

## **Тема 2. Представление данных**

---

- 2.1. Частоты и распределения частот
- 2.2. Относительные частоты
- 2.3. Таблицы
- 2.4. Графические представления

## 2.1. Частоты и распределения частот

---

**Частота** – количество наблюдений, в которых признак принимает определенное значение или находится в определенном интервале.

Если признак номинальный, порядковый или дискретный количественный, то **дискретное распределение частот** – это таблица, в которой приведены все значения признака и частоты, с которой они встречаются в выборке.

Если признак непрерывный количественный, то **интервальное распределение частот** состоит из некоторого количества интервалов равной длины, на которые делится весь диапазон изменения признака, и соответствующих этим интервалам частот.

# Пример. Политические предпочтения

---

Имеются исходные данные по респондентам:

<u>Респондент</u>	<u>Предпочтение</u>
1	демократы
2	демократы
3	либералы
4	демократы
5	коммунисты
6	либералы
7	коммунисты
...	...
<u>105</u>	<u>демократы</u>
всего 105 респондентов	

Требуется получить  
распределение частот.

# Дискретное распределение

---

## Номинальная шкала

### Политические предпочтения

<b>КАТЕГОРИИ</b>	<b><i>f</i></b>
<b>Демократы</b>	<b>45</b>
<b>Коммунисты</b>	<b>43</b>
<b>Либералы</b>	<b>15</b>
<b>Всего</b>	<b>105</b>

## Порядковая шкала

### Отношение к фильму

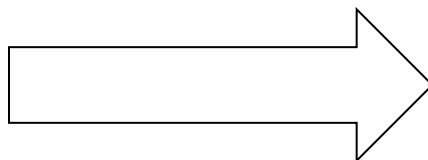
<b>КАТЕГОРИИ</b>	<b><i>f</i></b>
<b>Очень понравился</b>	<b>24</b>
<b>Понравился</b>	<b>12</b>
<b>Фильм средний</b>	<b>10</b>
<b>Не понравился</b>	<b>6</b>
<b>Очень плохой</b>	<b>8</b>
<b>Всего</b>	<b>60</b>

# Интервальное распределение частот

**Интервальное распределение** частот состоит из некоторого количества интервалов равной длины, на которые делится весь диапазон изменения признака, и соответствующих этим интервалам частот.

(n = 77)						
46	59	65	69	71	74	79
49	60	65	69	71	75	80
50	60	65	69	72	75	81
50	60	66	70	72	75	81
52	61	67	70	73	76	83
53	62	67	70	73	76	84
54	62	67	70	73	77	84
55	63	68	70	73	77	85
55	64	68	71	74	78	87
56	64	68	71	74	79	89
58	64	69	71	74	79	90

Данные о весе  
77 человек



ИНТЕРВАЛЫ	<i>f</i>
45-49	2
50-54	5
55-59	5
60-64	10
65-69	14
70-74	20
75-79	11
80-84	6
85-89	3
90-94	1
Итого	77

# Условия для выбора интервалов

---

1. Интервалы не должны пересекаться
2. Интервалы должны охватывать все значения признака
3. Интервалы должны иметь одинаковую длину, за исключением крайних
4. Интервалы не должны иметь пробелов

---

**Иногда вместо абсолютных частот в таблицах указываются относительные частоты или проценты**

# Расчет относительных частот

Результаты экзамена 115 студентов					
СУММА БАЛЛОВ	$f$	$P$	$CP$	%	$C\%$
100	7	0,06	0,06	6,1	6,1
99	9	0,08	0,14	7,8	13,9
98	12	0,10	0,24	10,4	24,3
97	15	0,13	0,37	13,0	37,3
96	18	0,16	0,53	15,7	53,0
95	17	0,15	0,68	14,8	67,8
94	12	0,10	0,78	10,4	78,2
93	10	0,09	0,87	8,7	86,9
92	6	0,05	0,92	5,2	92,1
91	3	0,03	0,95	2,6	94,8
90	6	0,05	1,00	5,2	100
ИТОГО	115	1		100,0	

## Обозначения:

$f$  = частота

$P$  = относительная частота

$CP$  = накопленная относительная частота

% = проценты

$C\%$  = накопленные проценты



# Пример. Политические предпочтения

КАТЕГОРИИ			
Демократы			
Коммунисты			
Либералы			
Всего			

$f$  – абсолютная частота (45)

$P$  – относительная частота ( $45/105 = 0,428$ )

% – процент ( $45/105 * 100\% = 42,8\%$ )

