Слайд 1. Тематический план лекции

Тема 5. Средние величины

План

1. Сущность и значение средней величины

2. Средняя арифметическая и ее свойства

3. Виды степенных средних

4. Структурные средние величины

1.

Статистика изучает массовые явления и процессы, каждое из которых обладает как общими для всей совокупности, так и особенными, индивидуальными свойствами. Различие между индивидуальными явлениями называют *вариацией*. *Взаимодействие элементов совокупности приводит к ограничению вариации хотя бы части их свойств.* Эта тенденция существует объективно, именно в ее объективности заключена причина широчайшего применения средних величин на практике и в теории.

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

На производство одного и того же количества товара определенного вида и качества разные производители (заводы, фирмы) затрачивают неодинаковое количество труда и материальных ресурсов. Но рынок осредняет эти затраты, и стоимость товара определяется средним расходом ресурсов на производство.

Если средняя величина обобщает качественно однородные значения признака, то она является типической характеристикой признака в данной совокупности. Так, можно говорить об измерении типичного роста русских девушек рождения 1986г. по достижении ими 20-летнего возраста.

Однако неправильно сводить роль средних величин только к характеристике типичных значений признаков в однородных по данному признаку совокупностях. На практике значительно чаще современная статистика использует средние величины, обобщающие явно неоднородные явления (как, например, урожайность всех зерновых культур по территории всей России, включая кукурузу и гречиху, и плодородные черноземы Кубани, и скудные почвы Архангельской области).

Средняя величина национального дохода на душу, средняя урожайность зерновых по всей стране, среднее потребление разных продуктов питания – это характеристики государства как единой народнохозяйственной системы, это так называемые *системные средние.*

Системные средние могут характеризовать как пространственные или объектные системы, существующие одномоментно (государство, отрасль, регион и т.п.), так и динамические системы, протяженные во времени (год, десятилетие, сезон и т.п.).

Типическая средняя может обобщать системные средние для однородной совокупности, или системная средняя может обобщать типические средние для единой, хотя и неоднородной, системы.

2.

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

Наиболее распространенным видом средних является *средняя арифметическая величина, под которой понимается такое среднее значение признака, при вычислении которого общий объем признака в совокупности сохраняется неизменным.*

В зависимости от характера исходных данных средняя арифметическая определяется следующим образом:

1. если дан ряд одиночных значений признака, то рассчитывается *средняя арифметическая простая*:

Слайд 2а. Средняя арифметическая простая

$$\bar{x} = (x_1 + x_2 + \ldots + x_n) : n = \frac{\sum_{i=1}^{n} x_i}{n},$$

где $\tilde{o}$ – средняя величина;

n – численность совокупности.

2. если расчет осуществляется по сгруппированным данным или вариационным рядам, то применяется *средняя арифметическая взвешенная*:

Слайд 2б. Средняя арифметическая взвешенная

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{\sum_{i=1}^{n} f_i},$$

где f_i – вес усреднения.

Название «вес» выражает тот факт, что разные значения признака имеют неодинаковую «важность» при расчете средней величины. Весом может быть:

1) частота повторения индивидуальных значений признака;

2) частость (q_i);

3) объемный показатель, логически связанный с усредняемым признаком (X);

4) доля объемного показателя в его суммарном объеме по совокупности в целом.

Простая и взвешенная средние величины различаются не только по величине (не всегда), по способу вычисления, но и по своей роли в решении различных задач статистического анализа. Например, рассмотрим среднюю величину урожайности картофеля в группе хозяйств. Если эта средняя при решении поставленной задачи входит в систему показателей площади посадки, валового сбора, себестоимости и других характеристик, то следует применять взвешенную среднюю. Если же задачей является измерение вариации урожайности между хозяйствами, то следует применять простую среднюю величину урожайности.

