

Показатели , характеризующие
центральную тенденцию ряда

Показатели , характеризующие
вариацию вокруг центральной
тенденции

Характеристики
рассеивания

Выход

Показатели , характеризующие центральную тенденцию ряда

Математическим ожиданием выборки называется сумма произведений всех ее возможных значений на соответствующие относительные частоты:

$$M(x) = x_1p_1 + x_2p_2 + \dots + x_n p_k$$

Т.е. математическое ожидание - это «среднее взвешенное» возможных значений.

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Пример 1. Найти математическое ожидание для следующих данных:

Варианта	2	6	10	12	14
Частота	1	5	7	3	4
Относительная частота	1/20	5/20	7/20	3/20	4/20

$$M(x) = x_1 p_1 + x_2 p_2 + \dots + x_n p_k$$

В этом случае: $M(X) = 2 * 1/20 + 6 * 5/20 + 10 * 7/20 + 12 * 3/20 + 14 * 4/20 = 9,7$.

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Пример 2. Студенты какой группы справились с контрольной лучше?

Оценка	2	3	4	5
1 группа	2	7	10	3
2 группа	1	9	10	2

$$\bar{x} = \frac{1}{n} * \sum_{i=1}^k x_i * n_i$$

$$\bar{x} = \frac{1}{n} (x_1 n_1 + x_2 n_2 + \dots + x_k n_k)$$

$$\bar{x}_1 = \frac{1}{15} (2 * 2 + 3 * 0 + 4 * 10 + 5 * 3) = \frac{4 + 40 + 15}{15} = \frac{59}{15} = 3.9$$

$$\bar{x}_2 = \frac{1}{20} (2 * 1 + 3 * 9 + 4 * 10 + 5 * 0) = \frac{2 + 27 + 40}{20} = \frac{69}{20} = 3.45$$

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Среднее арифметическое 2х групп

1 способ

$$\bar{x} = \frac{3.9 \cdot 15 + 3.45 \cdot 20}{35} = 3.6$$

2 способ

Оценка	2	3	4	5
1 группа	2		10	3
2 группа	1	9	10	

$$\bar{x} = \frac{1}{35} (2 \cdot 3 + 3 \cdot 9 + 4 \cdot 20 + 5 \cdot 3) = 3.6$$

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Мода - это наиболее часто встречающееся значение признака.

Рассмотрим случай точечного распределения.

В совокупности оценок успеваемости **2, 3, 4, 4, 4, 5, 5** модой является оценка **4**, потому, что эта оценка встречается чаще других.

Принято считать, что в случае, когда все значения оценок встречаются одинаково часто, совокупность данных моды не имеет.

Например, в совокупности **3, 3, 3, 4, 4, 4, 5, 5, 5** моды нет.

В примере совокупности **2, 3, 3, 4, 5, 5** модами являются оценки **3** и **5**. В этом случае говорят, что совокупность оценок является **бимодальной**.

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

При интервальном распределении моды подсчитываем следующим образом:

Количес твенное значение признака	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Число случаев	1	6	19	58	53	24	16	3

1. Определим модальный интервал.

$$[180;200]_{n_1} = 58$$

$$2. MoX = x_s + h \cdot (n_s - n_{s-1}) / ((n_s - n_{s-1}) + (n_s - n_{s+1}))$$

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

При интервальном распределении моды подсчитываем следующим образом:

Количес- твенное значение признака	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Число случаев	1	6	19	58	53	24	16	3

x_s — начало модального интервала

h -длина модального интервала

n_s — частота в модальном интервале

n_{s+1} — частота в последующем интервале

n_{s-1} — частота в предыдущем интервале

$$Mo = 180 + 20 \frac{(58 - 19)}{(58 - 19) + (58 - 53)} = 197.73$$

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Пример 3. Данные статистического исследования представлены в таблице:

Количес- твенное значение признака	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Число случаев	1	6	19	58	53	24	16	3

Найти $M_o X$.

