**Дата:** 20.03.2020.

**Тема:** Средние показатели и показатели вариации

**Литература:** Статистика: учеб. для студ. учреждений СПО/[В.С.Мхиторян, Т.А.Дуброва, В.Г.Минашкин и др.]; под ред. В.С.Мхиторян. –М.: Издательский центр «Академия»

**Глава 6**

**Краткие сведения по теме:**

**Средние показатели**

**Главное значение средних величин** состоит в их обобщающей функции, то есть замене множества различных индивидуальных значений признака средней величиной, характеризующей всю совокупность явлений.

Виды средних величин различаются, прежде всего, тем, какое свойство, какой параметр исходной варьирующей массы индивидуальных значений признака должен быть сохранен неизменным.

Средней арифметической величиной называется такое среднее значение признака, при вычислении которого общий объем признака (сумма значений признака) в изучаемой совокупности сохраняется неизменным. Иначе можно сказать, что средняя арифметическая величина – это среднее слагаемое, то есть при ее вычислении общий объем (сумма всех значений) признака мысленно распределяется поровну между всеми единицами совокупности. Исходя из определения, формула средней арифметической величины имеет вид $\overbar{Х}=\frac{х\_{1}+х\_{2}+х\_{3}+…+х\_{n}}{n} $**=**$ \frac{\sum\_{}^{}x}{n}$

По этой формуле вычисляются средние величины первичных признаков, если известны индивидуальные (отдельные) значения признака. Если изучаемая совокупность велика, то исходная информация чаще представляет собой ряд распределения или группировку, как, например, следующая таблица, где приведен условный пример дискретного ряда распределения студентов по возрасту:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Возраст, Х | 17 | 18 | 19 | 20 | 21 |
| Число студентов, f | 3 | 5 | 7 | 4 | 2 |

Средний возраст должен представлять собой результат равномерного распределения общего (суммарного) возраста всех студентов. Общий (суммарный) возраст всех студентов, согласно исходной информации в вышеприведенной таблице, можно получить как сумму произведений значений признака в каждой группе Xi, на число студентов с таким возрастом fi (частоты). Получим формулу:

$$\overbar{Х}=\frac{\sum\_{i=1}^{n}x\_{i}f\_{i}}{\sum\_{i=1}^{n}f\_{i}}$$

Такую форму средней арифметической величины называют ***взвешенной*** арифметической средней. В качестве весов здесь выступают количество единиц совокупности (fi) в разных группах. Название «вес» выражает тот факт, что разные значения признака имеют неодинаковую «важность» при расчете средней величины. «Важнее», весомее возраст студентов 18, 19, 20 лет, а такие значения возраста как 17, 20 или 21 при расчете средней не играют большой роли – их «вес» мал. По формуле средней арифметической взвешенной по данным в условном примере получим:

$\overbar{Х}=\frac{17∙3+18∙5+19∙7+20∙4+21∙2}{21}=\frac{396}{21} $=18,857 лет

Как видим, средняя арифметическая величина может быть дробным числом, если даже индивидуальные значения признака могут принимать только целые значения. Ничего необычного для метода средних в этом не заключено, так как из сущности средней не следует, что она обязана быть реальным значением признака, которое могло бы встретиться у какой-либо единицы совокупности.

Если при группировке значения осредняемого признака заданы интервалами, то при расчете средней арифметической величины в качестве значения признака в группах принимают середины этих интервалов, то есть исходят из предположения о равномерном распределении единиц совокупности по интервалу значений признака. Для открытых интервалов в первой и последней группе, если таковые есть, значения признака надо определить экспертным путем исходя из сущности, свойств признака и совокупности. При отсутствии возможности экспертной оценки значения признака в открытых интервалах, для нахождения недостающей границы открытого интервала применяют размах (разность между значениями конца и начала интервала) соседнего интервала (принцип «соседа»).

Например, по данным следующей таблицы минимальную и максимальную величину веса студентов определить затруднительно, поэтому воспользуемся принципом «соседа» – применим размах соседнего интервала, который у второго и предпоследнего составляет 10 кг, значит первый интервал будет от 55 до 65 кг, а последний – от 80 до 90 кг. Середины интервалов определяем как полусумму нижней и верхней границы интервалов.

|  |  |  |  |
| --- | --- | --- | --- |
| Группы студентов по весу, кг | Количество студентов, чел. | Середина интервала X | Xf |
| до 60 | 6 | 55 | 330 |
| 60 – 70 | 8 | 65 | 520 |
| 70 - 80 | 5 | 75 | 375 |
| более 80 | 5 | 75 | 170 |
| Итого | 21 | 66,429 | 1395 |

Средний вес студентов, рассчитанный по формуле средней арифметической взвешенной с заменой точных значений признака в группах серединами интервалов Xi, составит частное от деления итогов последнего и второго столбцов таблицы: = 1395/21 = 66,429 (кг).