Если при группировке значения осредняемого признака заданы интервалами, то при расчете средней арифметической величины в качестве значения признака в группах принимают середины этих интервалов. Для открытых интервалов в первой и последней группе, если таковые есть, значения признака надо определить экспертным путем исходя из сущности, свойств признака и совокупности. Например, можно минимальный возраст рабочих считать 17 лет, а максимальный возраст – 65 лет.

Слайд 3а. Средняя арифметическая для интервального ряда

$$\bar{x}_i = \frac{x_i^B + x_i^H}{2},$$

где $\bar{x}_i$ – серединное значение i-го интервала;

x_i^B, x_i^H – верхняя и нижняя граница i-го интервала.

$$\bar{x} = \frac{\sum_{i=1}^{n} \bar{x}_i \cdot N_i}{\sum_{i=1}^{n} N_i},$$

где N_i – численность *i*-й группы.

Средняя арифметическая обладает рядом свойств.

Сущностные свойства средней арифметической:

1) средняя арифметическая постоянной величины равна этой постоянной;
2) алгебраическая сумма линейных отклонений (разностей) индивидуальных значений признака от средней арифметической равна нулю:

Слайд 4а. 2-е свойство средней арифметической

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0 \qquad \text{и} \qquad \sum_{i=1}^{n}(x_i - \bar{x}) \cdot f_i = 0.$$

Это означает, что все отклонения от средней в ту и другую сторону (положительные и отрицательные), обусловленные случайными причинами, взаимно погашаются;

3) сумма квадратов отклонений индивидуальных значений признака от средней арифметической есть величина минимальная:

Слайд 4б. 3-е свойство средней арифметической

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

Вычислительные свойства средней арифметической:

1) если все значения признака уменьшить (увеличить) на одну и ту же величину, то и средняя арифметическая уменьшится (увеличится) на ту же самую величину;
2) если все значения признака разделить (умножить) на какое-либо постоянное число *A*, то средняя арифметическая уменьшится (увеличится) в *A* раз;

3) если вес каждого значения признака разделить на какое-либо постоянное число *A,* то средняя арифметическая не изменится.

3.

Степенные средние делятся на простые и взвешенные. Общая формула простой степенной средней записывается следующим образом:

Слайд 5а. Простая степенная средняя

$$\overline{X} = \left(\frac{\sum_{i=1}^{N} X_i^k}{N} \right)^{\frac{1}{k}} = \sqrt[k]{\frac{X_1^k + X_2^k + \ldots + X_N^k}{N}},$$

где k – показатель степени, определяющий вид степенной средней.

В случае простой средней все значения усредняемого признака X имеют одинаковую важность (вес). Если же значения X имеют неодинаковую важность (вес) при усреднении, то используется формула взвешенной степенной средней. В общем виде взвешенная степенная средняя имеет вид:

Слайд 5б. Взвешенная степенная средняя

$$\overline{X} = \left(\frac{\sum_{i=1}^{N} X_i^k \cdot f_i}{\sum_{i=1}^{N} f_i} \right)^{\frac{1}{k}},$$

где f_i – вес усреднения.

С изменением показателя степени k формула степенной средней меняется. При k=1 получаем арифметическую среднюю, при $k=2$ – квадратическую, при k=3 – кубическую, при $k=0$ – геометрическую, при k=-1 – гармоническую среднюю.

Если при замене индивидуальных величин признака на среднюю величину необходимо сохранить неизменной сумму квадратов исходных величин, то средняя будет являться *квадратической средней величиной* ($\widetilde{O}_{\hat{e}\hat{a}}$). Формула *средней квадратической простой*:

Слайд 6а. Средняя квадратическая простая и взвешенная

$$\overline{X}_{кв} = \sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N}}, \qquad \overline{X}_{кв} = \sqrt{\frac{\sum_{i=1}^{N} X_i^2 \cdot f_i}{\sum_{i=1}^{N} f_i}}.$$

