

**МЕТОДЫ ОПИСАТЕЛЬНОЙ  
СТАТИСТИКИ**

## 2.1. ПРЕДСТАВЛЕНИЕ КОЛИЧЕСТВЕННЫХ ДАННЫХ

Количественные данные для анализа и интерпретации необходимо обобщить.

Первый этап представления – это упорядочивание данных по величине от максимальной до минимальной.

Такое представление называют *несгруппированным рядом*.



Пример: Группа детей шестилетнего возраста была протестирована по методике Керна-Йерасика (тест на школьную зрелость). Результаты тестирования занесены в таблицу 1

№	Вербальный интеллект
1	14
2	13
3	14
4	14
5	14
6	13
7	12
8	12
9	15
10	13
11	13
12	13
13	13
14	9
15	13
16	13

Оценки проставлялись в алфавитном порядке так, как записаны дети.

Упорядочим ряд данных по убыванию:

15, 14, 14, 14, 14, 13, 13, 13, 13, 13, 13, 13, 13, 12, 12, 9

Получим несгруппированный ряд данных.

***Проранжируем полученные данные*** эти данные, присваивая 1 ранг наибольшему значению.

число 15 будет иметь ранг - 1;

14 принадлежит 4 ранга – 2, 3, 4 и 5.

Общий ранг вычисляем следующим образом:

$$(2+3+4+5)/4=3,5$$

т.е. складываем все ранги и делим на число повторений.

ранг числа 13 - ?

ранг числа 12 - ?

и числа 9 - ?

ТАБЛИЦА 2 РАНЖИРОВАНИЕ НЕСГРУППИРОВАННОГО  
УПОРЯДОЧЕННОГО РЯДА

№ п/п	Вербальный Интеллект	ранг
9	15	1
1	14	3,5
3	14	3,5
4	14	3,5
5	14	3,5
2	13	9,5
6	13	9,5
10	13	9,5
11	13	9,5
12	13	9,5
13	13	9,5
15	13	9,5
16	13	9,5
7	12	14,5
8	12	14,5
14	9	16

Данную таблицу можно записать с помощью, **распределения частот**.

Различные показатели вербального интеллекта размещаются по величине в данном случае от 15 до 9, а справа от каждой оценки указывается число ее повторений.

Каждое число справа называется частотой -  $f$ ,

сумма частот обозначается  $n$ .

Сгруппированные показатели	Частота ( $f$ )
15	1
14	4
13	8
12	2
9	1
$n=$	16

## 2.2. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ РАСПРЕДЕЛЕНИЯ ДАННЫХ

Частотное распределение ничего не говорит о статистических закономерностях, которые описывали бы числовые характеристики изучаемой совокупности.

К характеристикам распределения, описывающим количественно его структуру и строение, относятся:

- характеристики положения;
- рассеивания;
- асимметрии и эксцесса.



## Характеристики положения (центральной тенденции):

- мода ( $M_o$ ),
- медиана ( $M_e$ )
- квантили
- среднее арифметическое ( $\bar{M}$ ).



Определение: величина признака, которая встречается чаще всего в изучаемом ряду, или совокупности называется *модой (Mo)*.

В дискретном ряду Mo определяется без вычисления, как значение признака с наибольшей частотой

Mo = ?

# ПРИМЕР: НАЙТИ МОДУ РЯДА

12, 13, 14, 14, 14, 16, 16, 16, 18, 19

12, 13, 14, 14, 14, 16, 16, 18, 18, 18 19

12, 12, 12, 14, 14, 14



1. Два значения признака, стоящие рядом, встречаются одинаково часто.

мода равна среднему арифметическому этих двух значений

2. Два значения, встречаются также одинаково часто, но не стоят рядом.

ряд данных имеет две моды, т.е. он бимодальный.

3. Если все значения данных встречаются одинаково часто

ряд не имеет моды.

Если в ряду данных встречается два или более равных значений признака, то говорят о неоднородности совокупности.

Определение: *медианой (Me)* – называется такое значение признака, которое делит ряд пополам.

При нечетном числе элементов в ряду данных, медиана равна центральному члену ряда,

При четном среднему арифметическому двух центральных значений ряда.

$Me = ?$

Вычисление медианы имеет смысл только для порядкового признака.

*Среднее арифметическое значение признака вычисляется по формуле:*

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

где  $x_i$  – значения признака,  $n$  – количество данных в рассматриваемом ряду.

Среднее арифметическое значение признака, вычисленное для какой-либо группы, интерпретируется как значение наиболее типичного для этой группы человека.

Однако бывают случаи, когда подобная интерпретация несостоятельна (в случае, если существует большая разница между минимальным и максимальным значениями признака).


*Квантиль* – это такое значение признака, которое делит распределение в заданной пропорции:

слева 0,5%, справа 99,5%;

слева 2,5%, справа 97,5% и т.п.