Полученное значение записано в итоговую строку таблицы в 3-м столбце.

Следует обратить внимание, что объемного показателя – это сумма, а итог по столбцам относительных показателей или средних групповых величин – средняя.

Средняя арифметическая величина обладает 5 *свойствами*, знание которых полезно как при ее использовании, так и при ее расчете.

1. Сумма отклонений индивидуальных значений признака от его среднего значения равна нулю.
2. Если каждое индивидуальное значение признака умножить или разделить на постоянное число, то и средняя увеличится или уменьшится во столько же раз. Вследствие этого свойства индивидуальные значения признака можно сократить в *c* раз, произвести расчет средней и результат умножить на *c*.
3. Если к каждому индивидуальному значению признака прибавить или из каждого значения вычесть постоянное число, то средняя величина возрастет или уменьшится на это же число. Это свойство полезно использовать при расчете средней величины из многозначных и слабоварьирующих значений признака аналогично предыдущему свойству.
4. Если веса средней взвешенной умножить или разделить на постоянное число, средняя величина не изменится. Используя это свойство, при расчетах следует сокращать веса на их общий сомножитель либо выражать многозначные числа весов в более крупных единицах измерениях.
5. Сумма квадратов отклонений индивидуальных значений признака от средней арифметической меньше, чем от любого другого числа.

Если при замене индивидуальных величин признака на среднюю величину необходимо сохранить неизменную сумму квадратов исходных величин, то средняя будет являться **квадратической** средней величиной. Ее формула следующая:

Главной сферой применения квадратической средней в силу пятого свойства средней арифметической величины является измерение вариации признака в совокупности.

Аналогично, если по условиям задачи необходимо сохранить неизменной сумму кубов индивидуальных значений признака при их замене на среднюю величину, мы приходим к средней кубической величине, имеющей вид:

**

Если при замене индивидуальных величин признака на среднюю величину необходимо сохранить неизменным произведение индивидуальных величин, то следует применить **геометрическую** среднюю величину, имеющую следующий вид:

Основное применение средняя геометрическая находит при определении средних относительных изменений, о чем сказано в теме 6. Геометрическая средняя величина дает наиболее точный результат осреднения, если задача стоит в нахождении такого значения признака, который качественно был бы равноудален как от максимального, так и от минимального значения признака.

Когда статистическая информация не содержит частот f по отдельным вариантам Xi совокупности, а представлена как их произведение Xf, тогда применяется формула **средней гармонической взвешенной**, для получения которой обозначим Xf=w, откуда f=w/X, и, подставив эти обозначения в формулу средней арифметической взвешенной, получим формулу средней гармонической взвешенной:

Все рассмотренные выше виды средних величин принадлежат к общему типу степенных средних, имеющему следующий вид:

При m = 1 получаем среднюю арифметическую; при m = 2 – среднюю квадратическую; при m = 3 – среднюю кубическую; при m = 0 – среднюю геометрическую; при m = –1 – среднюю гармоническую. Чем выше показатель степени m, тем больше значение средней величины (если индивидуальные значения признака варьируют). В итоге, можно построить следующее соотношение, которое называется правилом мажорантности средних:



**Показатели вариации в статистике**

Признаки, изучаемые статистикой, варьируются (отличаются друг от друга) у различных единиц совокупности в один и тот же период или момент времени. Например, варьируется рост людей, их заработная плата т.п.

Причиной вариации являются разные условия существования разных единиц совокупности. Например, огромное число причин влияет на рост человека, его заработную платы и т.д.

Для управления и изучения вариации статистикой разработаны специальные методы исследования вариации, система показателей, с помощью которой вариация измеряется, характеризуются ее свойства.

**Первым этапом** статистического изучения вариации является построение *ряда распределения* (или вариационного ряда) – упорядоченного распределения единиц совокупности по возрастающим (чаще) или по убывающим (реже) значениям признака и подсчет числа единиц с тем или иным значением признака.

Ряд распределения бывает *дискретным* и *интервальным*.

