



Первичная обработка данных

Основные определения

- Генеральная совокупность –
множество объектов (людей), о которых будут
сделаны выводы в результате исследования
- Выборка –
часть генеральной совокупности, непосредственно
участвовавшая в обследовании

- **Объем выборки (N, n)** –

Общая численность выборки, показывает, сколько человек было обследовано

- **Абсолютная частота (f_i)** –

показывает, сколько человек имеет данный показатель

- **Относительная частота (f_i')** –
$$f_i' = \frac{f_i}{N} \cdot 100\%$$

показывает, какая часть группы имеет данный показатель

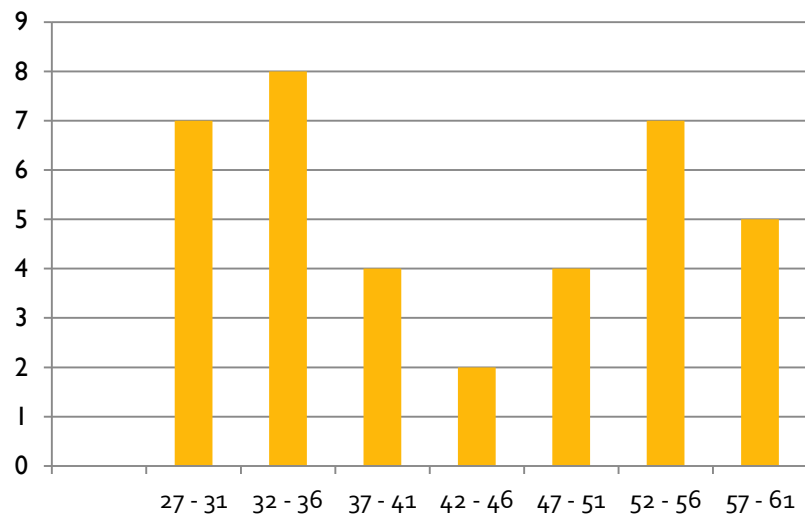
Таблица распределения частот

Личностная направленность	Абсолютная частота, f_i	Относительная частота, f'_i
личностная (на себя)	17	44,7%
коллективистская (на взаимодействие)	14	36,8%
деловая (на задачу),	7	18,4%
Всего	38	100%

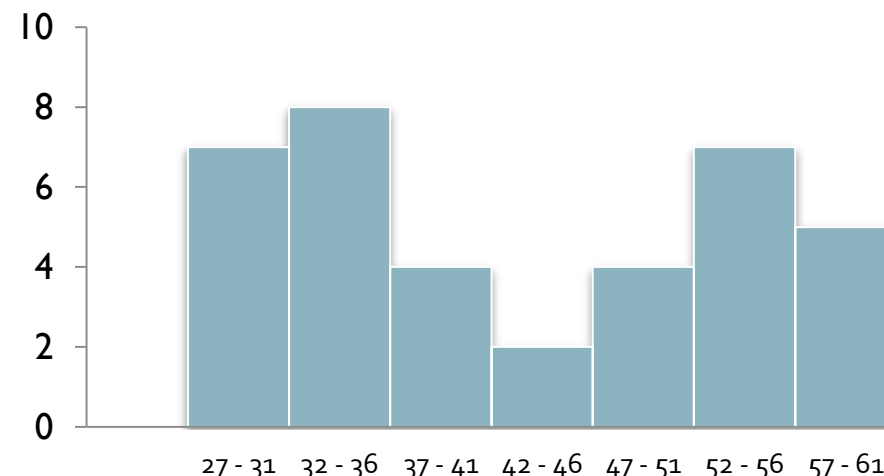
Таблица распределения частот

Тревожность			Абсолютная частота, f_i	Относительная частота, f'_i
уровень	интервал	средняя точка		
1	22 – 26	24	4	5,88%
2	27 – 31	29	6	8,82%
3	32 – 36	34	6	8,82%
4	37 – 41	39	12	17,65%
5	42 – 46	44	18	26,47%
6	47 – 51	49	15	22,06%
7	52 – 56	54	5	7,35%
8	57 – 61	59	2	2,94%
всего			68	100%