## Разновидности квантилей:

- 1) Квартили  $Q_1, Q_2, Q_3$  – они делят распределение на четыре части по 25% в каждой;
  - 2) Квинтили  $K_1, K_2, K_3, K_4$  – они делят распределение на пять частей по 20% в каждой;
  - 3) Децили  $D_1, \dots, D_9$ , их девять, и они делят распределение на десять частей по 10% в каждой;
  - 4) Перцентили  $P_1, P_2, \dots, P_{99}$ , их девяносто девять, и они делят распределение на сто частей по 1% в каждой части.
- 

# ХАРАКТЕРИСТИКИ РАССЕЙВАНИЯ

Изучается средний возраст в двух группах, состоящих каждая из 6-ти человек. Значения признака распределились следующим образом:

1 группа – 10, 10, 10, 50, 50, 50       $M_1 = \overline{30}$

2 группа – 30, 30, 30, 30, 30, 30       $M_2 = \overline{30}$

Ошибка произошла из-за разброса значений возраста в этих группах.



Существует несколько способов оценки степени разброса или рассеивания данных.

Основными характеристиками рассеивания являются:

- *размах* ( $R$ ),
- *дисперсия* ( $D$ ),
- *среднеквадратическое (стандартное) отклонение* ( $\sigma$  – сигма),
- *коэффициент вариации* ( $V$ ).

Определение: *размах* – это разность между максимальным и минимальным значениями признака:  $R = x_{max} - x_{min}$ .

Дисперсия показывает разброс значений признака относительно своего среднего арифметического значения, то есть насколько плотно значения признака группируются вокруг  $M$ ;

чем больше разброс, тем сильнее варьируются результаты испытуемых в данной группе, тем больше индивидуальные различия между испытуемыми:

$$D = \frac{\sum_{i=1}^n (x_i - M)^2}{n - 1}$$

Из формулы видно, что дисперсия имеет "квадратный размер": если величина измерена в баллах, то дисперсия характеризует ее разброс в "баллах в квадрате", и т.п.

Большую наглядность в отношении разброса имеет среднеквадратическое отклонение, так как его размерность соответствует размерности измеряемой величины:

$$\delta = \sqrt{D}$$

Коэффициент вариации не имеет размерности, что позволяет сравнивать вариативность случайных величин, имеющих различную природу:

$$V = \frac{\delta}{M} \cdot 100\%$$

# ХАРАКТЕРИСТИКИ АССИМЕТРИИ И ЭКСЦЕССА

Мера *асимметрии* – коэффициент асимметрии ( $As$ ), рассчитываемый по формуле

$$As = \frac{\sum_{i=1}^n (x_i - M)^3}{n\delta^3}$$

Асимметрия характеризует степень асимметричности распределения.

Коэффициент асимметрии изменяется ( $-\infty < As < +\infty$ ), для симметричных распределений  $As=0$ .

Мера *эксцесса* (островершинности) – коэффициент эксцесса ( $E_x$ ), рассчитываемый по формуле:

$$E_x = \frac{\sum_{i=1}^n (x_i - M)^4}{n\delta^4} - 3$$

У студентов первого и второго курса был исследован уровень депрессивного расстройства .  
Сделать сравнительный анализ, используя методы описательной статистики.

1 курс	2 курс
30	24
27	17
23	17
22	15
19	15
19	14
18	14
16	13
15	12
14	12
13	11
12	11
12	8
12	8
10	7
10	7
10	4
10	0

Оценка центральной тенденции:

	1 курс	2 курс
мода	10	17
медиана	15	12
среднее	16	12

Медиана и среднее арифметическое значение на первом курсе выше, чем на втором, следовательно уровень депрессивного расстройства на 1 курсе превышает уровень расстройства на втором курсе.

Однако мода на 2-ом курсе значительно выше, чем на первом, т.е. преобладают более высокие значения уровня депрессивного расстройства



## Оценка разброса данных

	1 курс	2 курс
дисперсия	37,0	23,9
ст. отклонение	6,1	4,9
к. вариации	37,5	39,4

Дисперсия и стандартное отклонение на первом курсе выше, чем на втором, что говорит о более широком разбросе данных и следовательно, можно сделать вывод о том, что вторая выборка более однородна.

### 3. Асимметрия и Эксцесс

	1 курс	2 курс
ассиметрия	0,9	0,5
эксцесс	0,06	0,64

Первая выборка(1 курс) имеет небольшую положительную левостороннюю асимметрию (0,9) и незначительный положительный эксцесс (0,06).

Вторая выборка (2 курс) также имеет небольшую положительную асимметрию (0,5) и положительный эксцесс(0,64).

# ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

1. Представление количественных данных.
  2. Различные этапы представления данных.
  3. Несгруппированные ряды. Упорядоченные ряды.
  4. Ранжирование данных.
  5. Распределение частот.
  6. Числовые характеристики распределения данных. Оценка средних величин. Мода, медиана и средняя арифметическая.
  7. Оценка разброса данных. Коэффициенты вариации. Асимметрия и Эксцесс.
- 