**Дискретный ряд распределения** – это таблица, состоящая из двух столбцов (строк) – конкретных значений варьирующего признака Xi и числа единиц совокупности с данным значением признака fi – частот; число групп в дискретном ряду определяется числом реально существующих значений варьирующего признака. В следующей таблице приведен пример дискретного ряда распределения:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вес студента, кг | 48 | 50 | 53 | 55 | 56 | 59 | 62 | 64 | 68 | 70 | 72 | 77 | 85 | 88 | Итого |
| Кол-во студентов, чел. | 1 | 3 | 2 | 1 | 1 | 2 | 3 | 2 | 2 | 3 | 5 | 2 | 2 | 1 | 30 |

*Интервальный ряд распределения* – это таблица, состоящая из двух столбцов (строк) – интервалов варьирующего признака Xi и числа единиц совокупности, попадающих в данный интервал (частот - fi), или долей этого числа в общей численности совокупностей (частостей - di).

Трансформируем дискретный ряд, представленный в таблице выше, в интервальный ряд распределения. Для этого необходимо выбрать оптимальное число групп (интервалов признака) и установить длину (размах) интервала. Поскольку при анализе ряда распределения сравнивают частоты в разных интервалах, необходимо, чтобы длина интервалов была постоянной. Если приходится иметь дело с интервальным рядом распределения с неравными интервалами, то для сопоставимости нужно частоты (f) или частости (d) привести к единице интервала, полученное значение называется *плотностью* ρ, то есть ρ = f/h.

Оптимальное число групп выбирается так, чтобы в достаточной мере отразилось разнообразие значений признака в совокупности и, в то же время, закономерность в распределении, а его форма не искажалась случайными колебаниями частот. Если групп будет слишком мало, то не проявится закономерность вариации, а если групп будет чрезмерно много, то случайные скачки частот исказят форму распределения.

Чаще всего число групп в ряду распределения определяют по **формуле Стерждесса**: 

где k – число групп (округляемое до ближайшего целого числа); N – численность совокупности.

Из формулы Стерджесса видно, что число групп k – это функция объема данных (N).

Зная число групп, рассчитывают длину (размах) интервала по формуле:

где Xмax и Xmin — максимальное и минимальное значения в совокупности.

В нашем примере про вес студентов по формуле Стерждесса определим число групп: k = 1 + 3,322lg30 = 1+ 3,322\*1,477 = 5,907. Так как число групп не может быть дробным, то необходимо округлить до ближайшего целого числа полученное значение 5,907. Таким образом получим k = 6.

Рассчитаем длину (размах) интервала: h = (88 – 48)/6 = 40/6 = 6,667 (кг).

Теперь построим интервальный ряд студентов по весу с 6 группами с интервалом 6,667 кг.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| i | 1 | 2 | 3 | 4 | 5 | 6 | Итого |
| Вес, кг | 48 - 54,667 | 54,667 - 61,333 | 61,333 - 68 | 68 - 74,667 | 74,667 - 81,333 | 81,333 - 88 | - |
| Число студентов, чел. | 6 | 4 | 7 | 8 | 2 | 3 | 30 |

Примечание к таблице: единицы совокупности, имеющие значение признака, равное границе интервала (в нашем примере это вес 68 кг), включаются в тот интервал, где это точное значение впервые указывается (то есть в интервал от 61,333 до 68, а в следующий интервал от 68 до 74,667 - не включается).

При изучении вариации применяются такие характеристики ряда распределения, которые описывают количественно его структуру, строение. Такова, например, **медиана** – величина варьирующего признака, делящая совокупность на две равные по численности части (со значением признака меньше медианы и со значением признака больше медианы).

При рассмотрении [дискретного ряда](http://чалиев.рф/statistics/lection7-pokazateli-variatsiy.html#diskretny) медиана определяется суммированием частот ранжированного ряда до N/2, то есть в нашем примере про студентов - до 30/2 = 15. Значение X, отделающее первые 15 студентов от других 15, может приходиться на конкретное значение X, которое и будет медианой, или между двумя значениями X - тогда медианой будет их полусумма.

В вышеприведенном примере медианным интервалом является 3-ий (от 61,333 до 68), так как накопленная сумма частот *f*' до него 6+4=10, а вместе с ним - 6+4+7=17, что больше половины всех частот 30/2=15.

В интервальном ряду распределения для нахождения медианы применяется формула:



где     *X*0 - нижняя граница интервала, в котором находится медиана;

*h* - размах медианного интервала (разность между его верхней и нижней границей);

*f'*Me-1 - накопленная частота в интервале, предшествующем медианному;

*f*Me – частота в медианном интервале.

**Задания:**

* 1. **Изучить главу 6, пункты 6.1-6.5 и составить словарь тематических терминов.**
	2. **Выполнить тест и задачи**

**Сдать 21.03.2020**