Наглядное представление распределения частот

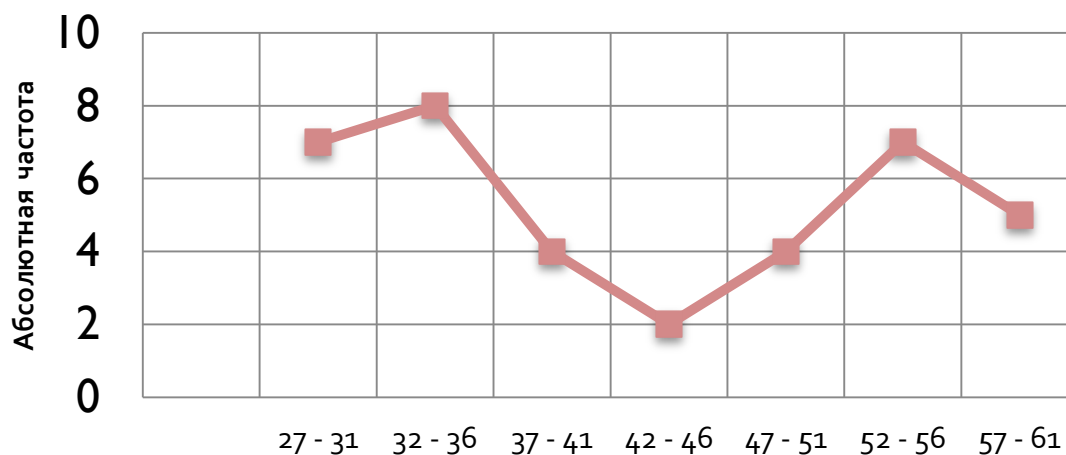


тревожность



- Вертикальные линии

Тревожность



- ПОЛИГОН

гистограмма

- 
- **Общая структура графиков должна предполагать чтение слева направо.**
 - Когда используется возможность изображать количества линейные величины с помощью, например, площадей или объемом. вероятнее всего, что их не удастся верно истолковать.
 - **Вертикальную шкалу следует выбрать так, чтобы на рисунке оказалась нулевая отметка.**
 - Нулевые линии шкал для кривой следует резко отграничивать, от других координатных линий.

- Для кривых, которые имеют шкалу, изображающую проценты, как правило, желательно выделить каким-то образом линию 100% или другие линии, используемые в качестве основы для сравнения.
- Рекомендуется *показывать* не больше координатных линий, чем это необходимо, чтобы облегчить чтение диаграммы
- Цифры на шкалах следует располагать слева и снизу или вдоль соответствующих осей
- Желательно включать в график цифровые данные или изображаемые формулы.

