

СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Лекция 1.

Общая характеристика статистического анализа данных.

1. Статистика. Предмет, объект и задачи статистического исследования.

2. Основные понятия и характеристики.

1. Статистика. Предмет, объект и задачи статистического исследования.

Статистика – это наука, изучающая количественную сторону массовых явлений и процессов в неразрывной связи с их качественной стороной, количественное выражение закономерностей общественного развития в конкретных условиях места и времени.

Объектом статистического исследования являются явления и процессы социально-экономической жизни общества, в которых отображаются и находят свое выражение социально-экономические отношения людей.

Предметом статистического исследования является количественная сторона массовых общественных явлений в неразрывной связи с их качественной стороной, которая отображается посредством объективных статистических показателей, то есть особых признаков.

Три этапа работы со статистическими данными:

- 1) сбор – это массовое научно-организованное наблюдение, посредством которого получают первичную информацию об отдельных фактах изучаемого явления;
- 2) сводка, группировка, обработка первичной информации (т.е. распределение множества фактов на однородные группы, итоговый подсчет по каждой группе и оформление полученных итогов в виде статистической таблицы);
- 3) обобщение и интерпретация статистической информации.

Задачи статистического исследования:

- 1) исследование происходящих в обществе преобразований социальных и экономических процессов на основе системы специальных показателей;
- 2) обобщение и прогнозирование тенденций развития народного хозяйства и его составляющих;
- 3) организация статистики отраслей народного хозяйства и общества (прикладной статистики). И т.д.

2. Основные понятия и характеристики.

Признаком называют свойство, характерную черту или иную особенность единиц объектов, явлений, которые могут быть наблюдаемы или измерены.

Например, к демографическим и социальным признакам относят пол, возраст...

Признаки бывают:

- *атрибутивными, т.е. качественными;*
- *количественными.*

Количественные признаки выражаются числами. Делятся на *дискретные* и *непрерывные*.

Дискретные признаки могут принимать только лишь отдельные, изолированные значения. (Например, целые числа: число комнат в квартире, число студентов в группе).

Непрерывные признаки способны принимать любые числовые значения в определенных границах. На практике значения непрерывных признаков округляют до определенной степени точности.

Атрибутивные (качественные) – это те признаки, отдельные значения которых отличаются друг от друга существенными (качественными) моментами. Эти признаки не измеряются, а описываются. Делятся на *альтернативные* и *номинальные*.

Альтернативные признаки принимают одно из двух значений (пол человека; да, нет).

Номинальные признаки способны принимать одно из нескольких значений (например, национальность).

Вариация – это колеблемость, изменчивость величины признака по отдельным единицам совокупности явлений.

Отдельное значение признака называют **вариантой** этого признака.

Статистическая совокупность – это совокупность социально-экономических объектов или явлений, объединенных некоей качественной основой, но отличающихся друг от друга отдельными признаками. Например, совокупность предприятий.

Единицей статистической совокупности является элемент статистической совокупности, который является носителем общих признаков.

Объем совокупности – общее число единиц совокупности.

Совокупности бывают однородными и разнородными.

Совокупность называют **однородной** если один или несколько изучаемых существенных признаков ее объектов являются общими для всех единиц.

Статистический показатель – это понятие (категория), отражающее количественные характеристики (размеры) социально-экономических явлений.

Статистический показатель следует отличать от признака. Признак – это свойство, присущее единице совокупности. Например, возраст. Показатель – характеристика группы единиц или совокупности в целом. Например, средний возраст работников фирмы.

Сводка и группировка статистических данных.

Статистическое наблюдение – планомерный, научно организованный сбор массовых данных о социально-экономических явлениях и процессах.

Объект статистического наблюдения – явление или процесс, обладающий свойствами однородности, воспроизводимости и устойчивости.

Статистическое наблюдение является первым этапом *статистического исследования*.

Получаемая в ходе статистического наблюдения информация характеризует единицы статистической совокупности с различных сторон и не позволяет сделать обобщающие выводы об объекте в целом (т.е. о всей статистической совокупности).

Статистическая сводка – комплекс последовательных операций по обобщению конкретных единичных фактов, образующих совокупность, для выявления типичных черт и закономерностей, присущих изучаемому явлению в целом.

По глубине обработки материала сводка бывает *простая* и *сложная*.

Простой сводкой называется операция по подсчету общих итогов по совокупности единиц наблюдения.