Например, имеются три участка земельной площади со сторонами квадрата: x_1=100м; x_2=200м; x_3=300м. Очевидно, что необходимо исходить из того, что, заменяя разные значения длины сторон на среднюю, мы должны сохранять общую площадь всех участков. Арифметическая средняя величина (100+200+300):3=200м не удовлетворяет этому условию, так как общая площадь трех участков со стороной 200 м была бы равна: 3*(200м)^2=120 000м^2. В то же время площадь исходных трех участков равна: (100м)^2+(200м)^2+(300м)^2= 140.000м^2. Правильный ответ дает квадратическая средняя:

$$\overline{X}_{кв} = \sqrt{\frac{(100)^2 + (200)^2 + (300)^2}{3}} = 216\,м.$$

Если необходимо сохранить неизменной сумму кубов индивидуальных значений признака при их замене на среднюю величину, тогда используют *среднюю кубическую*:

Слайд 6б. Средняя кубическая

$$\overline{X}_{куб} = \sqrt[3]{\frac{\sum_{i=1}^{N} X_i^3}{N}}.$$

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

Слайд 7а. Средняя геометрическая простая и взвешенная

$$\overline{X}_{геом} = \sqrt[N]{X_1 \cdot X_2 \cdot \ldots \cdot X_N} = \sqrt[N]{\prod_{i=1}^{N} X_i}\,, \qquad \overline{X}_{геом} = \sqrt[\sum f_i]{X_1^{f_1} \cdot X_2^{f_2} \cdot \ldots} = \sqrt[\sum f_i]{\prod_{i=1}^{N} X_i^{f_i}}\,.$$

Основное применение геометрическая средняя находит при определении средних темпов роста. Геометрическая средняя дает наиболее правильный по содержанию результат осреднения, если задача состоит в нахождении такого значения признака, который качественно был бы равно удален как от максимального, так и от минимального значения признака. Например, если максимальный размер выигрыша в лотерее составляет миллион рублей, а минимальный – сто рублей, то какую величину выигрыша можно считать средней между миллионом и сотней? Арифметическая средняя явно непригодна, она составляет 500 050 руб., а это, как и миллион, крупный, никак не средний выигрыш; он качественно однороден с максимальным и резко отличен от минимального. Только геометрическая средняя дает верный с точки зрения экономики и логики ответ: $\sqrt{100 \cdot 1000000} = 10000\ \text{ðóá}$. Десять тысяч – не миллион, и не сотня! Это, действительно, нечто среднее между ними.

Если необходимо, чтобы неизменной оставалась при осреднении сумма величин, обратных индивидуальным значениям признака, то средняя величина является *гармонической средней*. Применяется при усреднении относительных величин (за исключением относительных показателей динамики). Формула расчета:

Слайд 7б. Средняя гармоническая простая и взвешенная

$$\bar{X}_{гарм} = \frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}}, \qquad \bar{X}_{гарм} = \frac{\sum_{i=1}^{N} f_i}{\sum_{i=1}^{N} \frac{f_i}{X_i}}.$$

Чем выше показатель степени *k,* тем больше значение средней величины (если индивидуальные значения признака варьируют). Если все исходные значения признака равны, то и все средние равны этой константе. Следующее соотношение называется *правилом мажорантности средних:*

Слайд 8а. Правило мажорантности

$$\bar{X}_{гарм} \le \bar{X}_{геом} \le \bar{X}_{арифм} \le \bar{X}_{квадр} \le \bar{X}_{куб}.$$

Данное соотношение справедливо для совокупности с положительными значениями признака *X*.

Пользуясь этим правилом, статистика может в зависимости от настроения и желания ее «знатока» либо «утопить», либо «выручить» студента, получившего на сессии оценки 2 и 5. Каков его средний балл? Средняя гармоническая 2,86 балла. Средняя арифметическая 3,5 балла. Средняя кубическая 4,05 балла.

4.

При изучении вариации применяются такие характеристики вариационного ряда, которые описывают количественно его структуру, строение. К таким показателям относятся *медиана* и *мода*.

Медиана – величина варьирующего признака, делящая совокупность на две равные части – со значениями признака меньше медианы и со значениями признака больше медианы.

