

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

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

Порядковая шкала

Отношение к фильму

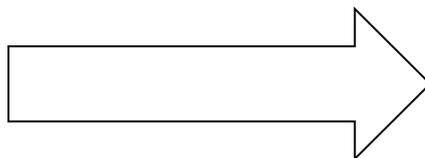
КАТЕГОРИИ	<i>f</i>
Очень понравился	24
Понравился	12
Фильм средний	10
Не понравился	6
Очень плохой	8
Всего	60

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	<i>f</i>	<i>P</i>	<i>CP</i>	%	<i>C%</i>
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

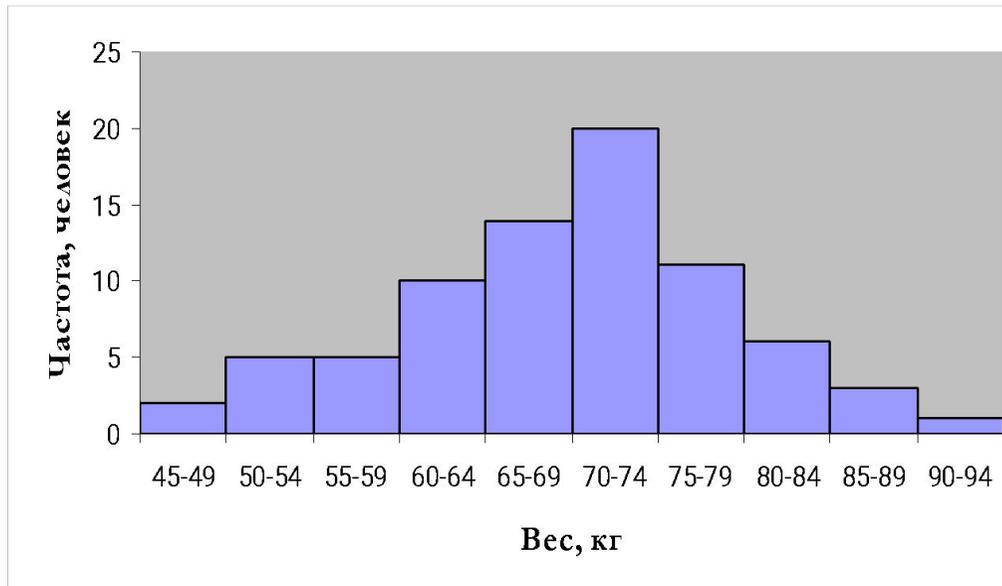
2.4. Визуальные представления

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?

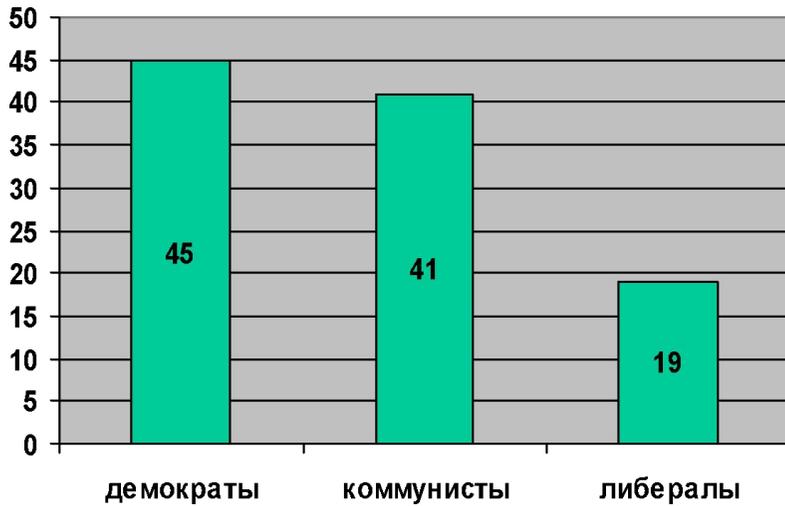
Гистограммы частот

Гистограмма частот – графическое представление, которое показывает распределение переменной.