Сложная сводка представляет собой комплекс операций, включающих группировку единиц наблюдения, подсчет итогов по каждой группе и по всему объекту и представление результатов группировки и сводки в виде статистических таблиц.

В сводке отдельные единицы статистической совокупности объединяются в группы при помощи метода группировок.

Группировка – это процесс образования однородных групп на основе разбиения статистической совокупности на части по существенным признакам.

С помощью метода группировок решаются задачи:

- 1) выделение социально-экономических типов явлений;
- 2) изучение структуры явления и структурных сдвигов, происходящих в нем;
- 3) выявление связи и зависимости между явлениями.

Различают следующие **виды группировок**:

- *типологическая группировка*, т.е. разделение качественно разнородной совокупности на классы или однородные группы;
- *структурная группировка*, в которой происходит разделение однородной совокупности на группы, характеризующие ее структуру по какому-либо варьируемому признаку;
- *аналитическая группировка*, выявляющая взаимосвязи между изучаемыми явлениями и их признаками. Особенностями аналитической группировки является то, что единицы группируются по факторному признаку; каждая выделенная группа характеризуется средними значениями результативного признака.

Признаки, оказывающие влияние на изменение других признаков, называются **факторными**. Признаки, изменяющиеся под влиянием факторных, называются **результативными**.

Группировка, в которой группы образованы по одному признаку, называется **простой**. **Сложной** называется группировка, в которой разделение на группы происходит по 2-м и более признакам, взятым в сочетании (*комбинированная группировка*).

Принципы построения статистических группировок.

1. *Выбор группировочного признака* – признака, по которому производится разбиение совокупности на отдельные группы.

2. *Определение количества групп*. Если в основание группировки положен качественный признак, то групп будет столько, сколько существует уровней признака.

Если в основании группировки дискретный признак, то, как правило, групп столько, сколько существует различных значений признака.

Если признак количественный и при этом непрерывный, число групп определяется двумя способами:

1) по формуле Стерджесса:

$$n = 1 + 3,322 \cdot \lg N,$$

где n – число групп, N – численность совокупности.

2) если N не превышает 25-30 единиц, то в расчет берутся 3-4 группы,

3. *Определение интервала группировки.* *Нижней границей* интервала называется наименьшее значение признака в интервале, а *верхней границей* – наибольшее значение признака в интервале.

Для группировки с равными интервалами величина интервала:

$$h = \frac{R}{n} = \frac{x_{\max} - x_{\min}}{n},$$

где x_{\max} , x_{\min} – максимальное и минимальное значения группировочного признака в совокупности; n – число групп, R – размах вариации.

Интервалы могут быть:

- *закрытыми*, когда у интервала указаны обе границы;
- *открытыми*, когда у первого интервала не указана верхняя граница, а у последнего – нижняя.

Результаты сводки и группировки оформляют, обычно, в виде статистических таблиц. В статистической таблице выделяют два элемента:

- подлежащее (в первой вертикальной или горизонтальной графе) – перечень единиц или групп, на которые подразделена вся масса единиц наблюдения;
- сказуемое – числа, при помощи которых характеризуются выделенные в подлежащем единицы или группы.

Примеры различных группировок.

Таблица 1 – Типологическая группировка

№ п/п	Группы предприятий по форме собственности	Число предприятий	
		единиц	в % к итогу
1	Республиканская собственность	26326	93,6
2	Муниципальная	420	1,5
3	Частная	1366	4,9
Всего		28112	100,0

Таблица 2.– Структурная группировка

№ п/п	Группы населения по размеру среднедушевого дохода, руб.	Численность населения	
		всего, млн. чел.	в % к итогу
1	до 1000 руб.	2,4	2,0
2	1000–1800	24,8	18,0
3	1800–2600	34,2	25,0
4	2600–3400	29,4	21,5
5	3400–10000	45,7	33,5
Всего		136,5	100,0

Таблица 3 – Аналитическая группировка

№ п/п	Группы банков по сумме активов, млн.руб.	Количество банков	В среднем на 1 банк	
			Численность занятых, чел.	Балансовая прибыль, млрд. руб.
1	до 20	29	184	22,5
2	20 – 30	8	313	31,6
3	30 – 40	7	374	36,0
4	40 – 50	9	468	69,2
5	50 и более	7	516	205,6
Всего		50	1855	360,0