# Пример. Отметки по математике

КАТЕГОРИИ	$f$	$Cf$	$P$	$CP$	$\%$	$C\%$
Отлично	17	17	0.200	0.200	20.0%	20.0%
Хорошо	41	58	0.482	0.682	48.2%	68.2%
Удовл.	20	78	0.236	0.918	23.6%	91.8%
Неуд.	7	85	0.082	1.000	8.20%	100%
Всего	85		1.0		100%	

- накопленные значения (сумма с 1-ой строки по данную)
- 58 учеников (или 0.682 всех учеников, или 68.2% всех учеников) имеют оценку хорошо и выше

## 2.4. Визуальные представления

---

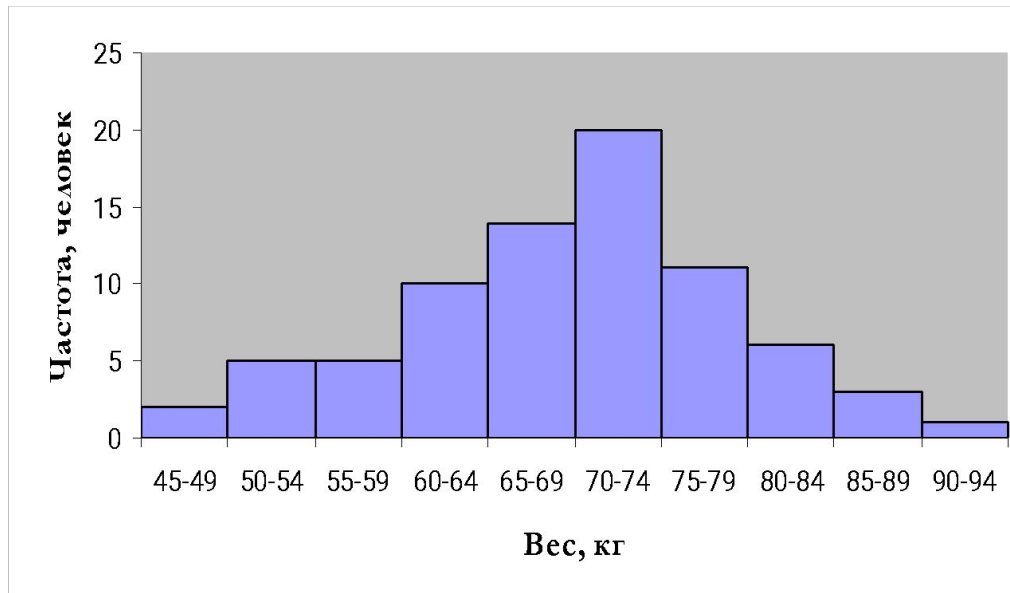
**Графические изображения данных** дают визуальное представление об основных свойствах имеющегося набора данных:

- 1. Размах значений.** Минимальное и максимальное значения.
- 2. Типичные значения.** Какие значения встречаются чаще всего?
- 3. Общая конфигурация данных.** Где сосредоточены основные данные? Симметрично ли они расположены вокруг типичного значения? В какую сторону смещены?
- 4. Характерные особенности.** Имеются ли выбросы? Есть ли значения, которые пропущены?

