Дисперсия вариационного ряда и её свойства

Занятие 8

Определение

Дисперсией σ^2 вариационного ряда называется средняя арифметическая квадратов отклонений вариантов от их средней

$$\sigma^2 = \frac{\sum_{i=1}^{m} (x_i - \overline{x})^2 \; n_i}{n}$$

Средним квадратическим отклонением σ называется арифметическое отклонение значение корня квадратного из дисперсии

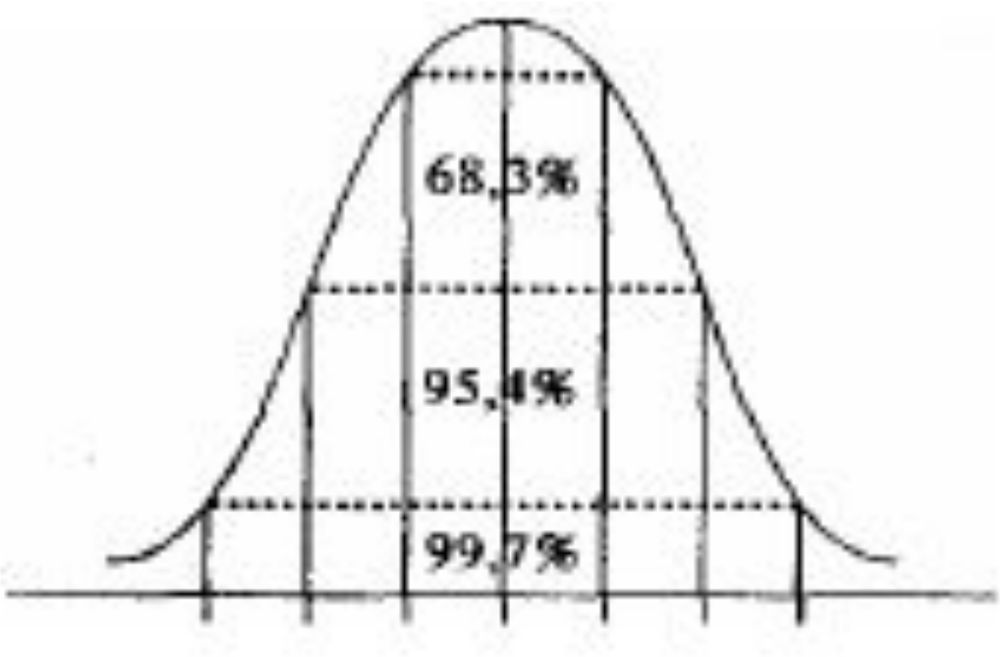
$$\sigma = \sqrt{\sigma^2}$$

Дисперсия и среднее квадратическое отклонение – мера рассеяния

Используется при сравнительных статистических исследованиях, для обоснования ошибки репрезентативности выборочного наблюдения, а также при изучении корреляционных и иных статистических связей между признаками фактора и признаками следствия, или между причиной и следствием.

Среднее квадратическое отклонение позволяет правильно оценить надежность выборочных показателей.

Правило трёх σ (характерно для нормального распределения)



Вычислить дисперсию и среднее квадратичное отклонение

Распределение рабочих предприятия по времени, затрачиваемому на обработку одной детали

Время, затрачиваемое на обработку одной детали, мин.	Число рабочих
2-4	42
4-6	73
6-8	154
8-10	205
10-12	26
Итого	500

$$\overline{X} = \frac{3*42+5*73+\ldots+11*26}{500} = 7.4,$$

$$\sigma^2 = \frac{(3-7.4)^2*42+(5-7.4)^2*73+\ldots+(11-7.4)^2*26}{500} = 4.24$$

$$\sigma = \sqrt{4.24} \approx 2.06$$

Свойства дисперсий

Теорема 1.

Если все варианты увеличить (уменьшить) в k раз, то дисперсия увеличится (уменьшится) в k^2 раз, а среднее квадратическое отклонение - в $|k|$ раз.

$$\frac{\sum_{i=1}^{m}\left(k\,x_i - k\,\bar{x}\right)^2 * n_i}{n} = k^2\,\sigma^2$$

где $\bar{x}$ - средняя арифметическая,

σ^2 - дисперсия вариационного ряда.

Свойства дисперсий

Теорема 2.

Если варианты увеличить или уменьшить на одну и ту же постоянную величину, то дисперсия не изменится.

$$\frac{\sum_{i=1}^{m}\left[(x_i+c)-(\bar{x}+c)\right]^2 * n_i}{n} = \frac{\sum_{i=1}^{m}(x_i-\bar{x})^2 * n_i}{n}$$

Свойства дисперсий

Теорема 3.

Если веса увеличить или уменьшить в одно и то же число раз, то дисперсия не изменится.

$$\frac{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2 k\, n_i}{\sum_{i=1}^{m} k\ n_i}=\frac{\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2\ n_i}{n}$$

Свойства дисперсий

Теорема 4.

