СТАТИСТИКА.

Описательная статистика.

Лекция 1. Абсолютные, относительные и средние величины. Мода и медиана.

Автор: Равичев Л.В.
РХТУ им. Д.И.Менделеева
Кафедра управления технологическими инновациями
Москва - 2013

Абсолютные величины

Абсолютные величины характеризуют численность совокупности и объём изучаемого явления в определенных границах времени и места.

Абсолютная величина

Объём явления на определённую дату

Объём явления за определённый период времени

Относительные величины

Относительная величина представляет собой результат сопоставления двух статистических показателей и даёт цифровую меру их соотношения.

$$\textit{Относительная величина} = \frac{\textit{Сравниваемый показатель}}{\textit{База сравнения}}$$

Относительная величина

Результат соотношения одноимённых статистических показателей

Результат соотношения разноимённых статистических показателей

Относительные величины одноимённых статистических показателей в экономике

1. Относительные величины динамики характеризует изменение явления во времени. Они показывают во сколько раз изменится объём явления за определённый период времени, т.е. ***темпы роста.***

Темпы роста с переменной базой (цепные темпы роста):

$$\boxed{T_{p_1} = \frac{y_2}{y_1}\times 100; \quad T_{p_2} = \frac{y_3}{y_2}\times 100; \; \ldots \; T_{p_{n-1}} = \frac{y_n}{y_{n-1}}\times 100}$$

Темпы роста с постоянной базой (базисные темпы роста):

$$\boxed{T'_{p_1} = \frac{y_1}{y_k}\times 100; \quad T'_{p_2} = \frac{y_2}{y_k}\times 100; \; \ldots \; T'_{p_n} = \frac{y_n}{y_k}\times 100}$$

Относительные величины одноимённых статистических показателей в экономике

Пример. Имеются следующие данные о стоимости основного капитала по фирме:

№ предприятия входящего в фирму	Стоимость основного капитала, тыс. руб.		
	на 1 января 1999 г.	на 1 января 2000 г.	на 1 января 2001 г.
1	22 150	24 855	26 970
2	7 380	9 100	12 550
3	13 970	16 700	20 800

Определить показатели динамики стоимости основного капитала фирмы.

Решение:

на 1 января 1999 г. – y_1 = 22 150 + 7 380 + 13 970 = 43 500

на 1 января 2000 г. – y_2 = 24 855 + 9 100 + 16 700 = 50 655

на 1 января 2001 г. – y_3 = 26 970 + 12 550 + 20 800 = 60 320

1) Темпы роста с переменной базой:

$$T_{p1} = \frac{y_2}{y_1} \times 100 = \frac{50655}{43500} \times 100 = 116,4\%; \quad T_{p2} = \frac{y_3}{y_2} \times 100 = \frac{60320}{50655} \times 100 = 119,1\%$$

2) Темпы роста с постоянной базой (за постоянную базу принимаем данные на 01.01.99г.) :

$$T_{p1} = \frac{y_2}{y_1} \times 100 = \frac{50655}{43500} \times 100 = 116,4\%; \quad T_{p2} = \frac{y_3}{y_1} \times 100 = \frac{60320}{43500} \times 100 = 138,7\%$$

Относительные величины одноимённых статистических показателей в экономике

2. Относительная величина структуры (удельный вес):

$$K_{c_i} = \frac{y_i}{\sum_{i=1}^{n} y_i}; \qquad \sum_{i=1}^{n} K_{c_i} = 1$$

3. Относительная величина координации:

$$K_{k_m} = \frac{y_i}{y_j}; \quad i = 1...n; \quad j = 1...n; \quad i \neq j$$

Относительные величины одноимённых статистических показателей в экономике

4. Относительная величина наглядности (сравнения):

$$K_{н_m} = \frac{y_i'}{y_j}; \quad i = 1 \ldots n_1; \quad j = 1 \ldots n_2$$

Относительные величины разноимённых статистических показателей в экономике

Эта группа статистических показателей носит название ***отно-сительных величин интенсивности.***