# Гистограммы частот

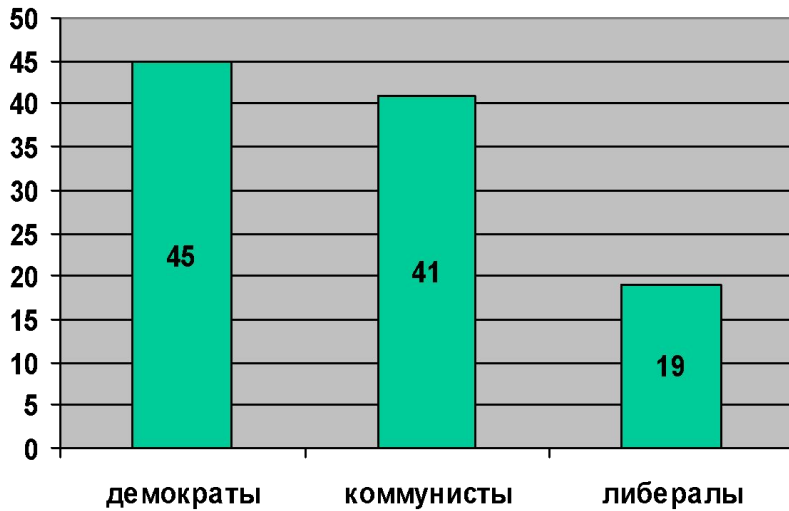
---

**Гистограмма частот** – графическое представление, которое показывает распределение переменной.

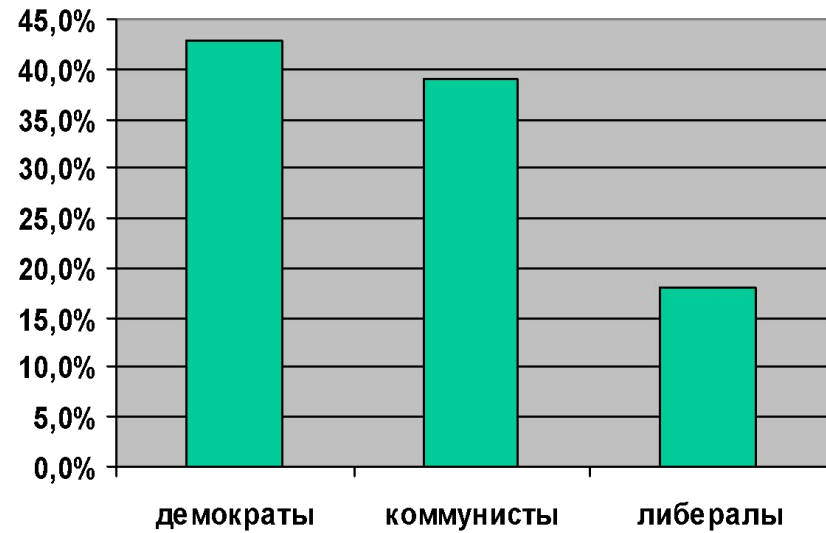


# Гистограммы. Примеры (1)

## Политические предпочтения



**Абсолютная частота**

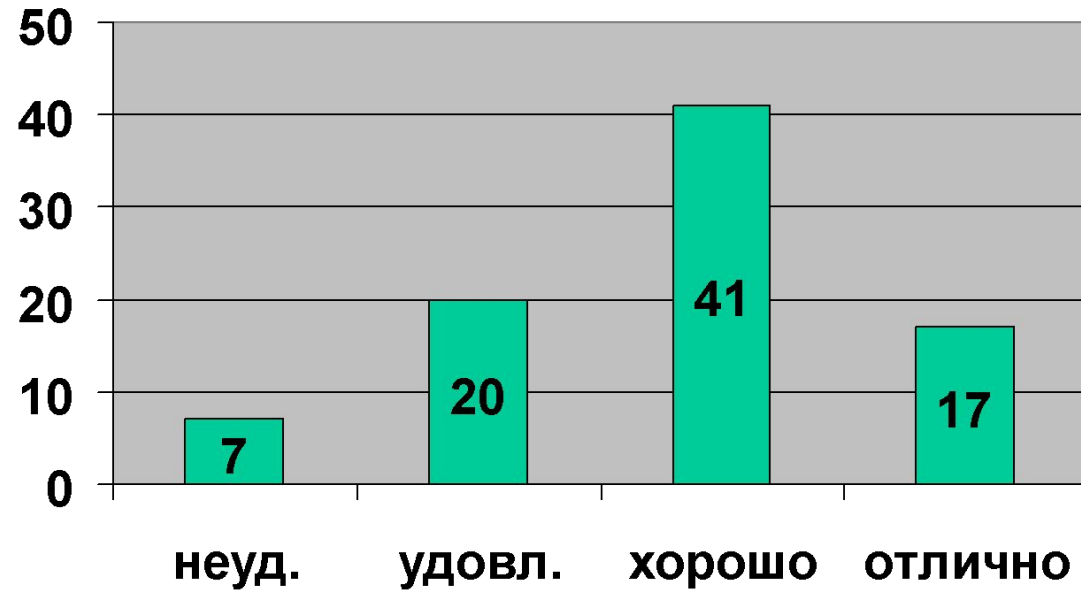


**Проценты**

# Гистограммы. Примеры (2)

---

Отметки по математике  
(абсолютная частота)

