

Группировка статистических данных

*Выполнила:
студентка 2 курса
группы «УчиА-17»
Жұмаш Арайлым*

Группировка статистических данных

Группировка — это метод, при котором вся исследуемая совокупность разделяется на группы по какому-то существенному признаку.

Признак, по которому осуществляется группировка называется *группировочным признаком* или основанием группировки.

Группировка представляет собой способ подразделения рассматриваемой совокупности данных на однородные по изучаемым признакам группы. Это делается с целью изучения структуры этой совокупности либо взаимосвязей между отдельными элементами этой совокупности. С помощью группировки можно выявить влияние отдельных единиц на средние итоговые показатели.

Три основных типа задач, решаемых с помощью метода статистической группировки:

- 1)** выделение социально–экономических типов явлений;
- 2)** изучение структуры явления и структурных сдвигов, происходящих в явлении;
- 3)** выявление взаимосвязей и взаимозависимостей между явлениями и признаками, характеризующими эти явления.



Виды группировок



Типологическая группировка решает задачу выявления и характеристики социально-экономических типов (частных подсовокупностей).

Структурная дает возможность описать составные части совокупности или строение типов, а также проанализировать структурные сдвиги.

Аналитическая (факторная) группировка позволяет оценивать связи между взаимодействующими признаками.

В зависимости от числа положенных в их основание признаков различают простые и многомерные группировки.

Группировка, выполненная по одному признаку, называется **простой**.

Многомерная группировка производится по двум и более признакам. Частным случаем многомерной группировки является **комбинационная группировка**, базирующаяся на двух и более признаках, взятых во взаимосвязи, в комбинации.

Типологическая группировка

Качественно однородные группы совокупностей, т. е. объекты, которые по своим группировочным признакам близки друг к другу, называют *типологической группировкой*.

Для построения типологической группировки необходимо воспользоваться количественными и качественными (атрибутивными) признаками.

Группировка по атрибутивному признаку предполагает, что число выделенных групп соответствует фактическому числу градаций этого признака. По количественному признаку необходимо правильно установить интервал группировки, определить необходимое число групп. Проблема определения интервалов типологической группировки решается на основании выделения таких количественных границ изменения группировочного признака, при которых явление изменяет или приобретает новое качество.

В типологической группировке от числа существующих социально–экономических типов зависит число групп. От состава, структуры однородных групп и изучения вариации признаков внутри однотипной совокупности и однотипных групп на основе построения структурной группировки зависят социально–экономические типы явлений. Разделение однородной совокупности на определенные группы, которые в дальнейшем будут характеризовать структуру по определенному группировочному признаку, называют *структурной группировкой*.

Структурная группировка

Структурная группировка применяется для характеристики структуры совокупности и структуры сдвигов.

Структурный называется группировка, в которой происходит разделение выделенных с помощью технологической группировки типов явлений, однородных совокупностей на группы, характеризующие их структуру по какому либо варьирующему признаку. Например, группировка населения по размеру среднедушевого дохода. Анализ структурных группировок взятых за ряд периодов или моментов времени, показывает изменения структуры изучаемых явлений, то есть структурные сдвиги. В изменении структуры общественных явлений отражаются важнейшие закономерности их развития.

Показатель численности групп представлен либо частотой (количеством единиц в каждой группе), либо частотностью (удельным весом каждой группы).

Аналитическая группировка

Аналитическая группировка – это группировка, выявляющая взаимосвязи и взаимозависимости между изучаемыми социально–экономическими явлениями и признаками, их характеризующими.

Все признаки в статистической науке можно подразделять на факторные и результативные. Признаки, которые оказывают большое влияние на изменение результативных признаков, называют факторными. Признаки, изменяющиеся под влиянием факторных признаков, называют *результативными*.

Важная задача при построении аналитической группировки – выбор числа групп, на которые необходимо разбить изучаемую совокупность единиц наблюдения, и определение их границ.

Требования, которые необходимо соблюдать в процессе построения аналитических группировок, это: каждая изучаемая группа должна содержать однородные единицы совокупности по группировочному признаку, и количество единиц в каждой изучаемой группе должно быть достаточным для того, чтобы получить статистические характеристики изучаемого объекта.

Простая и комбинированная группировка

Простой называется группировка, если группа образована только по одному признаку. Если разбить группу на подгруппу в соответствии с определенными признаками, то такую группировку называют *комбинированной*.

Комбинационной считается группировка, когда разбивка совокупности на группы производится по двум и более группировочным признакам, взятым в сочетании (комбинации) друг с другом.

При изучении сложных социально–экономических явлений и процессов применяются комбинационные группировки. Для того чтобы построить комбинационную группировку, необходимо выявить наличие достаточно большого числа наблюдений.

Для того чтобы найти скопление (в мерном пространстве) объектов (точек), необходимо применить многомерную группировку.

Различают группировки по используемой информации:

1) первичные – производятся на основе исходных данных которые были получены в результате статистического наблюдения;

2) вторичные – это результат соединения или расчленения группировки.

Ряд распределения

Ряд распределения – это группировка, в которой для характеристики групп применяется один показатель – численность группы. Другими словами, это ряд чисел, показывающий, как распределяются единицы некоторой совокупности по изучаемому признаку.