Относительная величина интенсивности показывает сте-пень распространённости данного явления в изучаемой среде и образуется в результате сравнения разноименных, но опреде-лённым образом связанных между собой абсолютных величин.

$$K_{u_m} = \frac{y_i}{z_j}; \quad i = 1 \ldots n_1; \quad j = 1 \ldots n_2$$

Категории средних

- Степенные средние
  - Средняя арифметическая
  - Средняя геометрическая
  - Средняя гармоническая
  - Средняя квадратическая
- Средняя хронологическая
- Структурные средние
  - Мода
  - Медиана

Степенная средняя случайной величины

Для степенной средней определяющей функцией является уравнение:

$$\sum_{i=1}^{n} x_i^k m_i = \sum_{i=1}^{n} \bar{x}_i^k m_i$$

откуда

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^{n} x_i^k m_i}{\sum_{i=1}^{n} m_i}}$$

где k может принимать значения -1; 1; 2.

Среднее арифметическое значение случайной величины (k=1)

Средним арифметичским значением дискретной случайной величины называют сумму произведений всех ее возможных значений на их вероятности. Если x имеет конечное число значений x_i, которые встречаются f_i раз то среднее значение x вычисляют по формуле:

$$\boxed{\bar{x}_{ap} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n}{f_1 + f_2 + \ldots + f_n}}$$

В самом простом случае, когда значения x_i встречаются только по одному разу, формула упрощается и принимает вид:

$$\boxed{\bar{x}_{ap} = \frac{x_1 + x_2 + \ldots + x_n}{n}}$$

Среднее арифметическое значение случайной величины (k=1)

Среднее средних значений

Если большое количество данных разбито на k групп, для которых подсчитаны групповые средние значения, то чтобы подсчитать общее среднее X нужно умножить групповые средние $\bar{x}_{ap1}, \bar{x}_{ap2}, \ldots, \bar{x}_{apk}$ на соответствующее количество данных в группах $n_1, n_2, \ldots n_k$ и сложить эти произведения, а затем разделить сумму на общее количество данных.

$$\boxed{\bar{x}_{ap.общ.} = \frac{\bar{x}_{ap1} \cdot n_1 + \bar{x}_{ap2} \cdot n_2 + \ldots + \bar{x}_{apk} \cdot n_k}{n_1 + n_2 + \ldots + n_k}}$$

Среднее значение суммы случайных величин

Среднее значение суммы случайных величин равно сумме средних значений случайных величин. Так, для двух наборов случайных величин $X_1, X_2, \ldots, X_k$ и $Y_1, Y_2, \ldots, Y_n$, с соответствующими вероятностями появления $p_1, p_2, \ldots, p_k$ и $q_1, q_2, \ldots, q_n$, расчетная формула имеет вид:

$$\boxed{\overline{X+Y} = \sum_{i=1}^{k} X_i p_i + \sum_{j=1}^{n} Y_j q_j = \overline{X} + \overline{Y}}$$

В случае большего количества наборов случайных величин формула имеет аналогичный вид:

$$\boxed{\overline{X+Y+Z} = \sum_{i=1}^{k} X_i p_i + \sum_{j=1}^{n} Y_j q_j + \sum_{j=1}^{m} Z_j r_j = \overline{X} + \overline{Y} + \overline{Z}}$$

Среднее значение произведения случайных величин

Среднее значение произведения взаимно независимых случайных величин равно произведению средних значений случайных величин. Так, для двух наборов независимых случайных величин $X_1, X_2,\ldots, X_k$ и $Y_1, Y_2,\ldots, Y_n$, с соответствующими вероятностями появления $p_1, p_2,\ldots, p_k$ и $q_1, q_2,\ldots, q_n$, расчетная формула имеет вид:

$$\boxed{\overline{X\times Y}=\sum_{i=1}^{k}X_i p_i\times\sum_{j=1}^{n}Y_j q_j=\overline{X}\times\overline{Y}}$$