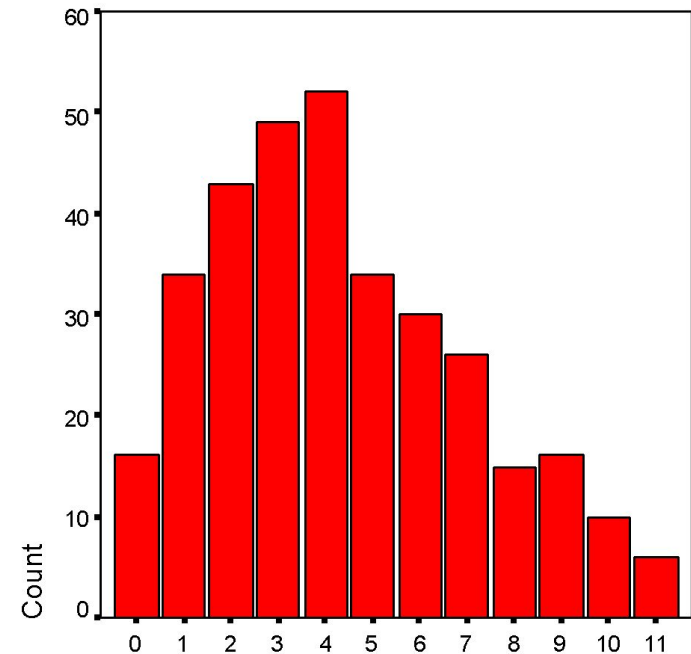


# Пример. Тест первокурсников

Первокурсники экономических факультетов написали тест по математике. В тесте приняли участие 331 человек. Оценивалось 11 заданий.

ВЕЗИИ 12

Ләйиқ	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0	10	25	34	30	34	30	25	10	10	10	10	10
1	0	10	25	34	30	34	30	25	10	10	10	10	10
2	0	10	25	34	30	34	30	25	10	10	10	10	10
3	0	10	25	34	30	34	30	25	10	10	10	10	10
4	0	10	25	34	30	34	30	25	10	10	10	10	10
5	0	10	25	34	30	34	30	25	10	10	10	10	10
6	0	10	25	34	30	34	30	25	10	10	10	10	10
7	0	10	25	34	30	34	30	25	10	10	10	10	10
8	0	10	25	34	30	34	30	25	10	10	10	10	10
9	0	10	25	34	30	34	30	25	10	10	10	10	10
10	0	10	25	34	30	34	30	25	10	10	10	10	10
11	0	10	25	34	30	34	30	25	10	10	10	10	10
12	0	10	25	34	30	34	30	25	10	10	10	10	10



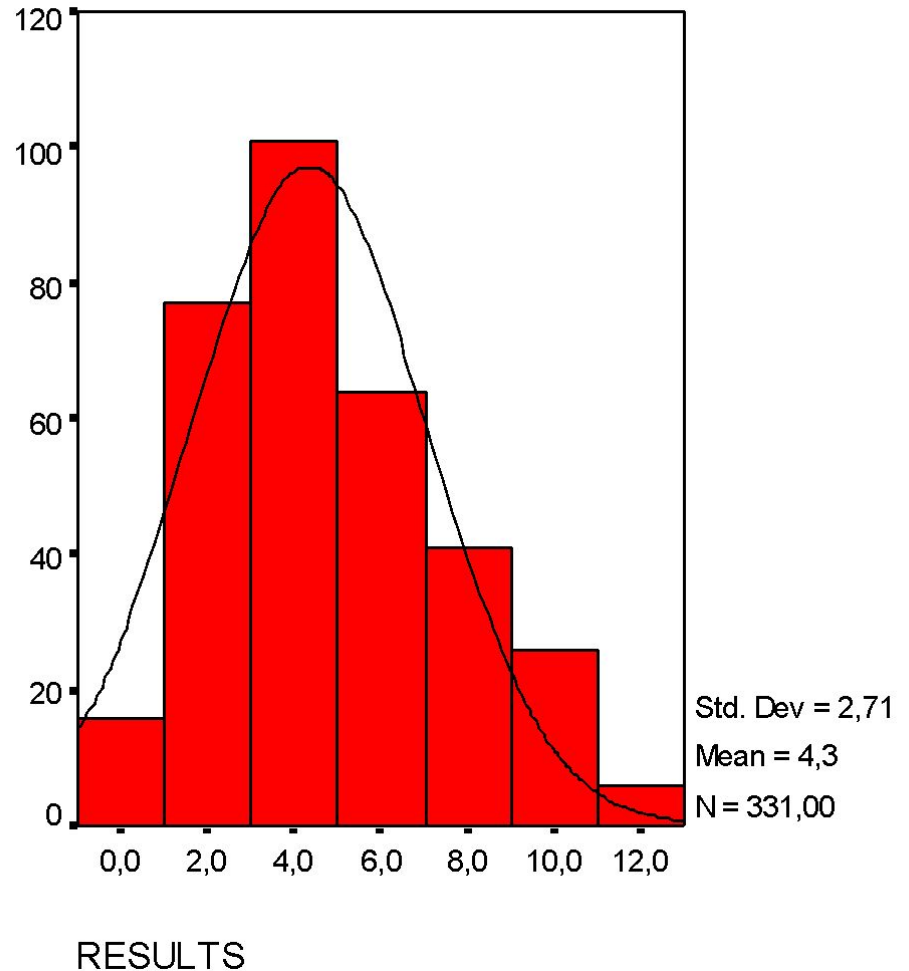
RESULTS

# Нормальное распределение

---

Анализ данных наиболее эффективен в случае, когда гистограмма близка к нормальному распределению.