Ряды, построенные по атрибутивному признаку, называются *атрибутивными рядами распределения*. Ряды распределения, построенные по количественному признаку, называются *вариационными рядами*.

Вариационные ряды распределения состоят из двух элементов вариантов и частот.

Вариантами называются числовые значения количественного признака в ряду распределения, они могут быть положительными и отрицательными, абсолютными и относительными.

Частоты – это численности отдельных вариантов или каждой группы вариационного ряда. Сумма всех частот называется объемом совокупности и определяет число элементов всей совокупности.

Принципы построения группировок

1. Выбор группировочного признака

В зависимости от вида группировочных признаков различают группировки по количественным и качественным (атрибутивным) признакам.

2. Определение числа групп:

Если в основании группировки атрибутивный (качественный) признак, то количество групп равняется количеству значений этого признака

Если в основании группировки лежит количественный признак, то число групп определяют по формуле Стерджесса:

$$n = 1 + 3.322 \lg N = \log_2 N + 1$$

n — число групп

N — число единиц совокупности

Принципы образования интервалов

Интервал группировки — это значение варьирующего признака, лежащее в определенных пределах.

Нижняя граница интервала — это значение наименьшего признака в интервале.

Верхняя граница — это наибольшее значение в интервале.

Величина интервала — это разница между верхней и нижней границами.

Интервалы группировок могут быть равными и неравными.

Равные интервалы применяются в тех случаях, когда значение количественного признака внутри совокупности изменяется равномерно.

Величина равных интервалов определяется по формуле:

$$i = (X_{max} - X_{min}) / n$$

i — величина интервала

X_{max} - максимальное значение признака в совокупности

X_{min} — минимальное значение признака в совокупности

n — число групп

Правила округления интервалов

- Если интервал имеет один знак ДО запятой, то полученное значение округляется до десятых ($0,88 = 0,9$; $8,715 = 8,7$)
- Если величина интервала имеет два знака ДО запятой, то полученное значение округляется до целых ($11,11 = 11$; $29,98 = 30$)
- Если интервал трех, четырех и более значимое число, то интервал принимают кратным 50 или 100
- Интервалы бывают открытые и закрытые. Закрытым считается интервал, в котором есть и нижняя и верхняя границы, в противном случае интервал считается открытым. При решении задач неизвестную границу открытого интервала определяют по величине смежного с ним интервала.

Классификация

От группировок следует отличать классификацию.
Классификация является основой группировок.

Классификацией называется систематизированное распределение явлений и объектов на определенные группы, классы, разряды на основании их сходства и различия. Отличительной чертой классификации является то, что в основу ее кладется качественный признак.

Пример. Распределение предприятий по региона по величине розничного товарооборота в текущем году.

№ п/п	Розничный товароборот, тыс.тг	№ п/п	Розничный товароборот, тыс.тг
1.	151331	16.	21253
2.	56440	17.	47248
3.	99212	18.	92955
4.	34088	19.	178291
5.	43520	20.	68865
6.	38196	21.	9767
7.	208492	22.	60674
8.	104518	23.	23944
9.	82972	24.	127725
10.	45561	25.	24559
11.	137445	26.	21946
12.	28970	27.	44876
13.	51387	28.	117021
14.	156775	29.	33775
15.	65680	30.	36637

Если заранее не задано число групп , то для определения количества групп можно использовать формулу Стерджесса:

$$n=1+3.322\lg N=\log_2 N+1$$

Величину интервала определяют по следующей формуле:

$$h = \frac{R}{n} = \frac{x_{\max} - x_{\min}}{n} = \frac{208492 - 9767}{1 + 3.322 \lg 30} = \frac{198725}{6} \approx 33121$$

Результаты представим в табличной форме.

Группы предприятий по размеру розничного товарооборота, тыс.тг	Число предприятий	№ предприятия	Общий товарооборот по группе, тыс.тг	Средний размер товарооборота по группам, тыс.тг
9767 – 42888	10	4, 6, 12, 16, 21, 23, 25, 26, 29, 30	273135	27313,5
42888 – 76009	9	2, 5, 10, 13, 15, 17, 20, 22, 27	484251	23805,7
76009 – 109130	4	3, 8, 9, 18	379657	94914,3
109130 – 142251	3	11, 24, 28	382191	127397
142251 – 175372	2	1, 14	308106	154053
175372 – 208493	2	7, 19	386783	193391,5
Итого:	30		2214123	73804,1

Вторичная группировка

На практике иногда приходится пользоваться уже имеющимися группировками, которые могут быть несопоставимы из-за неодинаковых границ интервалов или различного количества выделяемых групп. Для приведения таких группировок к сопоставимому виду используется метод вторичной группировки.

Вторичная группировка заключается в образовании новых групп на основе ранее произведенной группировки.

Во вторичной группировке применяются два способа образования новых групп:

Первый способ состоит в укреплении первоначальных интервалов. Это наиболее простой и распространенный способ вторичной группировки.

Второй способ называется методом долевого перегруппировки и состоит в том, что за каждой группой закрепляется определенная доля единиц совокупности.