Среднее гармоническое значение случайных величин (k= -1)

Если случайная величина ***x*** имеет конечное число значений x_i, которые встречаются f_i раз, то среднее гармоническое:

$$\bar{x}_{гарм} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

В самом простом случае, когда все f_i одинаковые.

$$\bar{x}_{гарм} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Среднее квадратическое значение случайных величин (k=2)

Если случайная величина ***x*** имеет конечное число значений x_i, которые встречаются f_i раз, то среднее квадратическое:

$$\overline{x}_{кв} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i}}$$

В самом простом случае, когда $f_i = 1$:

$$\overline{x}_{кв} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Среднее геометрическое значение случайных величин

Если случайная величина ***x*** имеет конечное число значений $\boldsymbol{x}_i$, которые встречаются $\boldsymbol{f}_i$ раз, то среднее геометрическое значение ***x*** вычисляют по формуле:

$$\boxed{\bar{x}_{геом} = \sqrt[n]{\prod_{i=1}^{n} x_i^{f_i}}}$$

В самом простом случае, когда значения $\boldsymbol{x}_i$ встречаются только по одному разу, формула упрощается и принимает вид:

$$\boxed{\bar{x}_{геом} = \sqrt[n]{\prod_{i=1}^{n} x_i}}$$

Среднее геометрическое значение случайных величин

Пример. Перевозка грузов по автотранспортному предприятию такова:

	Январь	Февраль	Март
Перевезено грузов, тыс. т	37,0	40,5	42,0

Определить среднемесячный темп роста объёма грузовых перевозок.

Решение: Коэффициенты роста объёма грузовых перевозок:

$$\boxed{K_1 = \frac{40{,}5}{37{,}0} = 1{,}095} \qquad \boxed{K_2 = \frac{42{,}0}{40{,}5} = 1{,}037}$$

Среднемесячный коэффициент роста определяется по формуле средней геометрической:

$$\boxed{\overline{K} = \sqrt[2]{K_1 \times K_2} = \sqrt[2]{1{,}095 \times 1{,}037} = 1{,}066}$$

или 106,6% (средний темп роста).

Средняя хронологическая случайных величин

Если случайные величины $y_1, y_2, \ldots, y_n$ представляют собой моментальный динамический ряд, то средний уровень такого ряда оценивается по формуле ***средней хронологической взвешенной:***

$$\boxed{\bar{y}_{хрон.вз.} = \frac{\sum_{i=1}^{n} y_i \Delta t_i}{\sum_{i=1}^{n} \Delta t_i}}$$

Где $\bar{y}_{хрон.вз.}$ - средний уровень ряда; y_i – уровни динамического ряда; Δt_i - время, в течение которого данный уровень ряда оставался неизменным.

Средняя хронологическая случайных величин

Пример №1. На 1 января 2001 года число сотрудников компании «Бест» состав--ляло 551 человек, 2 января уволился 1 сотрудник, 6 января было принято на работу 24 человека, 16 января было принято 6 человек, 25 января уволилось 10 сотрудников. Найти среднее значение числа сотрудников компании в январе 2001 года.

Численность сотрудников компании «Бест», чел. (y)	Число календарных дней, в течение которых данная численность сотрудников оставалась без изменения (Δt)	yΔt
551	2	1102
550	4	2200
574	10	5740
580	9	5220
570	6	3420
ИТОГО	31	17682

$$\bar{y}_{хрон.вз.} = \frac{17682}{31} = 570 \ чел.$$

Средняя хронологическая случайных величин

Пример №2. Определить на сколько рублей и на сколько процентов различаются средние остатки по вкладам за первый квартал, если на 1 января 2002 года остаток по первому вкладу составлял 500 руб., по второму вкладу – 700 руб. В течение первого квартала имели место следующие изменения величины остатков вкладов (руб.):