На гистограмме нарисована линия для нормального распределения со средним 4,3.





# Бимодальное распределение

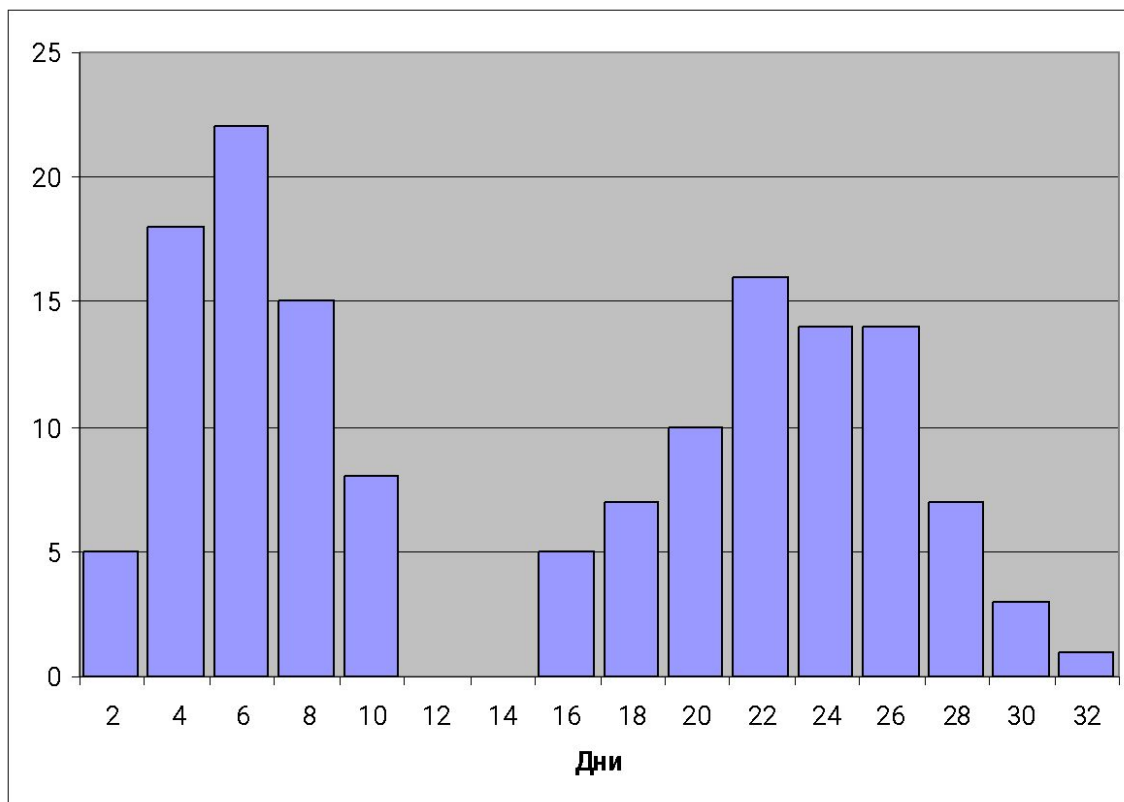
Если в гистограмме видны две отдельные группы данных, говорят о **бимодальном распределении данных**.

Возможно, следует изменить ход анализа:

