + 0,5 = \frac{189}{2} + 0,5 = 95.$$

Далее производят аккумуляцию частот (рассчитывают накопленные частоты), и варианта, имеющая накопленную частоту, равную или превышающую полученное ранее значение, является медианной.

В примере $2 + 8 + 20 + 91 = 121 > 95$, следовательно, медианой является варианта № 4, т.е. костюм 50 размера.

В *непрерывном (интервальном) вариационном ряду* медиана определяется по формуле:

$$M_e = X_{Me} + h_{Me} \cdot \frac{\frac{\sum F}{2} - S_{Me-1}}{f_{Me}}$$

де X_{Me} – нижняя граница медианного интервала (интервала, кумулятивная частота которого равна или превышает половину суммы частот);

h_{Me} – длина медианного интервала;

f_{Me} – частота медианного интервала;

$\sum f$ – сумма частот ряда;

S_{Me-1} – сумма накопленных частот интервала, предшествующего медианному.

ТЕМА 5 ПОКАЗАТЕЛИ ВАРИАЦИИ

К показателям вариации относят:

- размах вариации;
- дисперсию;
- среднее квадратическое (стандартное) отклонение;
- коэффициент вариации.

В отличие от средних величин, показатели вариации позволяют изучить степень распределения признака в совокупности.

Средняя величина не объясняет, как группируются около нее отдельные значения – лежат ли они вблизи или, наоборот, существенно отклоняются от средней.

В случае, когда отдельные значения вариант достаточно близко располагаются около средней, последняя достаточно надежно описывает всю совокупность.

В связи с этим необходимо дополнительное исследование отклонений отдельных вариант исследуемого признака от средней величины.

1) **Размах вариации** – разность между максимальным и минимальным значением признака (определяется в тех же единицах, что и значение признака):

$$R_x = x_{\max} - x_{\min}.$$

Недостатком этого показателя является то, что его величина зависит от крайних значений признака, т.е. он фиксирует только крайние отклонения и не учитывает отклонений остальных вариант от их средней.

2) **Среднее линейное отклонение** – частное от деления всех отклонений, взятых по модулю, на их число:

$$d_x = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (\text{данные не сгруппированы}),$$

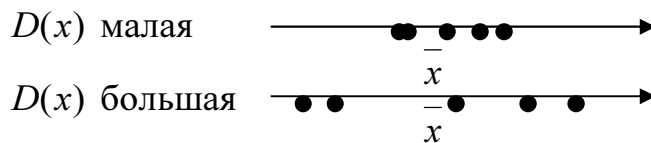
$$d_x = \frac{\sum_{i=1}^n |x_i - \bar{x}| \cdot f_i}{\sum_{i=1}^n f_i} \quad (\text{данные сгруппированы}).$$

3) **Дисперсия** – показатель среднего квадрата отклонения вариант признака от среднего значения:

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \text{ (данные не сгруппированы),}$$

$$D_x = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^n f_i} \text{ (данные не сгруппированы).}$$

Дисперсия характеризует, как сильно рассеяны значения выборки относительно среднего значения.



4) **Среднее квадратическое (стандартное) отклонение** – корень квадратный из дисперсии:

$$\sigma_x = \sqrt{D_x} = \sqrt{\sigma_x^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}.$$

Стандартное отклонения является мерилем надежности средней. Чем меньше среднее квадратическое отклонение, тем полнее средняя арифметическая отражает всю совокупность.

5) **Коэффициент вариации** – критерий типичности средней – отношение стандартного отклонения к центру распределения (средней) и выражается в %:

$$V_x = \frac{\sigma_x}{\bar{x}} \cdot 100\%.$$

Считается, что совокупность является однородной если $V_x \leq 33\%$. В противном случае, говорят, что средняя характеризует совокупность по признаку, который существенно изменяется у отдельных единиц и типичность такой средней – сомнительна.