Для ранжированного ряда с нечетным числом членов медианой является варианта, расположенная в центре ряда. Для ранжированного ряда с четным числом членов медианой будет средняя арифметическая из двух смежных центральных вариант.

В дискретном вариационном ряду медианой следует считать значение признака в той группе, в которой накопленная частота превышает половину численности совокупности.

В интервальном вариационном ряду порядок нахождения медианы следующий: индивидуальные значения признака располагают по ранжиру; определяют для данного ранжированного ряда накопленные частоты; по данным о накопленных частотах находят медианный интервал (медиана делит численность ряда пополам, следовательно, она там, где накопленная частота составляет половину или больше половины всей суммы частот), затем определяют по формуле *медиану*:

Слайд 8б. Медиана

$$Me = X_{Me} + i_{Me} \cdot \frac{\frac{\sum_{i=1}^{k} f_i}{2} - S_{Me-1}}{f_{Me}},$$

где Me – медиана;

X_{Me} – нижняя граница интервала, в котором находится медиана;

i_{Me} – величина медианного интервала;

S_{Me-1} – сумма накопленных частот, предшествующих медианному интервалу;

f_{Me} – частота медианного интервала;

k – число групп.

Аналогично медиане вычисляются значения признака, делящие совокупность на четыре равные по числу единиц части. Эти величины называются *квартилями* и обозначаются заглавной латинской буквой Q с подписным значком номера квартиля. Q_2 совпадает с Ме. Приведем формулы для первого и третьего квартилей:

Слайд 9а. Квартили

$$Q_1 = X_{Q_1} + i_{Q_1} \cdot \frac{\frac{\sum_{i=1}^{k} f_i}{4} - S_{Q_1-1}}{f_{Q_1}} \qquad Q_3 = X_{Q_3} + i_{Q_3} \cdot \frac{\frac{3 \cdot \sum_{i=1}^{k} f_i}{4} - S_{Q_3-1}}{f_{Q_3}}$$

Значения признака, делящие ряд на пять равных частей, называют *квинтилями,* на десять частей – *децилями,* на сто частей – *процентилями.*

Важное значение при характеристике структуры совокупности имеет такая величина признака, которая встречается в изучаемом ряду чаще всего. Такую величину называют *модой* (*Mo*).

В дискретном ряду мода определяется без вычисления как значение признака с наибольшей частотой.

Обычно встречаются ряды с одним модальным значением признака. Если два или несколько равных (и даже несколько различных, но больших, чем соседние) значений признака имеются в вариационном ряду, он считается соответственно бимодальным либо мультимодальным. Это говорит о неоднородности совокупности, возможно, представляющей собой агрегат нескольких совокупностей с разными модами.

В интервальном вариационном ряду, тем более при непрерывной вариации признака, когда каждое значение признака встречается только один раз. Модальным интервалом является интервал с наибольшей частотой. Внутри этого интервала находят условное значение признака, вблизи которого плотность распределения, т.е. число единиц совокупности, приходящееся на единицу измерения варьирующего признака, достигает максимума. Это условное значение и считается *точечной модой*, определяемой по формуле:

Слайд 10а. Мода

$$Mo = X_{Mo} + i_{Mo} \cdot \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

где Mo – мода;

X_{Mo} – нижняя граница модального интервала;

i_{Mo} – величина модального интервала;

f_{Mo} – частота в модальном интервале;

f_{Mo-1} – частота в предшествующем интервале;

f_{Mo+1} – частота в следующем интервале за модальным.

Соотношение между средней величиной, медианой и модой: если распределение по форме близко к нормальному закону, то медиана находится между модой и средней величиной, причем ближе к средней, чем к моде.

Слайд 10б. Соотношение между средней, медианой и модой

при правосторонней асимметрии: $\bar{X} > Me > Mo$;

при левосторонней асимметрии: $\bar{X} < Me < Mo$.

для умеренно асимметричных распределений справедливо равенство:

$$|Mo - \bar{X}| = 3 \cdot |Me - \bar{X}|.$$