1 вариант - отказаться от изучения одной из групп

2 вариант - изменить подход к сбору и анализу данных

**Пример.** Продолжительность пребывания больных в стационаре



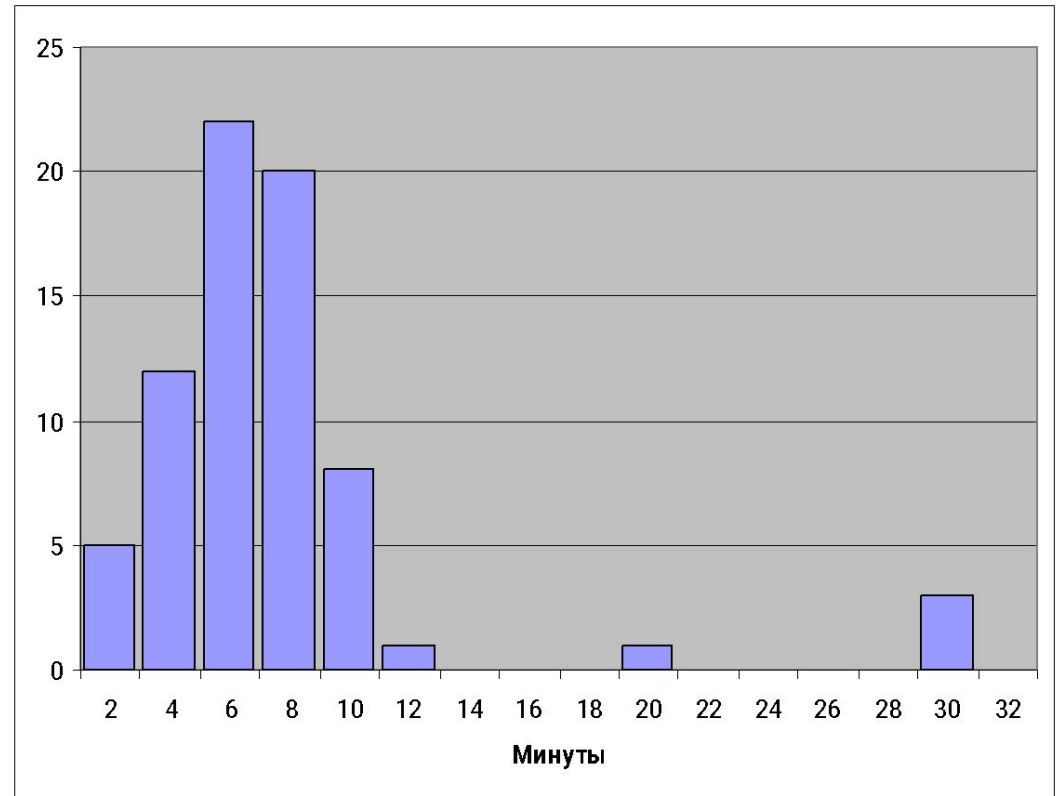
# Выбросы

---

Сильно отклоняющиеся значения называются **выбросами**.

Следует решить, отказаться от выбросов в дальнейшем исследовании или нет. Можно провести два исследования параллельно – с выбросами и без. Если выбросы исключаются, это должно быть аргументировано и детально описано в отчете.