Дисперсия равна средней арифметической квадратов вариантов на соответствующие им веса без квадрата средней арифметической, т.е.

$$\sigma^2 = \frac{\sum_{i=1}^{m} x_i^2 \, n_i}{n} - \left(\bar{x}\right)^2$$

Совокупность разбита на l непересекающихся групп

Групповой дисперсией σ_j^2, называется дисперсия распределения членов j -ой группы относительно их средней – т.е. групповой средней $\bar{x}_j$

$$\sigma_j^2 = \frac{\sum_{i=1}^{m} \left(x_i - \bar{x}_j\right)^2 m_i}{N_j}$$

где m_i - частоты вариантов в группе,

$N_j = \sum_{i=1}^{m} m_i$ - объем группы.

Определения

Межгрупповой дисперсией δ^2 называется средняя арифметическая квадратов отклонений групповых средних $\overline{x}_j$ всех непересекающихся групп от общей средней $\overline{x}$, т.е.

$$\delta^2 = \frac{\sum_{i=1}^{l} \left(\overline{x}_j - \overline{x}\right)^2 N_j}{n}$$

где $N_j \ (j = 1, 2, \dots, l)$ – объемы групп.

Определения

Средней групповых дисперсий $\overline{\sigma}^2$ называется средняя арифметическая групповых дисперсий, т.е.

$$\overline{\sigma}^2 = \frac{\sum_{j=1}^{l} \sigma_j^2 * N_j}{n}$$

где N_j $(j = 1, 2, \ldots, l)$ – объем непересекающихся групп.

Свойства дисперсий

Дисперсия распределения членов всей совокупности относительно общей средней называется ***общей дисперсией***.

Теорема 5 (правило сложения дисперсий).

Общая дисперсия равна сумме средней групповых дисперсий $\overline{\sigma}^2$ непересекающихся групп, на которые разбита совокупность, и межгрупповой дисперсии δ^2, т.е.

$$\sigma^2 = \overline{\sigma}^2 + \delta^2$$

Определение

Отношение среднего квадратичного отклонения к средней величине признака, вычисленное в процентах, называется ***коэффициентом вариации***:

$$v = \frac{\sigma}{\bar{x}} * 100\ \%$$

Пример. Вычислить межгрупповую дисперсию распределения рабочих по заработной плате по цехам.

Заработная плата, руб.	Число рабочих в цехах			Всего
	№ 1	№ 2	№ 3	
70-80	7	1	-	8
80-90	12	5	-	17
90-100	15	9	4	28
100-110	6	18	8	32
110-120	-	12	32	44
120-130	-	5	16	21
Итого	40	50	60	150

Решение

Вычислим среднюю заработную плату рабочих цеха № 1. Для этого переходим к соответствующему дискретному распределению.

Заработная плата, x_i руб.	Число рабочих	$x_i * n_i$
75	7	75*7
85	12	85*12
95	15	95*15
105	6	105*6
Итого	40	3600/40=90

Решение

Аналогично вычисляем средние групповые для цеха № 2 и № 3 – 105 и 115 соответственно.

Вычисляем общую среднюю – это средняя заработная плата всех рабочих предприятия

$$\bar{x} = \frac{90*40+105*50+115*60}{40+50+60} = 105\ (руб.)$$

В соответствии с формулой вычисления межгрупповой дисперсии, получаем

$$\delta^2 = \frac{(90-105)^2*40+(105-105)^2*50+(115-105)^2*60}{150} = 100$$

Моменты вариационного ряда

Моментом k-го порядка $M_k(a)$ варьирующего признака X по отношению к значению a называют среднее математическое из k-х степеней отклонений значений признака от a, т. е.

$$M_k(a) = \overline{(X-a)^k} = M(X-a)^k$$

Моменты вариационного ряда

Если a = 0, момент называется *начальным* ν_k ,
а при $a = \overline{X}$ его называют *центральным* μ_k .

$$\nu_k = M\left(X^k\right) = \sum_{i=1}^{l} x_i^k \; \frac{m_i}{n}$$

$$\mu_k = M\left(X - \overline{X}\right)^k = \sum_{i=1}^{l} \left(x_i - \overline{X}\right)^k \; \frac{m_i}{n}$$

Определения

За показатель отклонения распределения признака X от симметрии относительно X принимают величину

$$\alpha = \frac{\mu_3}{\sigma^3}$$

называемую ***асимметрией***

Определения

Эксцессом называют величину

$$\varepsilon = \frac{\mu_4}{\sigma^4} - 3$$

Он показывает степень крутости кривой распределения признака X по сравнению с крутостью нормального распределения дисперсия которого равна σ^2.

Замечание

Если $\varepsilon = 0$, то распределение нормальное.

Если $\varepsilon > 0$, то крутость положительная и кривая распределения имеет более острую вершину, чем при нормальном распределении.

Если $\varepsilon < 0$, то крутость отрицательная и кривая имеет более плоскую вершину. Возможно даже, что в центре распределения будут выемки (двухмодальная кривая).

Значения эксцесса лежат на полусегменте $[-3;\ +\infty)$.

Замечание

Ошибки асимметрии и эксцесса вычисляются соответственно по формулам:

$$E_\alpha = \sqrt{\frac{6*(n-1)}{(n+1)(n+2)}}$$

$$E_\varepsilon = \sqrt{\frac{24*n*(n-2)(n-3)}{(n+1)^2\ (n+3)(n+5)}}$$

СПАСИБО ЗА ВНИМАНИЕ