

PSYCHOLOGIE



«Математические методы в психологии»

«Первичные описательные статистики»

*Подготовил: ст. преподаватель
Дмитриева С.Ю.*



Определение

К первичным описательным статистикам (Descriptive Statistics) обычно относят числовые характеристики распределения измеренного на выборке признака.



Каждая такая характеристика отражает *в одном числовом значении* свойство распределения *множества результатов измерения*: с точки зрения их *расположения на числовой оси* либо с точки зрения их **ИЗМЕНЧИВОСТИ.**



Основное назначение каждой из первичных описательных статистик

Замена множества значений признака, измеренного на выборке, одним числом (например, средним значением или мерой центральной тенденции).

Компактное описание группы при помощи первичных статистик позволяет интерпретировать результаты измерений, в частности, путем сравнения первичных статистик разных групп.

PSYCHOLOGIE



МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ



Определение

Мера центральной тенденции – это число, характеризующее выборку по уровню выраженности измеренного признака.



Существуют **3** способа определения «центральной тенденции»



Мода



Медиана



Среднее



МОДА

(МОДАЛЬНЫЙ ИНТЕРВАЛ ПРИЗНАКА-номинативные данные)

Это такое *значение* (*не частота признака*) из множества измерений, которое встречается наиболее часто.

Моде соответствует наибольший подъем (вершина) графика распределения частот.



Если график распределения частот имеет одну вершину, то такое распределение называется *унимодальным*.

Когда два соседних значения встречаются одинаково часто и чаще, чем любое другое значение, мода есть *среднее этих двух значений*.



Бимодальное распределение имеет на графике распределения 2 вершины, даже если частоты этих двух вершин не строго равны.



Пример

Среди 8 значений признака (3, 7, 3, 5, 7, 8, 7, 6)

$$M_0 = 7$$



МЕДИАНА

Это такое значение признака, которое делит упорядоченное (ранжированное) множество данных пополам так, что одна половина всех значений меньше медианы, другая – больше.



Т.о., первым шагом при определении медианы является упорядочивание (ранжирование) всех значений по возрастанию или убыванию.



Медиана определяется следующим образом:

Если данные содержат нечетное число значений

(8, 9, 10, 13, 15), то медиана есть центральное значение, т. е. $Md=10$

Если данные содержат четное число значений

(5, 8, 9, 11), то медиана есть точка, лежащая посередине между двумя центральными значениями, т.е. $Md=(8+9) : 2 = 8,5$



СРЕДНЕЕ (выборочное среднее, среднее арифметическое)

Определяется как сумма всех значений измеренного признака, деленная на количество суммированных значений.

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



Как образуется выборочная средняя?

Допустим, дана какая-то совокупность из числовых значений, которая состоит из n единиц.

Все эти единицы образуют так называемую выборку.

Сумма всех этих чисел будет формулой выражаться как $\sum X_i$ (X_i - это какое-либо из значений этой выборки, где $i = 1, 2, 3 \dots i-1, i$.

То есть i - это номер значения из выборки).

Тогда, для того чтобы найти выборочную среднюю, необходимо сложить все значения из данной выборки и поделить на их количество n .



Выборочное среднее

– **Выборочное среднее** есть просто среднее арифметическое наблюдаемых значений

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N}$$

– Допустим что у нас есть 5-летняя выборка из 60 месячных доходностей r_1, r_2, \dots, r_{60} некоторой акции

– Тогда выборочное среднее вычисляется по формуле

$$\frac{r_1 + r_2 + \dots + r_{60}}{60}$$





Сумма всех отклонений от среднего равна **0**

Сумма всех отклонений от среднего равна **0**

PSYCHOLOGIE



КВАНТИЛИ РАСПРЕДЕЛЕНИЯ



Квантиль

Это точка на числовой оси измеренного признака, которая делит всю совокупность упорядоченных измерений на 2 группы с известным соотношением их численности (один из квантелей – медиана).



Процентили

Это 99 точек – значений признака (P_1, \dots, P_{99}), которые делят упорядоченное (по возрастанию) множество наблюдений на 100 частей, равных по численности. Определение конкретного значения процентиля аналогично определению медианы.