**Пример.** Время обслуживания одного клиента

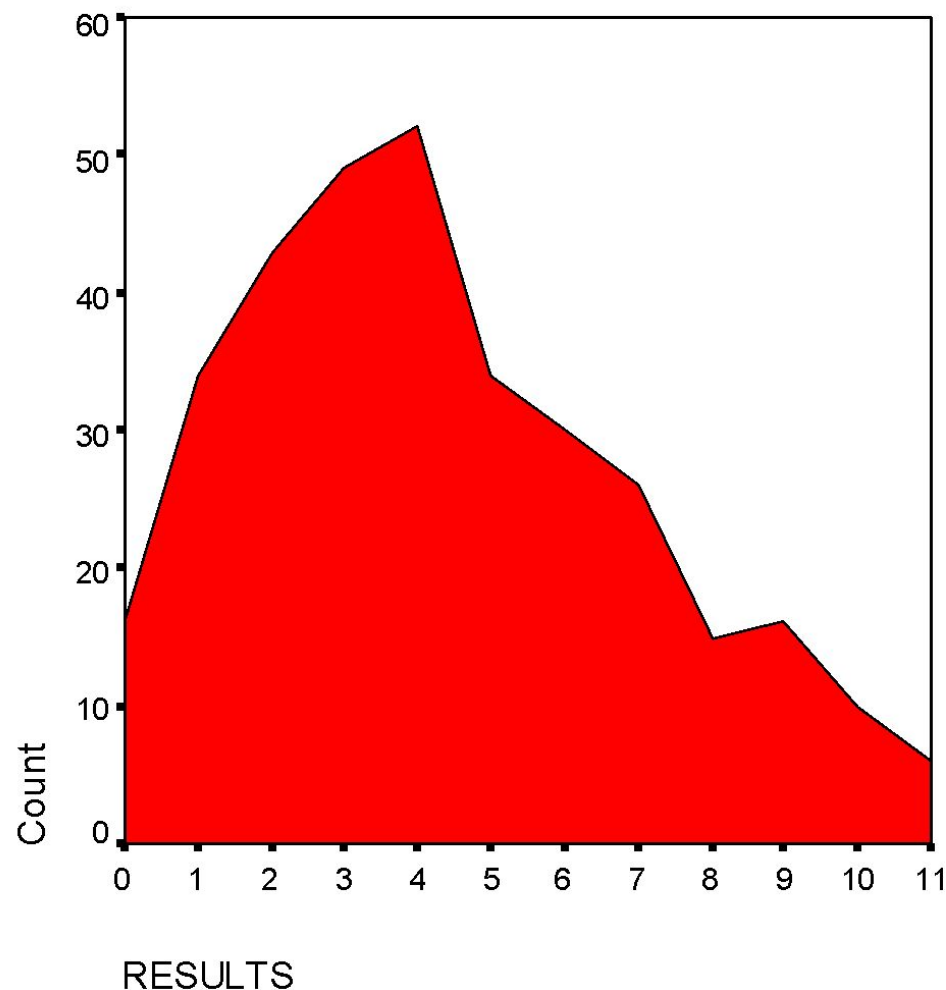


# Полигоны частот

---

**Полигон**, в отличие от гистограммы, строится в виде линии, проходящей по точкам, соответствующим серединам интервалов и частотам.

Зрительное представление о распределении частот, полученное при помощи полигона, довольно сильно отличается от гистограммы даже при одних и тех же данных.

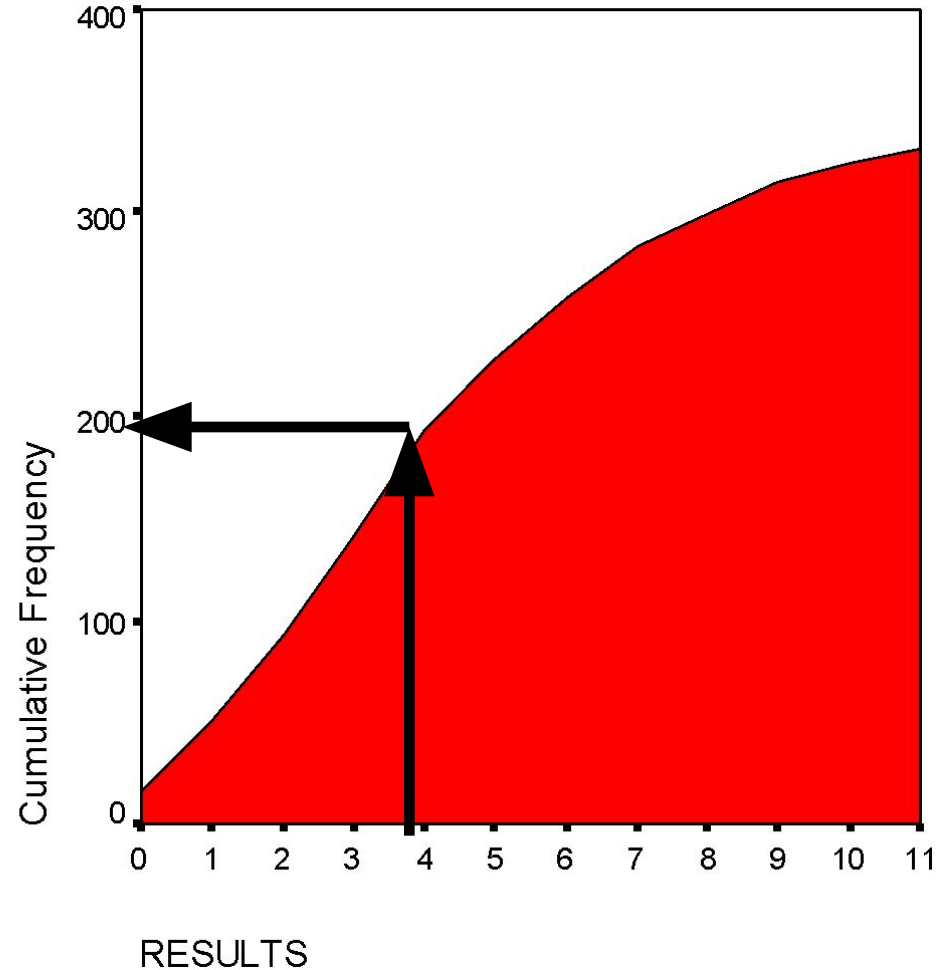


# Кумулята

Кумулята позволяет при помощи графика ответить на вопрос, сколько человек имеет 4 и менее решенных задач.