Т.к. максимальная частота ($n_j = 58$) соответствует интервалу 180-200, то $X'_s = 180$, $n_{s-i} = 19$, $n_{s+i} = 53$. Значит,
 $M_o X = 180 + 20 - (58 - 19) / (39 + 5) = 197,73$.

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Медианой MeX называется значение признака, относительно которого генеральная совокупность делится на две равные по объему части, причем в одной из них содержатся члены, у которых значение признака не превосходит MeX , а в другой - не меньше MeX .

Математическое
ожидание

Мода

Медиана



Выход

Показатели , характеризующие центральную тенденцию ряда

Пример 4.

Варианты	2	3	4	5
Количество учеников	2	7	10	3
Отн. частоты	2/22	7/22	10/22	3/22

Всего-22 ученика

$$\sum_{i=2}^3 \frac{2}{22} + \frac{7}{22} = \frac{9}{22} < \frac{11}{22}$$

$$\sum_{i=2}^4 \frac{2}{22} + \frac{7}{22} + \frac{10}{22} = \frac{19}{22} > \frac{11}{22}$$

$$\sum_{i=4}^5 \frac{10}{22} + \frac{3}{22} = \frac{13}{22} > \frac{11}{22}$$

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Для определения медианы по интервальному признаку используется следующая формула:

$$MeX = X_p + h * (n/2 - w(X_p)) / n_p$$

где h - ширина интервала,
 n - объем генеральной совокупности,
 $w(X_p)$ - накопленная частота до p -то интервала,
 n_p - частота интервала,
 p - номер медианного интервала.

Математическое
ожидание

Мода

Медиана

Выход



Показатели , характеризующие центральную тенденцию ряда

Пример 5. Вычислим для данного вариационного ряда медиану. Для ее нахождения строим кумулятивный ряд:

Кол-во значения признака	120-14 0	140-16 0	160-18 0	180-200	200-220	220-24 0	240-26 0	260-280
Число случаев	1	6	19	58	53	24	16	3
Накопленная частота	1	7	26	84	137	16	177	180

$$MeX = X_p + h * (n/2 - w(X_p)) / n_p$$

Сначала определим медианный интервал.

$$w(X_p) < n/2 < w(X_p)$$

$$n=180$$

$$90 < 90 < 90$$

$$84 < 90 < 137$$

$$h=20 \quad X_p=200$$

$$n_p=53 \quad w(X_p)=84$$

$$MeX = 200 + 20 * (90 - 84) / 53 = 202,26.$$



Математическое
ожидание

Мода

Медиана

Выход

Показатели , характеризующие центральную тенденцию ряда

Найдем для каждой группы точное значение медианы.

Для контрольной: $X_p = 3, h=1, n=33, w(X_p)=5, n_p=13.$

Значит,

$$MeX = 3 + 1 * (33/2 - 5) / 13 = -3.9.$$

Для экспериментальной группы: $X_p = 4, h=1, n=33, w(X_p)=11,$

$n_p=15.$

Значит,

$$MeX = 4 + 1 * (33/2 - 11) / 15 = -4.37.$$

Таким образом, мы можем сказать, что среднее число посещений музеев в контрольной группе - **3,9**, а в экспериментальной группе - **4,37**.

Математическое
ожидание

Мода

Медиана



Выход

Показатели , характеризующие вариацию вокруг центральной тенденции

Дисперсия выборки («рассеивание») - это величина, характеризующая разброс ее значений вокруг среднего. Обозначается $D(X)$.

Рассмотрим, как вычисляется дисперсия.

Для точечного распределения имеем

$$D(X) = (x_1 - M(X))^2 * n_1 + (x_2 - M(X))^2 * n_2 + \dots + (x_n - M(X))^2 * n_n,$$

где x_i - значения вариантов,

p_i - значения соответствующих относительных частот.

Для примера 1 вычислим дисперсию.

Напомним, что $M(X) = 9,7$.