Например

При определении 10-го перцентиля P_{10} , сначала все значения признака упорядочиваются по возрастанию. Затем отсчитывается 10 % испытуемых, имеющих наименьшую выраженность признака.

P_{10} будет соответствовать тому значению признака, который отделяет эти 10 % испытуемых от остальных 90 %.



Квартили

Это 3 точки – значения признака (P_{25} , P_{50} , P_{75}), которые делят упорядоченное по возрастанию множество наблюдений на 4 равные части.

Первый квартал = 25 процентилю, второй = 50 процентилю или медиане, третий = 75 процентилю.



Где используются?

Квартили и процентиля используются для определения частоты встречаемости тех или иных значений (интервалов) измеренного признака и для выделения подгрупп и отдельных испытуемых, наиболее типичных или нетипичных для данного множества наблюдений.

PSYCHOLOGIE



МЕРЫ ИЗМЕНЧИВОСТИ



Что это и зачем это нужно!

Меры центральной тенденции отражают уровень выраженности измеренного признака.

Не менее важной характеристикой является **выраженность индивидуальных различий** испытуемых по измеренному признаку.

Меры изменчивости (Dispersion) применяются в психологии для численного выражения величины межиндивидуальной вариации признака.



Размах (**R**)

Указывает на диапазон изменчивости значений.

Размах – это разность максимального и минимально значения и все!

$$R = X_{\max} - X_{\min}$$



Дисперсия (**D**)

Дисперсия так же входит в перечень терминов теории вероятности - в теории вероятностей - **наиболее употребительная мера отклонения от среднего** (мерарассеяния).



Дисперсия случайной величины – это мера разброса значений случайной величины относительно её математического ожидания.

Обозначается в статистике как сигма в квадрате.



Дисперсия

- **Дисперсия** – σ^2 - это квадрат среднего квадратического отклонения.
- Она представляет собой средний квадрат отклонений вариант от их средней величины.
- Она может быть также вычислена, как разность среднего квадрата значения признака и квадрата среднего арифметического значения признака:

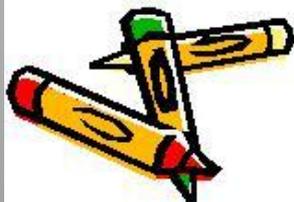
$$\sigma^2 = \overline{X^2} - (\overline{X})^2$$



Дисперсия:

1. Дисперсия постоянной величины равна **0**.
2. Если все значения вариантов признака **X** уменьшить на постоянную величину **A**, то дисперсия не изменится.
3. Если все значения вариантов **X** уменьшить в **K** раз, то дисперсия уменьшится в **K²** раз.
4. На практике часто используют более простую формулу для расчета дисперсий:

5. При малом числе наблюдений (< 30):



$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \quad \text{или} \quad \sigma^2 = \frac{n}{n-1} \left(\overline{X^2} - (\bar{X})^2 \right)$$





Пример:

Вычислим дисперсию признака x для выборки $N=6$.

N	X_i	$(x_i - M_x)$	$(x_i - M_x)^2$
1.	4.	4-3	1
2.	2	2-3	1
3.	4	4-3	1
4.	1	1-3	4
5.	5.	5-3	4
6.	2	2-3	1
	18	0	12



- $M_x = 18 : 6 = 3$. $D_x = 12 : (6-1) = 2, 4$



Стандартное отклонение (сигма,
среднеквадратическое отклонение)

На практике чаще используется стандартное отклонение, а не дисперсия! Это связано с тем, что сигма выражает изменчивость в исходных единицах измерения признака, а дисперсия – в квадратах исходных единиц.

Квадратный корень от дисперсии!

СТАНДАРТНОЕ ОТКЛОНЕНИЕ

$$\sigma = \sqrt{\frac{\sum (x_i - M)^2}{n - 1}}$$

x_i – каждое наблюдаемое значение признака;

M – среднее арифметическое;

n – количество наблюдений (объем выборки).

