**Группировка статистических данных**

1. **Понятие о статистической сводке и группировке.**

Изучая массовые общественные явления, статистика собирает сведения о каждой единице совокупности но не для характеристики отдельной единицы, а для получения обобщающих показателей по совокупности в целом. Например, при переписи населения получают сведения о каждом человеке (пол, возраст, образование, национальность и др.). Однако сбор этих сведений не имеет целью изучение возраста или образования отдельно каждого человека: их используют для получения обобщающих показателей по стране в целом о численности населения и его составе.

Чтобы полученные в результате статистического наблюдения данные могли быть использованы для характеристики изучаемой совокупности в целом, они должны быть научно обработаны, приведены в определенную систему, подсчитаны, обобщены. Такая систематизация и подсчет статистических данных является статистической сводкой. Цель сводки заключается в обобщении результатов статистического наблюдения. Например, сводка отчетных данных предприятий местной промышленности:

План перевыполнен на 10000 тыс. грн. Такая сводка называется простой, т.к. для получения итоговых показателей ограничились суммированием показателей по каждому предприятию.

Простая сводка дает видимость благополучия: план по управлению в целом перевыполнен, а в действительности, не все предприятия выполнили план. Чтобы выделить эти предприятия, отделить их от предприятий, выполнивших план, нужно все предприятия разделить на две группы:

а) выполнившие план реализации продукции;

б) не выполнившие план реализации.

|  |  |  |
| --- | --- | --- |
| НаименованиеПредприятий | Объем реализованной продукции(тыс.грн.) | |
| По плану | По факту |
| Кирпичный завод  Леспромхоз  Стекольный завод  Цементный завод  «Стройдеталь»  «Промтара»  Фанерная тара  Мебельная фабрика  Хлебозавод  Мелькомбинат | 20000  42000  25000  71000  35000  104000  93000  61800  84000  113000 | 22000  43200  25000  71200  34800  105100  92300  60700  86000  119500 |
| Итого | 649800 | 659800 |

В этом случае сводка примет следующий вид:

Таблица 2. Группировка предприятий по выполнению плана реализации продукции.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Группы предприятий | Число предприятий | Объем реализованной продукции | | Реализовано сверх или недодано |
| По плану | По факту |
| Выполнившие и перевыполнившие план | 6 | 434000 | 447000 | +13000 |
| 1 | 2 | 3 | 4 | 5 |
| 1 | 2 | 3 | 4 | 5 |
| Не выполнившие  план | 4 | 215800 | 212800 | -3000 |
| Итого | 10 | 649800 | 659800 | +10000 |

Произведенная подобным образом сводка показывает, что, хотя по управлению в целом план выполнен, значительная часть предприятий (4 из 10) не выполнила его, из-за чего недодано народному хозяйству продукции на 3000 тыс. грн. Если бы не была проведена группировка, то за «благополучными данными» по управлению в целом была бы скрыта плохая работа отстающих предприятий.

Такая сводка, включающая группировку данных, называется групповой. Как правило, статистическая сводка включает группировку данных, характеристику каждой группы системой показателей и подсчет групповых (по каждой группе) и общих (итоговых) показателей.

Сводка данных является одним из основных этапов статистической работы. От организации сводки в значительной мере зависит успех всего исследования. Статистическая сводка может дать большой материал, правильно характеризующий каждую единицу совокупности, но если сводка его произведена неправильно, но выводы, характеризующие совокупность в целом и ее отдельные группы, будут ошибочны.

Сводка собранных сведений должна быть произведена таким образом, чтобы наиболее существенные различия между отдельными группами явлений выделялись с целью их изучения.

Следовательно, научное исследование массовых явлений невозможно без разграничения групп, существенно различающихся межде собой, и объединения в группы явлений, сходных в существенном отношении. Поэтому статистическая группировка является основой научной сводки.

Статистическая группировка представляет собой расчленение совокупности общественных явлений на однородные типичные группы по существенным для них признакам.

Метод группировок является основой для научного применения других методов статистики – статистики, индексов и т.д. В развитие метода группировок огромный вклад внесли отечественные статистики. Им принадлежит первенство в применении комбинационных таблиц, в разработке классификации таблиц, и в проведении многочисленных группировок материалов аграрных переписей и обследований, которые оказали благотворное влияние на другие отраслевые статистики.

**2. Задачи и виды группировок.**

Приступая к группировке статистического материала, нужно четко определить ее цели и задачи. Важнейшей задачей статистических группировок является выделение существующих в действительности общественно экономических типов явлений.