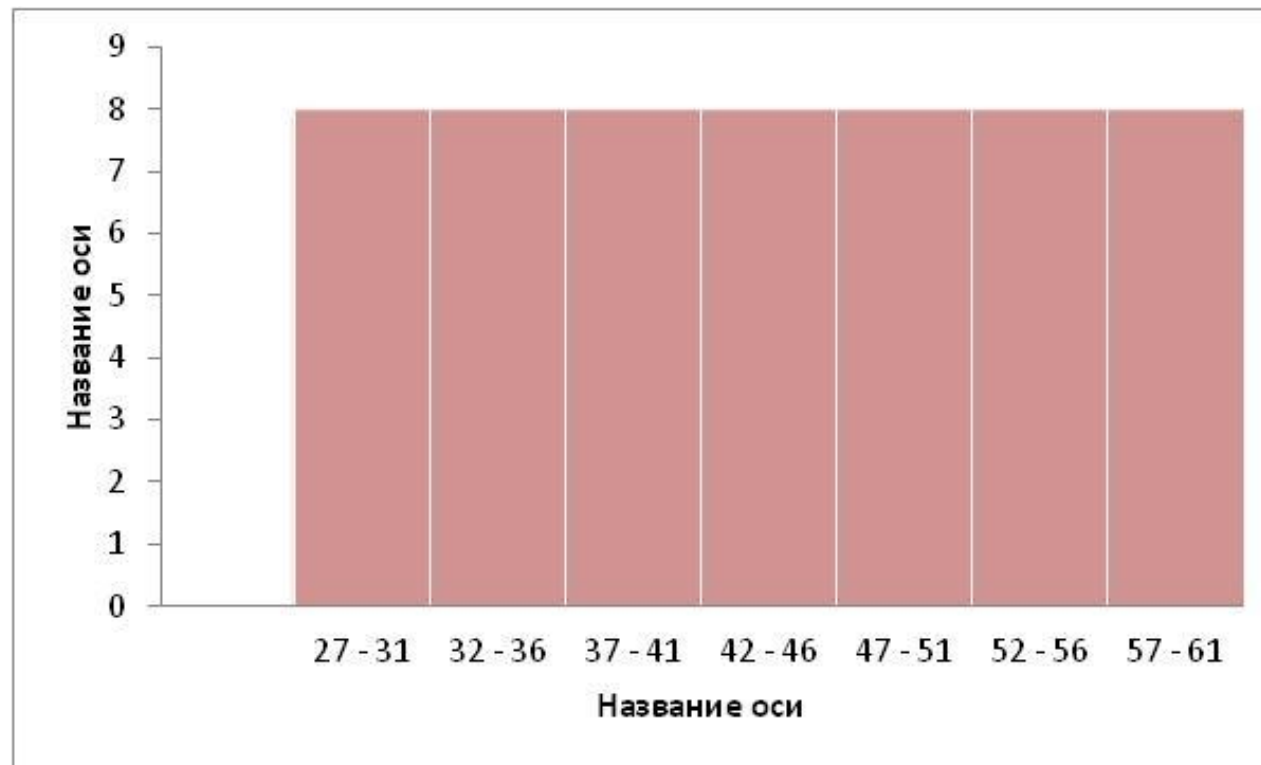
- 
- Если цифровые данные не попали на график, желательно привести данные в таблице, сопровождающей график.
 - Все обозначения и цифры для удобства чтения следует располагать от основания как начала или с правого края как начала.
 - Наименования следует делать возможно яснее и полнее. Если это требуется, необходимо дополнительно вводить подзаголовки или пояснения.

Основные характеристики распределения частот

- Форма графиков распределения
- Характеристики центральных тенденций
- Характеристики вариации данных
- Дополнительные характеристики распределения

