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:

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

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

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

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

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

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

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

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

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

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 моду подсчитываем следующим образом:

Количественное 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 моду подсчитываем следующим образом:

Количественное 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:

Количественное 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 *Мо X.*

Т.к. максимальная частота (я *=58)* соответствует интервалу 180-200, то *X's=180, ns-i =19, ns+i =53*. Значит,

Мо 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*.

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

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

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

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

Математическое ожидание

Мода

Медиана

Выход

Показатели , характеризующие центральную тенденцию ряда

Для определения медианы по интервальному признаку используется следующая формула:

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

где h - ширина интервала,
 n - объем генеральной совокупности,
$w\ (Xp)$ - накопленная частота до p-*то* интервала,
np – частота интервала,
/? - 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	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Число случаев	1	6	19	58	53	24	16	3
Накопленная частота	*1*	*7*	*26*	*84*	*137*	*16*	*177*	*180*

$$MeX = Xp + h^* (n/2 - w(X_P))/n_P$$

Сначала определим медианный интервал.

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

$$n=180$$

$$?<90<?$$

$$84<90<137$$

$$h=20 \qquad Xp=200$$

$$np=53 \quad w(Xp)=84$$

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

Математическое ожидание

Мода

Медиана

Выход

Показатели , характеризующие центральную тенденцию ряда

Найдем для каждой группы точное значение медианы.

Для контрольной: $Xp = 3, h=1, n=33, w(X_p)=5, np=13.$
Значит,
$MeX = 3+1*(33/2-5)/13 = -3.9.$

Для экспериментальной группы: $Xp = 4, h=1, n=33, w(Xp)=11, n_p=15.$
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 *Д(Х)*.

Рассмотрим, как вычисляется дисперсия.

Для точечного распределения имеем
$Д(X)= (x1 - M(X))^2*n_1+ (за - M(X))^2*n_2 +\ldots +(xn -M(X)$
$^2*n_n$,

где *xi* - значения вариант,
pi - значения соответствующих относительных частот.

Для примера 1 вычислим дисперсию.
Напомним, что *М(Х)=9,7.*
По формуле:
Д(Х)= (2-9,7)0-1/20+ (6-9,7)0-5/20+ (10-9,7)0-7/20+ (12-9,7)0 -3/20+(14-9,7) Ш/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:

Д(Х) =1/n Z (zк -М(Х))

Опк =1/n Z (zo +kh-zo -kh)

Oik =hO/n Z (k-k)

Oik = h ЦЯ/n Z

kOi -кQ,

где *k=l/n E knk* и суммирование по *к*.

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*
ni	*2*	*3*	*6*	*10*	*17*	*2*
ki	*-4*	*-3*	*-2*	*-1*	*0*	*1*
nШ	*-8*	*-9*	*-12*	*-10*	*0*	*2*
niMU	*32*	*27*	*24*	*10*	*0*	*2*

E=40

S=-37 S=95

Для данного примера

Д(Х)= 50(1/40-95- (37/40)^-37,98; о -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:

$$Я(X)=M(X)/Д(X)*100\%$$

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 У	Признак X			
		x_t = o	x_t = i	
	Уг=О	a	b	*a+b*
	У>=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:

№ испытуемого	*Переменная X*	*Переменная Y*
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 X			
		$x_i = 0$		
	УГО	*2*	*3*	*5*
	Уг=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