Группировка, в результате которой выделяются социально-экономические типы, называется типологической. Придерживаясь этого требования, статистика дает возможность видеть общественные явления такими, какие они есть, во всей их сложности, чтобы за грудами цифр не исчезли экономические типы явлений. При сводке данных переписей (или учетов) скота, выделяют скот, принадлежащий различным категориям (типам) хозяйств.

Примерами типологических группировок могут служить группировки хозяйств по формам собственности; населения – по общественным группам; работников – на занятых физическим или умственным трудом.

Кроме типологической группировки существуют также аналитическая и структурная группировка. Аналитическая группировка – группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками.

Группировка структурная – группировка, выявляющая состав, строение однородной в качественном отношении совокупности статистической по определенным признакам. Например, состав населения по полу, возрасту, уровню образования, роду занятий и т.д.

Сопоставление данных структурных группировок во времени дает представление о структурных сдвигах.

Многообразию общественных явлений, изучаемых статистикой, соответствует и многообразие применяемых статистикой группировок. В зависимости от сущности изучаемых явлений с задач исследования необходимы различные приемы группировок:

* по разным признакам, с различными интервалами, с большим или меньшим числом выделяемых групп.
* Различными могут быть показатели, применяемые для характеристики каждой группы.

Например, при группировке отраслей промышленности по экономическому назначению продукции, выделяются отрасли, производящие средства производства, и отрасли, производящие предметы потребления. В большинстве случаев качественные отличия не выступают столь отчетливо. Например, выделение в отраслях промышленности крупных, средних и мелких предприятий является достаточно сложной в методологическом отношении задачей. В подобных случаях после предварительной наметки возможных типов на основе четкой формулировки задач необходимо определить те признаки, которые положены в основу выделения типов – так называемые группировочные признаки.

Группировка может производится как по одному так и по нескольким признакам одновременно. Группировку по одному признаку называют простой или одномерной, а группировку по двум или нескольким признакам – комбинационной или многомерной. Выбор группировочных признаков всегда должен быть основан на анализе количественной природы исследуемого явления.

Использование одного признака может привести к искажению действительности, поскольку в ней, как правило, переплетаются противоположные тенденции и направления. Множественность признаков, характеризующих объекты, является следствием многообразия реальных связей между объектами.

В зависимости от вида группировочных признаков различают группировки по количественным и качественным признакам. Качественный – отражает определенные свойства, качества данного явления и записывается в виде текста.

Если качественный признак имеет мало разновидностей, то количество групп определяется числом этих разновидностей. Например, группировки населения по полу, семейному положению, образованию, деление на сельское и городское. Но нередки случаи, когда качественный признак может иметь большое число разновидностей и перечислить их все не возможно. Например, профессии рабочих, номенклатура выпускаемой продукции, виды основных фондов. В таких случаях разрабатывают классификацию разновидностей, т.е. сходные по основным особенностям разновидности объединяются в группы (классы).

Под классификацией обычно понимается более устойчивое разграничение единиц наблюдения, чем при группировке.

Кроме группировок по качественным признакам часто приходится производить группировку по качественным признакам. Например, группировка населения по возрасту, рабочих по разрядам, по степени выполнения норм и т.д. В ряде случаев группировки, на первый взгляд, качественные, в действительности основываются на количественных признаках. Например, при группировке детей по возрасту выделяют: 1) ясельный возраст; 2) дошкольный; 3) школьный. При отнесении в ту или иную группу руководствуются тем, что в яслях дети находятся до 3 лет, а в дошкольную группу относят детей в возрасте до 6-7 лет и, наконец, в третью группу дутей возрастом до 17 лет.

При группировке по количественному признаку нужно установить количество групп, на которые следует разбить весь диапазон изменения количественного признака, и в соответствие с числом групп определить интервалы группировки.

**3. Техника выполнения группировок.**

Группировка, показывающая распределение элементов совокупности по какому-то одному признаку, называется рядом распределения. Если ряд распределения построен по признаку, имеющему качественное выражение, то он является вариационным рядом. Вариационные ряды бывают интервальными и дискретными.

Интервальный ряд содержит распределение совокупности явлений на группы, ограниченные определенными интервалами. (от – до).

Примером интервального ряда может быть распределение колхозов и совхозов на группы по урожайности сельскохозяйственных культур, по размерам валового дохода на 100 га пашни. Рассмотрим пример группировки по урожайности.

Таблица 1. Распределение колхозов района по урожайности.

|  |  |  |
| --- | --- | --- |
| Урожайность озимой пшеницы с 1 га | Число колхозов | Удельный вес в% |
| До 15  15-20  20-25  25-30  30 и выше | 2  6  18  9  5 | 5,0  15,0  45,0  22,5  12,5 |
| Итого | 40 | 100 |

Предполагается, что группировочный признак может принимать любое значение в пределах каждого интервала.