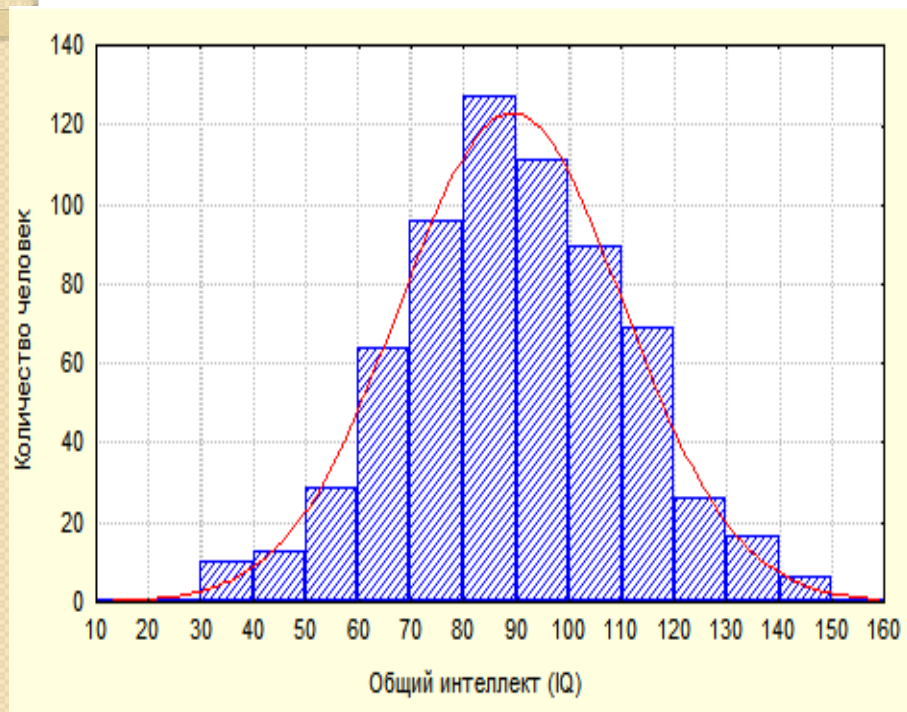
Форма графиков распределения

- Равномерное распределение



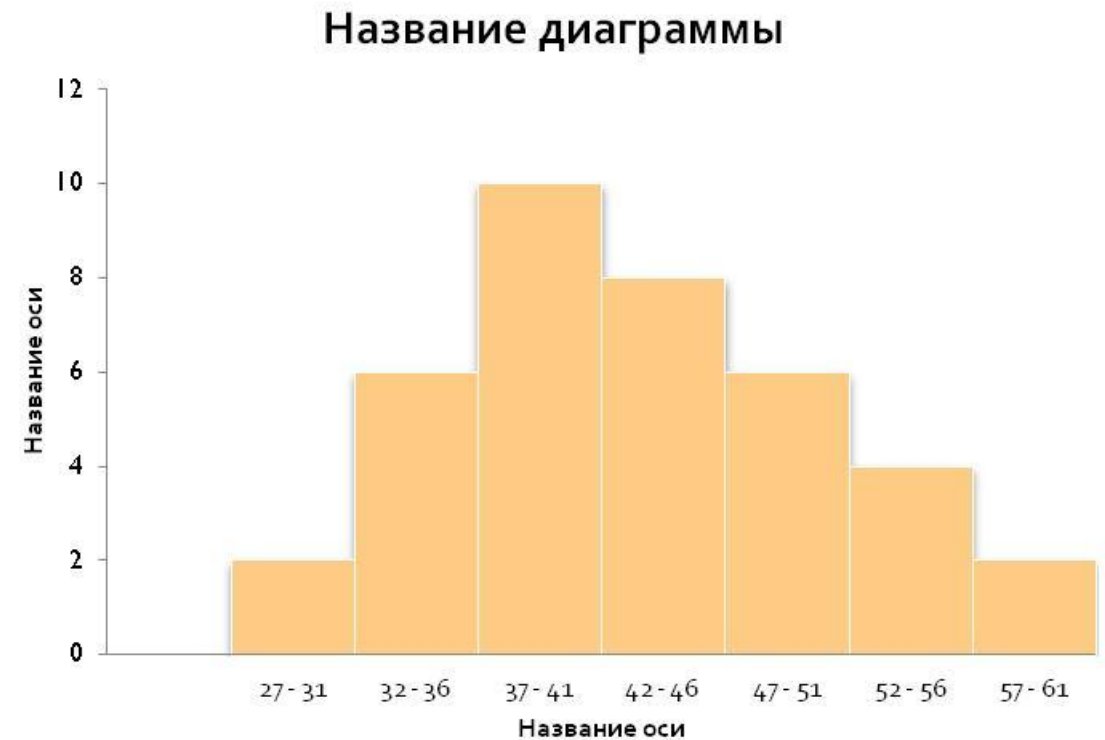
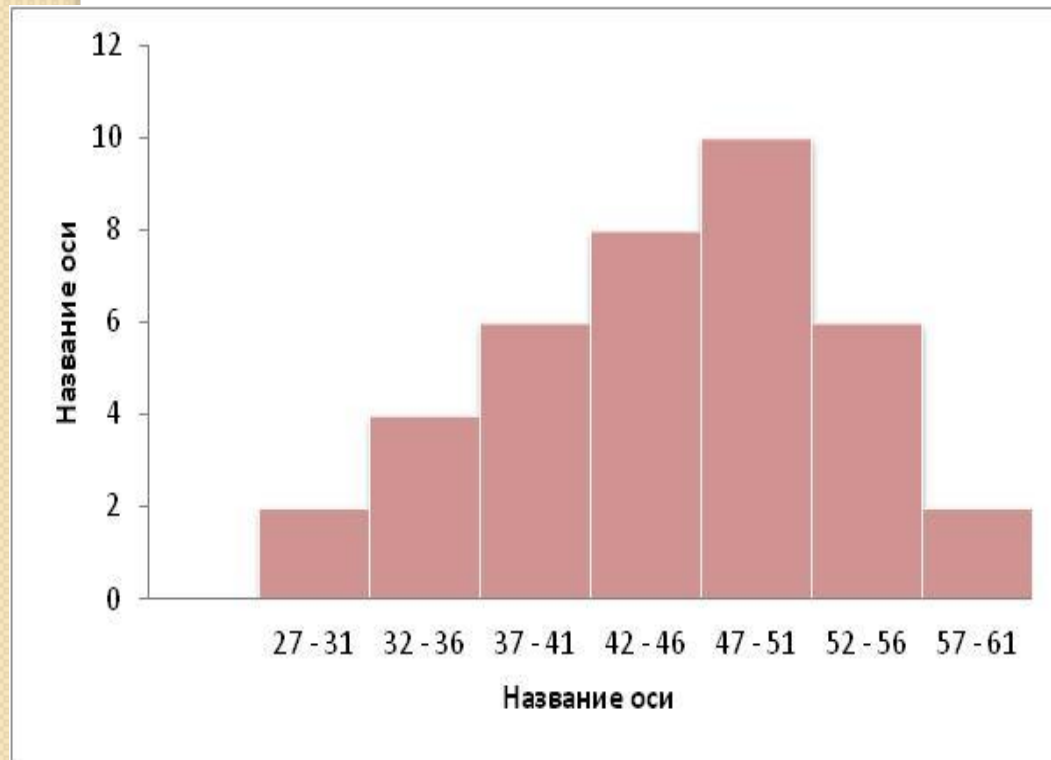
Форма графиков распределения