Вклады	Дата изменения размера вклада						
	05.01	17.01	02.02	21.02	13.03	20.03	28.03
1	+150	-200	-	+500	-	-	+100
2	-	-	+300	+150	-550	-200	+400

Средняя хронологическая случайных величин

Вклад №1

Периоды	Число дней в периоде (Δt)	Размер вклада (руб.)	yΔt
01.01-05.01	4	500	2 000
05.01-17.01	12	650	7 800
17.01-21.02	35	450	15 750
21.02-28.03	35	950	33 250
28.03-1.04	4	1 050	4 200
Итого	90	-	63 000

$$\bar{y}_1 = \frac{63000}{90} = 700 \;\; руб.$$

Вклад №2

Периоды	Число дней в периоде (Δt)	Размер вклада (руб.)	yΔt
01.01-02.02	32	700	22 400
02.02-21.02	19	1 000	19 000
21.02-13.03	20	1 150	23 000
13.03-20.03	7	600	4 200
20.03-28.03	8	400	3 200
28.03-01.04	4	800	3 200
Итого	90	-	75 000

$$\bar{y}_2 = \frac{75000}{90} = 833 \;\; руб. \; 33 \;\; коп.$$

Средняя хронологическая случайных величин

В случае, если характер изменения уровней ряда в рассматриваемые периоды неизвестен, и уровни ряда отстоят друг от друга на неравные промежутки времени, то средняя хронологическая взвешенная вычисляется по формуле:

$$\bar{y}_{хрон.вз.} = \frac{\frac{y_1 + y_2}{2}\Delta t_1 + \frac{y_2 + y_3}{2}\Delta t_2 ... + \frac{y_{n-1} + y_n}{2}\Delta t_{n-1}}{\sum_{i=1}^{n-1}\Delta t_i}$$

Средняя хронологическая случайных величин

Пример. Средняя численность работников предприятий розничной торговли Российской Федерации характеризуется следующими данными:

Годы	Средняя численность работающих в розничной торговле, тыс. чел.
1970	2203
1980	2802
1990	2768
1995	3136
2000	3109

$$\bar{y}_{хрон.вз.} = \frac{\frac{2203+2802}{2}10 + \frac{2802+2768}{2}10 + \frac{2768+3136}{2}5 + \frac{3136+3109}{2}5}{30} = 2603 \ тыс.чел.$$

Средняя хронологическая случайных величин

В случае, если промежутки времени между датами, на которые имеются данные одинаковы, и при равномерном изменении размера показателя между датами средняя хронологическая ряда вычисляется по формуле:

$$\bar{y}_{хрон.вз.} = \frac{\frac{y_1}{2} + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

где y_1 и y_n – начальный и конечный уровни ряда, n – число дат.

Средняя хронологическая случайных величин

Пример №1. Товарные запасы ОАО «Золотой век» на конец года представлены в следующей таблице:

	Годы				
	2007	2008	2009	2010	2011
Товарные запасы, тыс. руб.	26528	27567	29073	31561	35253

Среднегодовой запас товаров ОАО «Золотой век» за пятилетний период составил:

$$\bar{y}_{хрон.вз.} = \frac{\frac{26528}{2} + 27567 + 29073 + 31561 + \frac{35253}{2}}{4} = 29772{,}9 \ \text{тыс. руб.}$$

Средняя хронологическая случайных величин

<u>Пример №2.</u> Имеются следующие данные о стоимости имущества предприятия (млн. руб.):

Год	Отчётные данные			
	1.01	1.04	1.07	1.10
1999	62	65	70	68
2010	68	70	75	78
2001	80	84	88	90
2002	95	-	-	-

Определить абсолютное и относительное изменение среднегодовой стоимости имущества предприятия в 2001 г. по сравнению с 1999 и 2000 гг.

