

Лекция – 3 (2ч)

Тема: Определение основных статистических показателей для характеристик данных

План:

1. Среднее арифметическое значение.
2. Среднее квадратическое (или стандартное) отклонение.
3. Стандартная ошибка средней арифметической или ошибка репрезентативности.
4. Коэффициент вариации

Цель: научиться определять основные статистические характеристики вариационного ряда, делать по ним выводы о компактности группы по заданному признаку.

Теоретические сведения

Наиболее распространенными характеристиками совокупности являются:

- 1) среднее арифметическое значение (\bar{X})
- 2) среднее квадратическое (или стандартное) отклонение (σ);
- 3) стандартная ошибка среднего арифметического ($m_{\bar{X}}$);
- 4) коэффициент вариации (V).

1. Среднее арифметическое значение. Средним арифметическим называется частное от деления суммы всех значений вариант рассматриваемой совокупности на их число (n):

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{или} \quad \bar{X} = \frac{\sum x_i}{n},$$

где: \sum — знак суммирования,
 x_i — варианты или значения признака,
 n — объем выборки.

Среднее арифметическое обладает следующими свойствами:

- а) измеряется в тех же единицах, что и основные варианты;
- б) если каждое число совокупности уменьшить (увеличить) на одно и то же число, то её среднее значение уменьшится (увеличится) на это же число;
- в) если каждое число совокупности увеличить (уменьшить) в несколько раз, то её среднее значение увеличится (уменьшится) в такое же число раз;

г) сумма отклонений статистических данных совокупности от их точного среднего значения всегда равна нулю:

$$\sum_{i=1}^n (x_i - \bar{X}) = 0 \quad (2)$$

2. Среднее квадратическое (или стандартное) отклонение (σ).

Основной мерой статистического измерения изменчивости признака у членов совокупности служит среднее квадратическое отклонение (σ) (сигма) или, как часто ее называют, стандартное отклонение. Теория вариационной статистики показала, что для характеристики любой генеральной совокупности, имеющей нормальный тип распределения достаточно знать два параметра: среднюю арифметическую и среднее квадратическое отклонение. Эти параметры заранее не известны и их оценивают с помощью выборочной средней арифметической и выборочного стандартного отклонения, которые вычисляются при обработке случайной выборки.

В основе среднего квадратического отклонения лежит сопоставление каждой варианты (x_i) со средней арифметической данной совокупности. Так как в совокупности всегда будут варианты как меньше, так и больше, чем она, то сумма отклонений ($x_i - \bar{X}$), имеющих знак "-", будет погашаться суммой отклонений, имеющих знак "+", т.е. $\sum(x_i - \bar{X}) = 0$.

Отклонение вариант от своей средней арифметической выражает изменчивость признака. Если бы изменчивость признака у членов совокупности отсутствовала, тогда разность ($x_i - \bar{X}$) = 0. Но т.к. $\sum(x_i - \bar{X})$ всегда равна нулю, то для измерения изменчивости берут отклонение в квадрате, т.е. $(x_i - \bar{X})^2$. Если просуммировать квадраты отклонений, то эта сумма не будет равна нулю. А чтобы получить коэффициент, способный измерить изменчивость, берут среднее отклонение из выражения:

$$\sigma^2 = \frac{\sum(x_i - \bar{X})^2}{n}$$

Величина σ^2 называется девиантой (или взвешенной дисперсией), вариантой (или средним квадратом). Тогда среднее квадратическое отклонение имеет формулу:

$$\sigma = \pm \sqrt{\frac{\sum(x_i - \bar{X})^2}{n}} \quad \text{при } n > 30 \quad (3)$$

или

$$\sigma = \pm \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}} \quad \text{при } n \leq 30. \quad (4)$$

Свойства среднего квадратического (стандартного) отклонения:

1. Стандартное отклонение всегда измеряется в тех же единицах измерения, что и основные варианты.

2. Чем больше (σ) , тем больше изменчивость признака.