- Нормальное распределение



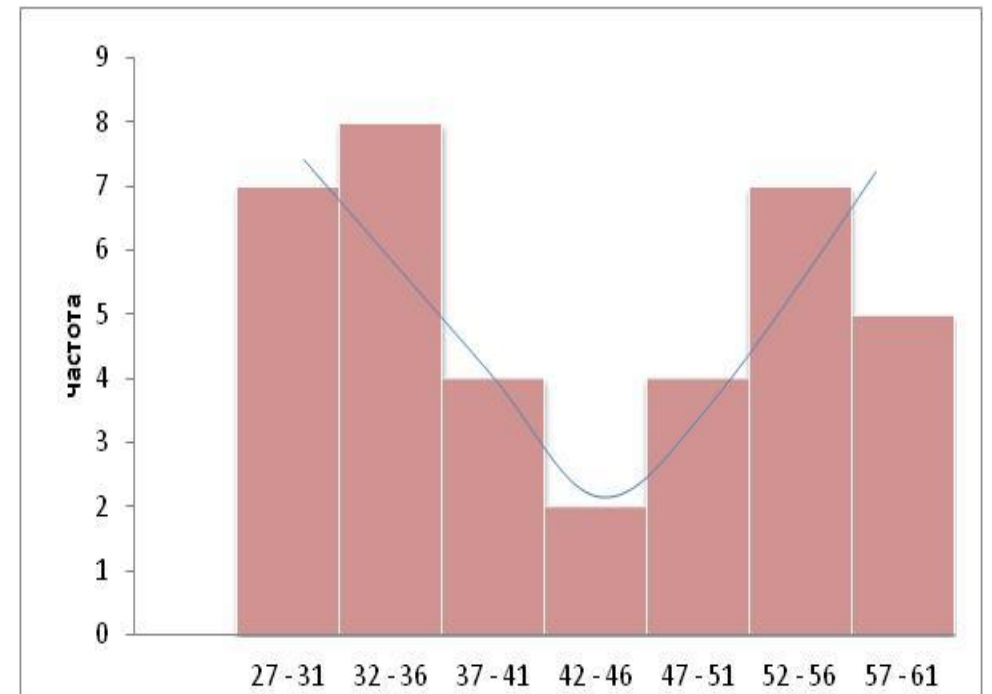
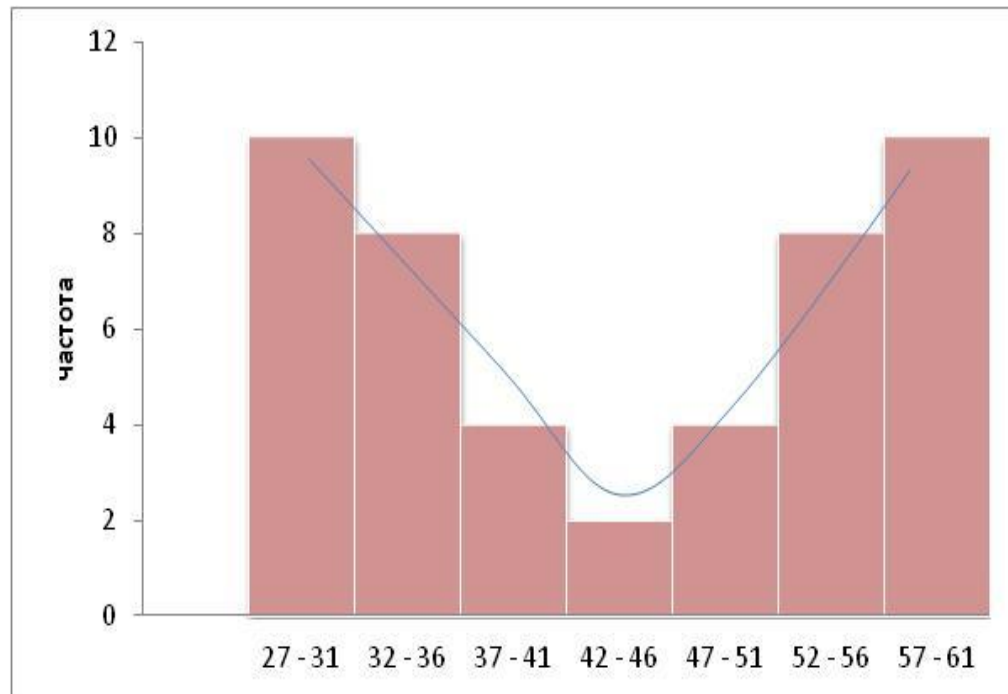
Форма графиков распределения