5) **Коэффициент осцилляции** – отражает относительное колебание крайних значений вокруг среднего:

$$O_x = \frac{R_x}{\bar{x}} \cdot 100\%.$$

ТЕМА 6 КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Статистика призвана изучать коммерческую деятельность с количественной стороны. Это осуществляется с помощью соответствующих приемов и методов статистики.

Важнейшей целью статистики является изучение объективно существующих связей между явлениями. В ходе статистического исследования этих связей необходимо выявить причинно-следственные зависимости между показателями, т.е. насколько изменение одних показателей зависит от изменения других показателей.

Статистические показатели коммерческой деятельности могут состоять между собой в следующих основных видах связи: балансовой, компонентной, факторной и др.

Факторные связи могут рассматриваться как функциональные и корреляционные.

Если с изменением одной из переменных вторая изменяется строго определенным образом, т.е. значению одной переменной обязательно соответствует одно или несколько точно заданных значений другой переменной, связь между ними является функциональной.

При корреляционные связи с изменением значения одной переменной вторая может в определенных пределах принимать любые значения с некоторыми вероятностями, но ее среднее значение или иные статистические (массовые) характеристики изменяются по определенному закону, т.е. разным значениям одной переменной соответствуют разные распределения значений другой переменной.

Термин "корреляция" означает "связь". В эконометрике этот термин обычно используется в сочетании "коэффициенты корреляции"

Слово correlation (корреляция) состоит из приставки «со-», которая обозначает совместность происходящего (по аналогии с «координация») и корня «relation», переводится как «отношение» или «связь» (вспомним public relations – связи с общественностью). Дословно correlation переводится как взаимосвязь.

Корреляция — статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). При этом, изменения одной или нескольких из этих величин приводят к систематическому изменению другой или других величин.

Корреляционная связь - это связь, где воздействие отдельных факторов проявляется только как тенденция (в среднем) при массовом наблюдении фактических данных.

Наиболее простым вариантом корреляционной зависимости является парная корреляция, т.е. зависимость между двумя признаками (результативным и факторным или между двумя факторными). Математически эту зависимость можно выразить как зависимость результативного показателя y от факторного показателя x .

Характерной особенностью корреляционных связей является то, что они проявляются не в единичных случаях, а в массе.

Для того, чтобы установить, есть ли зависимость между величинами, используются многообразные статистические методы, позволяющие определить, во-первых — какие связи; во-вторых — тесноту связи (в одном случае она сильная, устойчивая, в другом — слабая); в-третьих — форму связи (т.е. формулу, связывающую величину x и y).

Математической мерой корреляции двух случайных величин служит коэффициент корреляции.

Коэффициент корреляции — это мера взаимосвязи измеренных явлений.

Коэффициент корреляции (обозначается « r »)

Наиболее удобной формулой для расчета коэффициента корреляции является:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2}} \quad \text{или} \quad (2.4)$$

$$r_{xy} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{X^2 - (\bar{X})^2} \cdot \sqrt{Y^2 - (\bar{Y})^2}}, \quad \text{где} \quad \overline{XY} = \frac{1}{n} \sum_{i=1}^n x_i y_i; \quad \bar{Y}^2 = \frac{1}{n} \sum_{i=1}^n y_i^2; \quad X^2 = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

(2.5)

Коэффициент корреляции является показателем плотности (тесноты) линейной взаимосвязи.

Свойства коэффициента корреляции:

- значения коэффициента всегда находятся в пределах $-1 \leq r_{xy} \leq 1$;
- если $r_{xy} > 0$, то зависимость между фактором X и Y прямая, т.е. с ростом X показатель Y также возрастает;
- если $r_{xy} < 0$, то зависимость между фактором X и Y обратная;
- если $|r_{xy}| \rightarrow 1$, то плотность связи между X и Y велика – связь почти линейная (рис. 2.3);
- если $|r_{xy}| \rightarrow 0$, либо связи нет (рис. 2.2), либо связь резко нелинейная