3. В вариационных рядах с нормальным распределением частот 99,7% всех членов совокупности находящихся в границах от x_1 до x_2 , которые отстоят от средней арифметической на величину от -3σ до $+3\sigma$. За пределами $\pm 3\sigma$ находятся только 0,3% всех членов совокупности.

4. При вычислении стандартное отклонение определяют с точностью на один десятичный знак больше, чем точность, которую применяют для вычисления средней арифметической для того же ряда.

3. Стандартная ошибка средней арифметической или ошибка репрезентативности (m) характеризует колебания средней. При этом необходимо отметить, что чем больше объем выборки, тем меньше разброс средних величин.

Стандартная ошибка средней вычисляется по формуле:

$$m_{\bar{X}} = \pm \frac{\sigma}{\sqrt{n}}. \quad (5)$$

В современной научной литературе средняя арифметическая представляется вместе с ошибкой репрезентативности:

$$\bar{x} = \bar{X} \pm m_{\bar{X}} \quad (6)$$

4. Коэффициент вариации (V).

Изложенные выше характеристики совокупности (средняя арифметическая и среднее квадратическое отклонение) имеют один недостаток: они дают показатель изменчивости признака в именованных величинах, а не в относительных. Поэтому сопоставление (или сравнение) разноименных признаков по этим параметрам невозможно.

В этом случае удобно пользоваться коэффициентом изменчивости признака, который выражается в относительных величинах, а именно в процентах, и вычисляется по формуле:

$$V = \frac{\sigma}{\bar{X}} \cdot 100\% \quad (7)$$

Чем больше V, тем более изменчив признак. Значения коэффициента вариации, не выходящие за пределы **10%**, принято считать нормальными.

Если **V>20%**, то выборка некомпактна по заданному признаку.

Пример 5.1.

Определить основные статистические показатели результатов прыжка в длину группы спортсменов, если данные выборки таковы:

x_i , см ~ 185, 171, 190, 170, 190, 178, 188, 174, 193, 178, 176, 180, 175, 176, 180, 192 (n=16).

Решение:

1. Для вычисления основных статистических показателей составим рабочую таблицу следующего вида:

(x_i)	$(x_i - \bar{X})$	$(x_i - \bar{X})^2$
185	4	16
171	-10	100
190	9	81
170	-11	121
190	9	81
178	-3	9
188	7	49
174	-7	49
193	12	144
178	-3	9
176	-5	25
180	-1	1
175	-6	36
176	-5	25
180	-1	1
192	11	121
$\bar{X}=181$	$\Sigma(x_i - \bar{X})=0$	$\Sigma(x_i - \bar{X})^2=868$

2. Для определения среднего арифметического используем формулу:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n},$$

$$\bar{X} = \frac{2896 \text{ см}}{16} \approx 181 \text{ см}$$

3. Для вычисления среднего квадратического отклонения определяем разности $(x_i - \bar{X})$ и заносим во 2 столбец. Находим квадраты этих разностей, записываем их в 3-ий столбец и рассчитываем $\sum(x_i - \bar{X})^2$ Тогда:

$$\sigma = \pm \sqrt{\frac{\sum(x_i - \bar{X})^2}{n-1}},$$

$$\sigma = \pm \sqrt{\frac{868}{15}} \approx \pm 7,6 \text{ см.}$$

4. Определить стандартную ошибку среднего арифметического, используя формулу:

$$m_{\bar{X}} = \pm \frac{\sigma}{\sqrt{n}},$$

5. Определим ошибку репрезентативности по формуле:

$$m_{\bar{X}} = \pm \frac{7,6 \text{ см}}{\sqrt{16}} = \pm \frac{7,6 \text{ см}}{4} = \pm 1,9 \text{ см.}$$

6. Определим коэффициент вариации по формуле:

$$V = \frac{\sigma}{\bar{X}} \cdot 100\%,$$

$$V = \frac{7,6}{181} \cdot 100\% \approx 4,2\%.$$

Выводы: 1) среднее значение высоты прыжка составляет 181 см;

2) по показателю высоты прыжка группа компактна, т.к. стандартное отклонение составляет $\pm 7,6$ см, а коэффициент вариации — 4,2 % (менее 10%).