- Скошенное распределение



Форма графиков распределения

- U - образное распределение



Средние (центральные) характеристики

- **Мода (M_o)** - наиболее часто встречающееся значение;

его называют также модальным значением

1, 2, 3, 4, 4, 5, 7, 7, 7

$$M_o = 7$$

2, 2, 2, 4, 4, 5, 6, 7, 8

$$M_o = 2$$

Средние (центральные) характеристики

- **Медиана (M_e)** - это такое значение случайной величины, которое делит упорядоченную (в порядке возрастания или убывания величины) выборку пополам
- При нечетном количестве измерений за медиану принимается непосредственно центральное значение, справа и слева от него располагается по $(n-1)/2$ значений
- в выборке из 15 упорядоченных значений это будет восьмое значение, а в выборке из 23 значений - двенадцатое и т.д.

- 10,11,3,6,8,6,20,5,5

- 3,5,5,6,6,8,10,11,20

$M_e = 6$

- 3,5,5,6,6,8,10,11,20,22

$M_e = 7$

Средние (центральные) характеристики

- Среднее арифметическое значение M_x ;

$$M_x = \frac{\sum x_i}{n}$$

10т.р, 15т.р., 25т.р, 40т.р., 50т.р.

$$M_x = \frac{\sum x_i}{n} = \frac{10 + 15 + 25 + 40 + 50}{5} = 28$$

10т.р, 15т.р., 25т.р, 40т.р., 500т.р.

$$M_x = \frac{\sum x_i}{n} = \frac{10 + 15 + 25 + 40 + 500}{5} = 118$$

Характеристики вариации (рассеивания) данных

- размах d

$$d_x = x_{max} - x_{min} + 1$$

- Межквартильный размах

- Среднее отклонение MD

$$MD_x = \frac{\sum |x_i - M_x|}{n}$$

Характеристики вариации (рассеивания) данных

- дисперсия σ^2 или D

$$D_x = \sigma_x^2 = \frac{\sum (x_i - M_x)^2}{n} \qquad D = \sigma_x^2 = \frac{\sum (x_i)^2 - n \cdot M_x^2}{n}$$