Дискретный ряд отличается тем, что в нем группировочный признак не может принимать промежуточных значений. Примером такого ряда служит распределение магазинов по числу рабочих мест в них:

|  |  |
| --- | --- |
| Число рабочих мест | Число магазинов |
| 1  2  3 | 6  85  140 |

Ясно, что группировочный признак не может иметь здесь дробных значений.

В таблице 2 представлены данные по использованию производственного оборудования, занятого в основном производстве на машиностроительных предприятиях.

Таблица 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Группы Предприятий по кол-ву установл.ст. | Кол-во Предпр | Кол-во установленного оборудования тыс.ед. | Удельный вес работавшего оборудования% к установл | Ксм работы оборудования |
| До 50  51-100  101-200  201-500  501-1000  1001-2000  2001 и выше | 173 300  537  867  512  241  142 | 6,4  26,3  82,9  277,9  350,1  310,4  488 | 80  82  83  85  85  85  85 | 1,2  1,26  1,30  1,38  1,40  1,41  1,48 |

Как видно из таблицы 2, группы предприятий намечены с помощью неравных интервалов, т.е. интервалов, у которых разности между верхней и нижней границами неодинаковы в разных группах. Во второй группе – 50 ед., в третьей – 100, в четвертой – 300 и т.д., т.е. в приведенной группировке величина интервала постепенно увеличивается, а в последней группе верхняя граница не указывается совсем. Интервалы, в которых указана лишь одна граница (верхняя или нижняя), называются открытыми, остальные интервалы являются закрытыми.

Неравные интервалы применяются при группировках, которые охватывают массу единиц неоднородной совокупности с неравномерными и значительными колебаниями признака.

Внутри типичных групп для характеристики количественных различий единиц, составляющих соответствующую группу, могут быть применены равные интервалы. В этом случае величина интервала определяется делением размаха варьирования на принятое число групп.

Выше рассмотрены примеры группировок, произведенных на основе непосредственного обобщения первичных данных статистического наблюдения. Такие группировки можно назвать первичными. Однако на практике иногда приходится перегруппировать раннее сгруппированный материал для обеспечения сопоставимости данных двух или нескольких группировок, увеличения (укрупнения) или уменьшения интервалов и т.д. Такие группировки принято называть вторичными. Укрупнение интервалов – наиболее простой прием вторичной группировки.

Величина интервала i=()/h, где

hmax,hmin – максимальное и минимальное значение пизнака;

h – число групп, определяется по формуле Стэрджесса,

h=1+3,332ln(n), где n – число изучаемых единиц

совокупности.

**4.Статистические таблицы, их виды и принципы построения.**

Статистические таблицы являются средством оформления результатов сводки и группировки, а также анализа статистических данных. Статистические таблицы дают возможность представить результаты сводки и группировки в таком виде, который облегчает их чтение и анализ. Без статистических таблиц пришлось бы сопровождать каждый показатель громоздкими пояснениями, а с помощью статистических таблиц статистические материалы располагаются в определенном порядке, удобном для их сравнения между собой и для исчисления производных показателей. Чтобы соответствовать своему назначению, таблица должна быть по возможности небольшой, компактной, и, как говорят, удообозримой.

По форме таблица состоит из горизонтальных строк и вертикальных граф, пересечения которых образуют клетки, в которых размещаются показатели. Поэтому каждому показателю в таблице соответствует два заголовка: вверху – наименование граф, слева – наименование строк. Поскольку каждый заголовок вверху относится ко всем показателям данной графы, а каждый заголовок слева – ко всем показателям данной строки, то это дает возможность избежать повторения наименований и обозначений после каждого показателя.

По своему содержанию статистическая таблица напоминает грамматическое предложение: в ней приводятся наименования изучаемых явлений (о ком или о чем говорится в таблице) и показатели, характеризующие состояние или развитие этих явлений (что именно говорится об изучаемом явлении). Поэтому в статистических таблицах, как и в грамматическом предложении, принято выделять подлежащее и сказуемое.

Подлежащим таблицы является перечень явлений или их групп, которые характеризуются показателями таблицы. Например, в таблице «зависимость себестоимости молока от продуктивности коров», подлежащее – группы хозяйств по годовому надою. Сказуемым таблицы являются показатели, характеризующие подлежащее.

Таблица. Макет статистической таблицы.