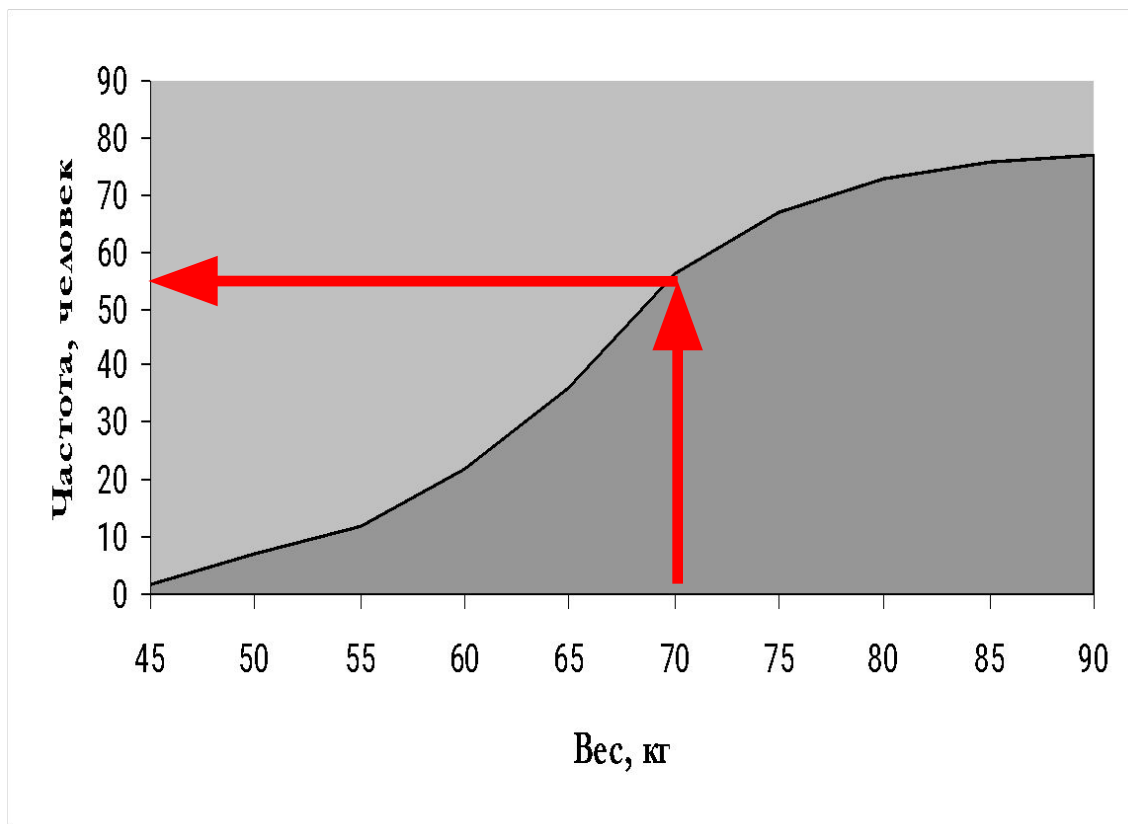
ВЕЗНН ТЗ

Ләйиқ	0	1	2	3	4	5	6	7	8	9	10	11	12
Сипшлрәтләр	0	34	43	42	25	34	30	29	12	11	10	9	331
Бәлсәтләр	0.000	0.103	0.130	0.148	0.121	0.103	0.081	0.112	0.042	0.048	0.030	0.018	1.000
Ләйиқ Бәлсәтләр	0.000	0.103	0.130	0.148	0.121	0.103	0.081	0.112	0.042	0.048	0.030	0.018	1.000



# Кумулята

Кумулята позволяет при помощи графика ответить на вопрос, сколько человек имеет вес до 70 кг.



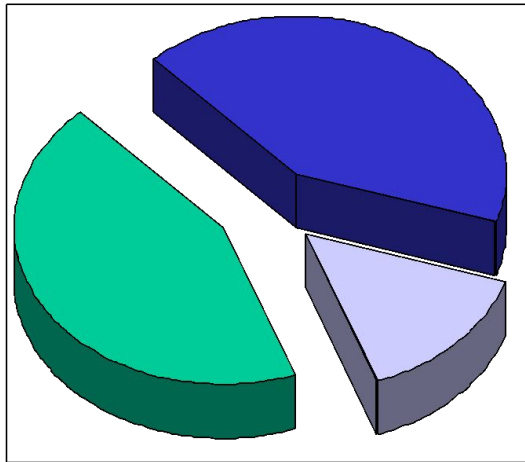
ИНТЕРВАЛЫ	$f$
45-49	2
50-54	5
55-59	5
60-64	10
65-69	14
70-74	20
75-79	11
80-84	6
85-89	3
90-94	1
Итого	77

# Круговая диаграмма

---

Круговая диаграмма полезна для представления частот переменной, измеряемой по номинальной шкале.

Объемная



Кольцевая