- среднеквадратическое (стандартное) отклонение σ

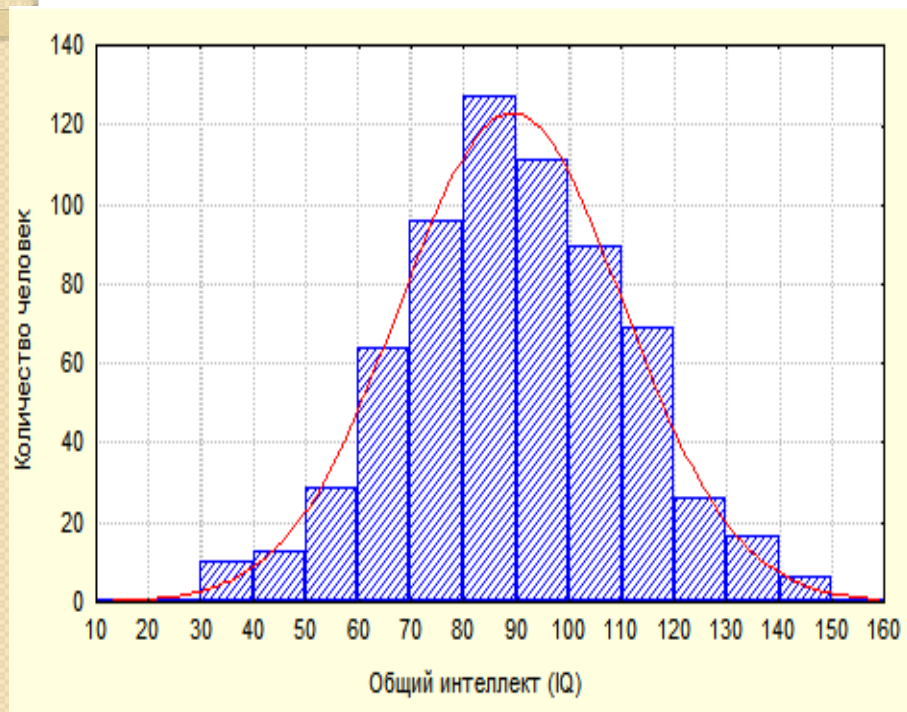
$$\sigma_x = \sqrt{D}$$

- коэффициент вариации V.