Общий заголовок.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Сказуемое  Подлежащее | Верхние внутренние заголовки | | | | |
| А | 1 | 2 | 3 | 4 | 5 |
| Боковые заголовки |  |  |  |  |  |
| Итог |  |  |  |  |  |

* Примечание к таблице (используется только тогда когда необходимо специально пояснить прием расчета, указать источник данных, оговорить специфический характер отдельных показателей.

Пример.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Группы хозяйств по годовому удою на корову, ц | Число хозяйств в группе | Среднегодовой надой на одну корову, ц | Себестоимость 1ц молока, руб. | Расход кормов на 1ц молока |
| До 40  40-50  45-50  свыше 50  в среднем | 17  26  43  14  100 | 37,2  42,4  48,1  53,3  45,5 | 313  275  248  225  263 | 1,39  1,31  1,22  1,08  1,25 |

По характеру подлежащего различают таблицы трех видов: простые, групповые, комбинационные. Подлежащее простой таблицы представляет собой перечень изучаемых явлений:

Таблица. Показатели обеспеченности населения медицинской помощью на конец года.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Показатели | 1940 | 1950 | 1960 | 1972 |
| Численность врачей, всего  Число врачей на 10000 человек населения  Число больничных коек на 10000 человек населения | 155,3  7,9  40,2 | 265  14,6  56,0 | 431,7  20  80,4 | 731,8  29,4  112,3 |

Таблица по своему построению является простой, так как в подлежащем ее нет группировки. Простые таблицы широко распространены в статистике, часто они недостаточны для анализа, поскольку не содержит группировки статистических данных.

Таблицы, которые содержат в подлежащем группировку, называются групповыми:

Таблица. Выпуск специалистов по отраслевым группам «Высших учебных заведений», тыс.чел.

|  |  |  |  |
| --- | --- | --- | --- |
| Группы ВУЗов | 1960 | 1972 | 1986 |
| Промышленности и строительства  Транспорта и связи  Сельского хозяйства  Экономики и права  Здравоохранения, физкультуры и спорта  Просвещения  Искусства и кинематографии | 95,2  16,1  34,7  25,0  30,7  139,1  2,5 | 250,4  31,2  59,6  54,5  53,3  227,6  7,7 | 510,5  64,1  120,3  104,1  132,3  478,5  12,2 |
| Всего | 343,3 | 684,3 | 1422,1 |

Еще больше возможностей для углубленного анализа статистического материала дают таблицы, содержащие результаты сложной группировки по двум или нескольким признакам. Такие таблицы называются комбинационными:

Таблица. Состав учащихся общеобразовательных школ на начало 1996/1997 учебного года.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Распределение учащихся по группам классов | Численность учащихся, млн. | Из них | | | |
| В городских поселениях | | В сельской местности | |
| Млн. | % | Млн. | % |
| 1-3  4-8  9-10  итого | 13,9  25,4  5,1  44,4 | 6,8  13,0  2,9  22,7 | 49  51  57  51 | 7,1  12,4  2,2  21,7 | 51  49  43  49 |

В приведенной таблице учащиеся общеобразовательных школ распределены на группы по двум признакам:

1. по группам классов;
2. по месту нахождения.

Число учащихся в городских поселениях больше, чем в сельской местности за счет учащихся старших классов. Число учащихся в 1-3 классах в сельской местности больше. Если бы группировка данных производилась по каждому признаку отдельно, этот факт мог остаться незамеченным.

Построению таблицы предшествует подбор макета для нее. Рекомендуется составить несколько макетов с целью выбора одного из них, в наибольшей мере соответствующего содержанию материала.

При составлении таблиц необходимо придерживаться следующих правил:

1. Таблица должна быть по возможности небольшой но размеру, включать только те данные которые необходимы для изучения данного явления. Такую таблицу проще читать и анализировать.
2. Общий заголовок, заголовки подлежащего и сказуемого должны формулироваться четко и коротко.
3. Если число показателей сказуемого велико, их необходимо пронумеровать. При этом графы, в которых содержится перечень объектов или групп обозначаются большими буквами алфавита, а графы с показателями сказуемого – арабскими цифрами.
4. Если нет сведений о размере явления, то в соответствующей клетке делается отметка «нет сведений» или ставятся точки (…). В случае невозможности заполнения какой-либо клетки, ввиду отсутствия соответствующего явления, делается прочерк (–).
5. Показатели каждой графы должны приводится с одинаковой степенью точности, т.е. до 1; 0,1; 0.01 и т.д.
6. Если приведенные показатели имеют различные единицы измерения, им выделяют специальную графу.
7. Таблицы должны быть замкнутыми, т.е. с итоговыми результатами.