$$\bar{y}_{1999} = \frac{\frac{62}{2} + 65 + 70 + 68 + \frac{68}{2}}{4} = 67\,млн.руб.$$

$$\bar{y}_{2000} = \frac{\frac{68}{2} + 70 + 75 + 78 + \frac{80}{2}}{4} = 74{,}25\,млн.руб.$$

$$\bar{y}_{2001} = \frac{\frac{80}{2} + 84 + 88 + 90 + \frac{95}{2}}{4} = 87{,}375\,\text{ìëí .ðóá.}$$

$\Delta_{2001\text{-}1999}$ = 87,375-67=20,375

$К_{2001/1999}$ = 87,375/67=1,304

(или увеличилась на 30,4%)

$\Delta_{2001\text{-}2000}$ = 87,375-74,25=13,125

$К_{2001/2000}$ = 87,375/74,25=1,177

(или увеличилась на 17,7%)

Мода

Модой называется значение признака, которое наиболее часто встречается в совокупности (в статистическом ряду).

1. Нахождение модальной величины в ***дискретном ряду***.

<u>Пример №1.</u> Обувной фабрикой проведено выборочное исследование потребляемой женщинами обуви, результаты которого приведены в таблице:

Размер обуви	32	33	34	35	36	37	38	39	40	41	42
Количество женщин	6	33	247	910	2093	2696	1923	1196	283	51	55

Мода этого ряда

Мода

<u>Пример №2.</u> Проведена малая выборка из партии электрических лампочек для определения продолжительности их службы. Результаты выборки приведены в таблице:

№ лампочки	1	2	3	4	5	6	7	8	9
Срок горения, час.	1450	1400	1370	1430	1400	1380	1270	1420	1400

Ранжированный ряд:

1270	1370	1380	1400	1400	1400	1420	1430	1450

Мода этого ряда

Мода

2. Нахождение модальной величины в ***интервальном вариационном ряду***.

$$M_o = x_{mo} + i\frac{f_2 - f_1}{(f_2 - f_1) + (f_2 - f_3)}$$

где: x_{mo}- нижняя граница модального интервала; i – разность между верхней и нижней границей модального интервала; f_1 – частота интервала, предшествующая модальному; f_2 – частота модального интервала; f_3 – частота интервала, следующего за модальным.

Мода

<u>Пример.</u> В таблице приведены данные о торговой площади магазинов:

Торговая площадь магазинов, м^2	Число магазинов
До 100	3
От 100 до 120	13
От 120 до 140	15
От 140 до 160	20
От 160 до 180	8
Свыше 180	1
ИТОГО	60

Необходимо рассчитать моду из интервального ряда.

$$M_o = 140 + 20\frac{20-15}{(20-15)+(20-8)} = 145{,}88\ \text{м}^2$$

Медиана

Медианой называется серединная варианта упорядоченного вариационного ряда, расположенного в возрастающем или убывающем порядке (ранжированный вариационный ряд).

1. Нахождение медианы в ***дискретном ранжированном вариа ционном ряду.***

<u>Пример.</u>

а) дан нечетный ранжированный вариационный ряд роста студенток:

156	158	160	161	166	168	172

M_e=161; место медианы N_{me}=(n+1)/2=4.

б) дан четный ранжированный вариационный ряд роста студенток:

155	156	158	160	161	166	168	172

$$\boxed{M_e = \frac{160+161}{2} = 160{,}5}$$

Медиана

2. Нахождение медианы ***интервального ряда.***

$$M_e = x_o + i\frac{\frac{\sum_{i=1}^{n} f_i}{2} - S_{m-1}}{f_m}$$

где: x_o – нижняя граница медианного интервала; i – величина медианного интервала; f_i – частоты интервального ряда; S_{m-1} – сумма накопленных частот в интервалах предшествующих медианному; f_m – частота медианного интервала.

Медиана

<u>Пример.</u> В таблице даны группы семей по среднемесячному доходу на 1 человека. Требуется для приведенного интервального ряда определить серединное значение, т.е. медиану.