$$V := \frac{\sigma}{M_x} \cdot 100 \%$$

Форма графиков распределения

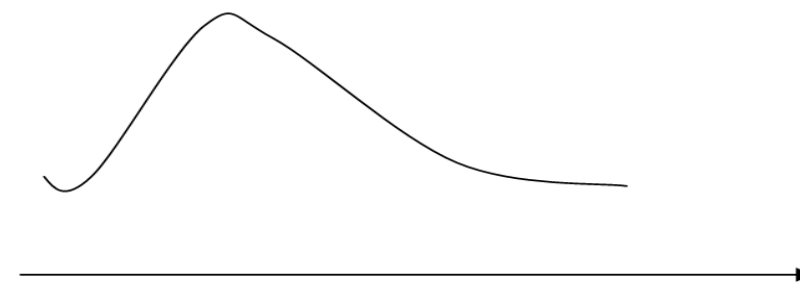
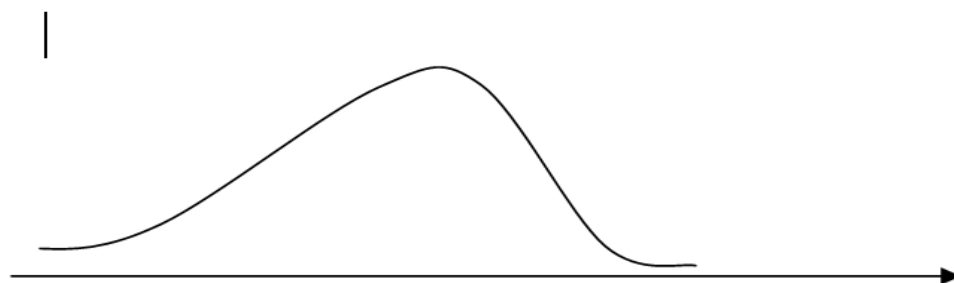
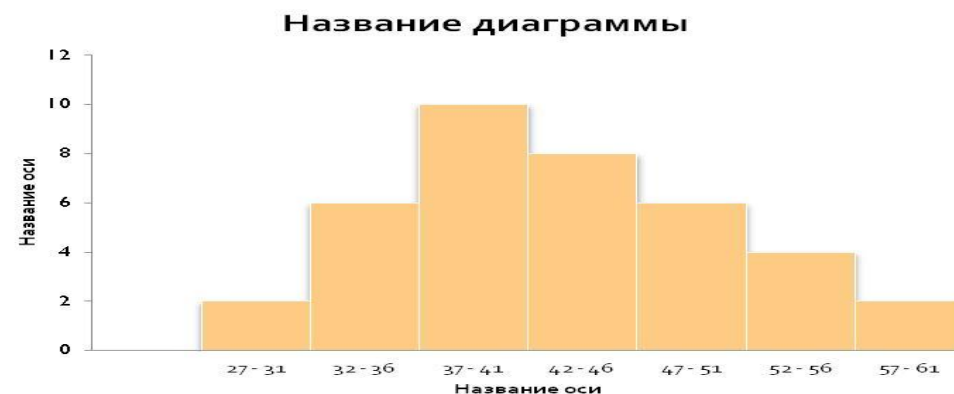
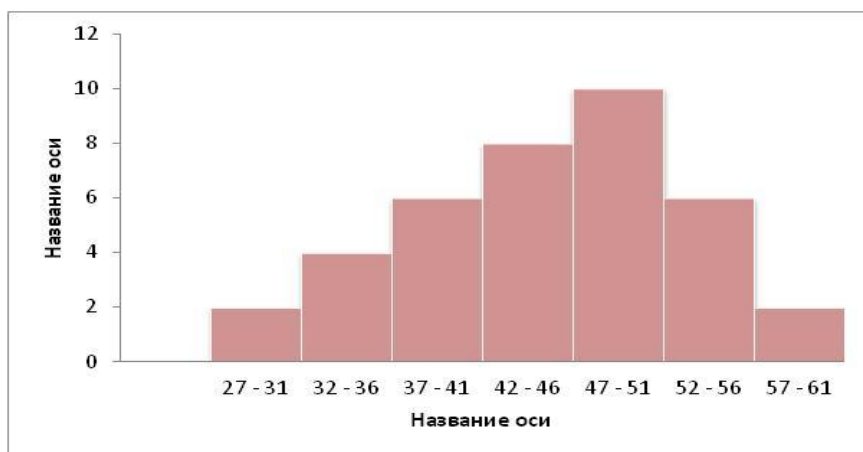
- Нормальное распределение



Дополнительные характеристики

- коэффициент асимметрии A_s

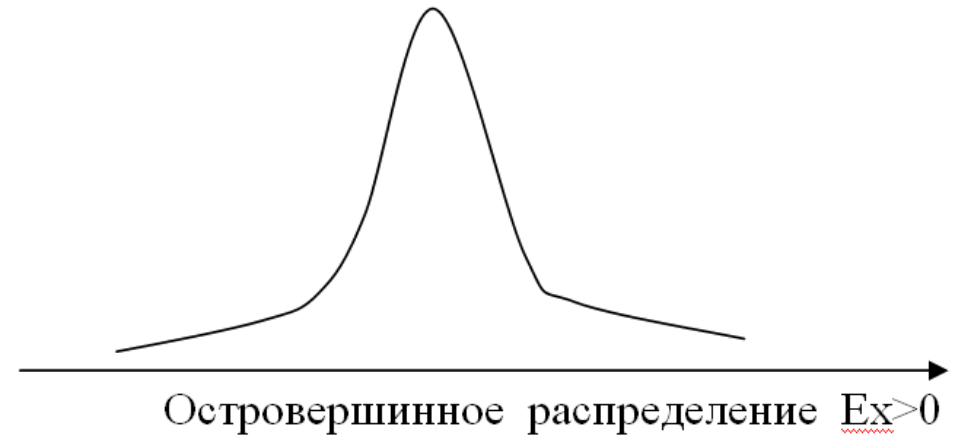
$$A_x = \frac{\sum (x_i - M_x)^3}{n \cdot \sigma_x^3}$$



Дополнительные характеристики

- Коэффициент эксцесса

$$E_x = \frac{\sum (x_i - M_x)^4}{n \cdot \sigma_x^4} - 3$$



Причины эксцесса:

- большая или меньшая степень тяготения переменных к центральной тенденции;
- неоднородность выборки, наложение друг на друга нескольких распределений с одинаковой модой и разной дисперсией