По формуле:

$$D(X) = (2 - 9,7)^2 * 1/20 + (6 - 9,7)^2 * 5/20 + (10 - 9,7)^2 * 7/20 + (12 - 9,7)^2 * 3/20 + (14 - 9,7)^2 * 3/20 = 10,91.$$

Дисперсия

Коэффициент
вариации

Корреляция

Выход



Показатели , характеризующие вариацию вокруг центральной тенденции

Вычислим дисперсию в случае интервального распределения изучаемого признака.

Каждый интервал мы заменяем его средним значением, а далее пользуемся формулой, которая использовалась для точечного распределения:

$$D(X) = 1/n \sum (z_k - M(X))^2$$
$$O_{nk} = 1/n \sum (z_0 + kh - z_0 - kh)^2$$
$$O_{ik} = h^2/n \sum (k - k)^2$$
$$O_{ik} = h^2/n \sum k^2 - kQ$$

где $k=1/n \sum knk$ и суммирование по k .

Дисперсия

Коэффициент
вариации

Корреляция



Выход

Показатели , характеризующие вариацию вокруг центральной тенденции

Интервалы (классы)	20-25	25-30	30-35	35-40	40-45	45-50
n_i	2	3	6	10	17	2
k_i	-4	-3	-2	-1	0	1
n_{II}	-8	-9	-12	-10	0	2
$n_i MU$	32	27	24	10	0	2

$$E=40$$

$$S=-37 \quad S=95$$

Для данного примера

$$D(X) = 50(1/40 - 95 - (37/40)^{-37,98}; \sigma - 6,16.$$

Дисперсия

Коэффициент
вариации

Корреляция

Выход



Показатели , характеризующие вариацию вокруг центральной тенденции

Коэффициент вариации - это числовая характеристика выборки, которая показывает соотношение между математическим ожиданием выборки и ее дисперсией:

$$V(X) = M(X) / D(X) * 100\%$$

Дисперсия

Коэффициент
вариации

Корреляция

Выход



Показатели , характеризующие вариацию вокруг центральной тенденции

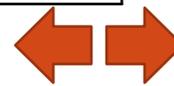
Связи (зависимости) между двумя и более переменными в статистике называются **корреляцией**.

Пример 6. Рассмотрим два ряда данных: X - семейное положение, Y - исключение из университета.

Предположим, что измеряются они по шкале наименований (0-нет, 1-да для каждой из переменных).

В силу того, что данные получены в результате использования такой шкалы наименований, пары (x, y) могут быть только вида $(0,0)$; $(0,1)$, $(1,0)$, $(1,1)$. Составим таблицу:

Признак Y	Признак X		
	$x_t = 0$	$x_t = 1$	
$Y=0$	a	b	$a+b$
$Y=1$	c	d	$c+d$
Итого	$a+c$	$b+d$	N



Дисперсия

Коэффициент
вариации

Корреляция

Выход

Показатели , характеризующие вариацию вокруг центральной тенденции

В общем виде формула корреляции Пирсона для такого вида данных имеет вид:

$$\underline{(be-ad)}$$

Вернемся к нашему примеру. Получены данные по шкале наименований:

<i>№ испытуемого</i>	<i>Переменная X</i>	<i>Переменная Y</i>
1	0	0
2	1	1
3	0	1
4	0	0
5	1	1
6	1	0
7	0	0
8	1	1
9	0	0
10	0	1

Дисперсия

Коэффициент
вариации

Корреляция

Выход



Показатели , характеризующие вариацию вокруг центральной тенденции

Составим таблицу, удобную для вычисления коэффициента корреляции:

Признак У	Признак Х			
		$X_i = 0$		
U_{10}	2	3	5	
U_2^{-1}	4	1	5	
Итого	6	4	10	

Подставим в формулу данные из этой таблицы:
(12-2)
, = 0,32.

Таким образом, коэффициент корреляции Пирсона для выбранного примера равен 0,32, т.е. зависимость между семейным положением студентов и фактами исключения из университета незначительная.

Дисперсия

Коэффициент
вариации

Корреляция

Выход