Группы семей по среднемесячному доходу на 1 человека, руб.	Число семей
До 900	10
От 900 до 1200	20
От 1200 до 1500	40
От 1500 до 1800	10
Свыше 1800	20
ИТОГО	100

$$M_e = 1200 + 300 \frac{\frac{100}{2} - 30}{40} = 1350 \;\; руб.$$

Следовательно, 50% семей имеют доход на одного человека <1350 руб.

Медиана

Свойство медианы:*сумма абсолютных величин линейных отклонений от M_e минимальна.*

<u>Пример.</u> Филиалы торговой фирмы «Элегант» расположены на расстоянии 10, 30,70, 90, 100 км от неё. Где построить склад фирмы для оптимального снабжения филиалов (минимум пробега автомобильного транспорта):

Расстояние, км	$x - M_e$	$x - \bar{x}$
10	-60	-40
30	-40	-20
70	0	+20
90	+20	+40
100	+30	+50
ИТОГО	$\sum\lvert x - M_e\rvert=150$	$\sum\lvert x - \bar{x}\rvert=170$

Квартили

Более общая постановка вариант, занимающих определённое место в ранжированном ряду, называется ***порядковой статистикой.***

Квартиль – значения признака, которые делят ранжированный ряд на четыре равные по численности части. Таких величин будет три: первая квартиль (Q_1), вторая квартиль (Q_2), третья квартиль (Q_3). Вторая квартиль является медианой.

Место квартили:

$$N_{Q_1} = \frac{\sum_{i=1}^{n} f_i}{4}$$

$$N_{Q_2} = \frac{2\sum_{i=1}^{n} f_i}{4}$$

$$N_{Q_3} = \frac{3\sum_{i=1}^{n} f_i}{4}$$

Квартили

Нижний квартиль:

$$Q_1 = x_o + i\frac{\frac{\sum_{i=1}^{n} f_i}{4} - S_{Q_1}}{f_{Q_1}}$$

Верхний квартиль:

$$Q_3 = x_o + i\frac{\frac{3\sum_{i=1}^{n} f_i}{4} - S_{Q_3}}{f_{Q_3}}$$

где: x_o – нижняя граница квартильных интервалов; i – величина интервала; f_i – частоты интервального ряда; S_{Q1} – сумма накопленных частот в интервалах предшествующих нижнему квартилю; S_{Q3} – сумма накопленных частот в интервалах предшествующих верхнему квартилю; f_{Q1}, f_{Q3} – частота квартильного интервала.

Квартили

Пример. Дан интервальный ряд распределения 50 учащихся по росту:

Рост, см	Число учащихся	Накопленные частоты
160-165	3	3
165-170	7	10
170-175	16	26
175-180	10	36
180-185	9	45
185-190	3	48
190-195	2	50
Всего	50	-

Определить нижний и верхний квартиль.

Квартили

Место нижнего квартиля:

$$N_{Q_1} = \frac{50}{4} = 12{,}5$$

Место медианы ранжированного интервального ряда:

$$N_{Q_2} = \frac{50}{2} = 25$$

Место верхнего квартиля:

$$N_{Q_3} = \frac{50}{4}3 = 37{,}5$$

Квартили

$$N_{Q_1} = 12{,}5$$

$$N_{Q_3} = 37{,}5$$

Нижний квартильный интервал

Верхний квартильный интервал

Рост, см	Число учащихся	Накопленные частоты
160-165	3	3
165-170	7	10
170-175	16	26
175-180	10	36
180-185	9	45
185-190	3	48
190-195	2	50
Всего	50	-

Квартили

Нижний квартиль:

$$Q_1 = 170 + 5\frac{\frac{50}{4} - 10}{16} = 170{,}8$$

Верхний квартиль:

$$Q_3 = 180 + 5\frac{\frac{150}{4} - 36}{9} = 180{,}8$$