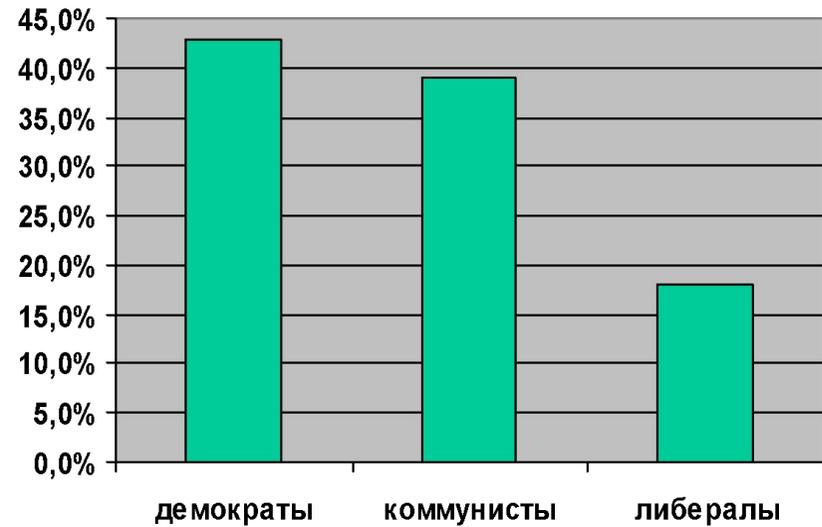


Гистограммы. Примеры (1)

Политические предпочтения



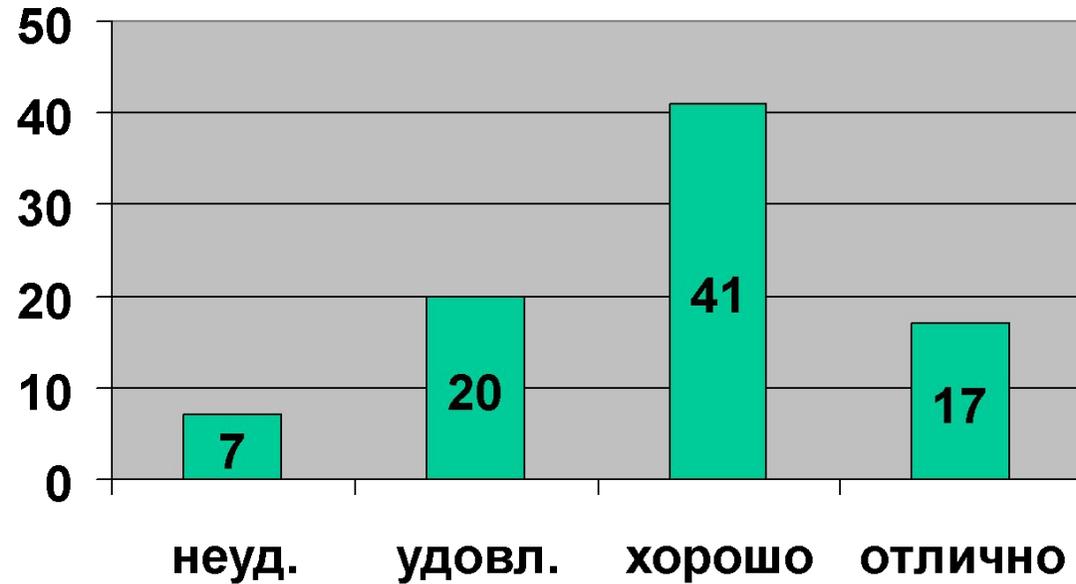
Абсолютная частота



Проценты

Гистограммы. Примеры (2)

Отметки по математике
(абсолютная частота)

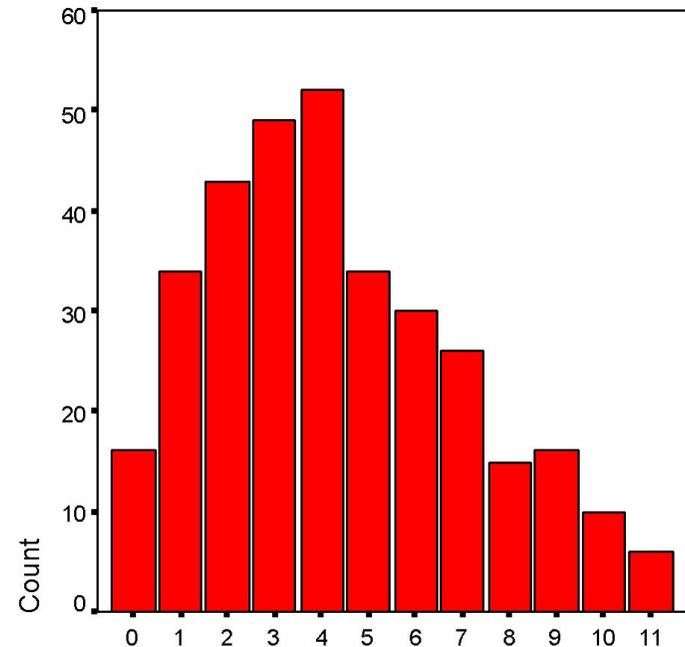


Пример. Тест первокурсников

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

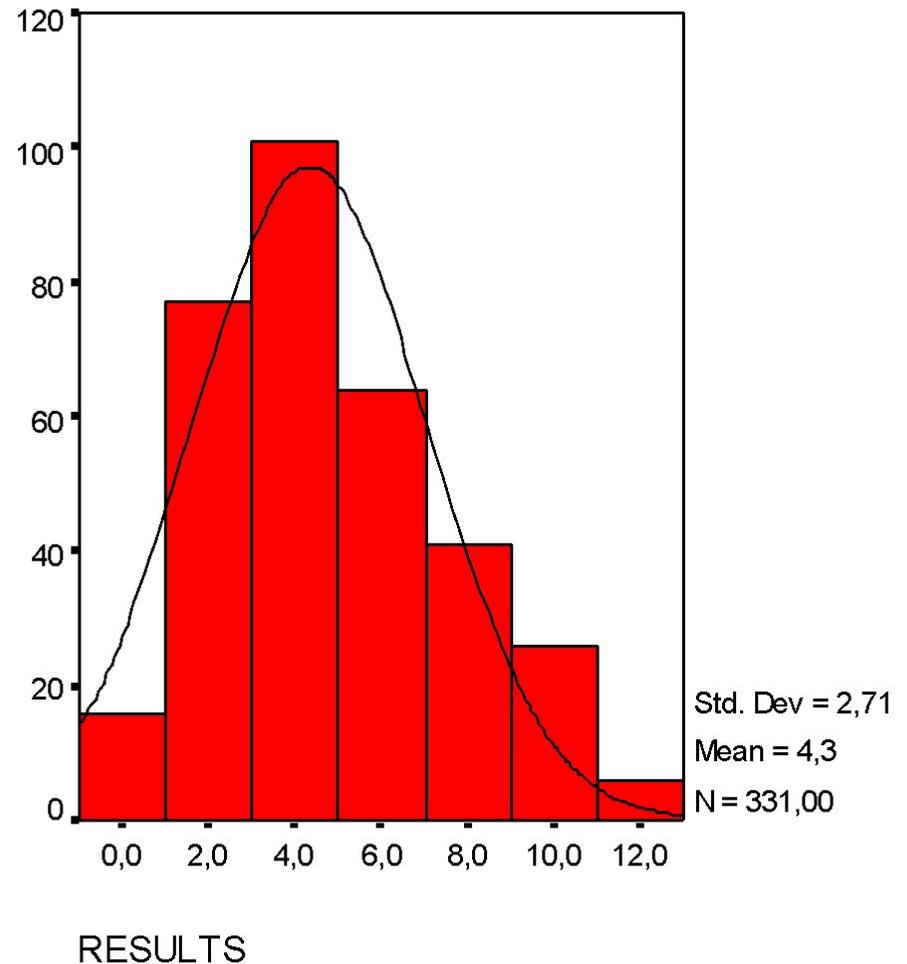


RESULTS

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

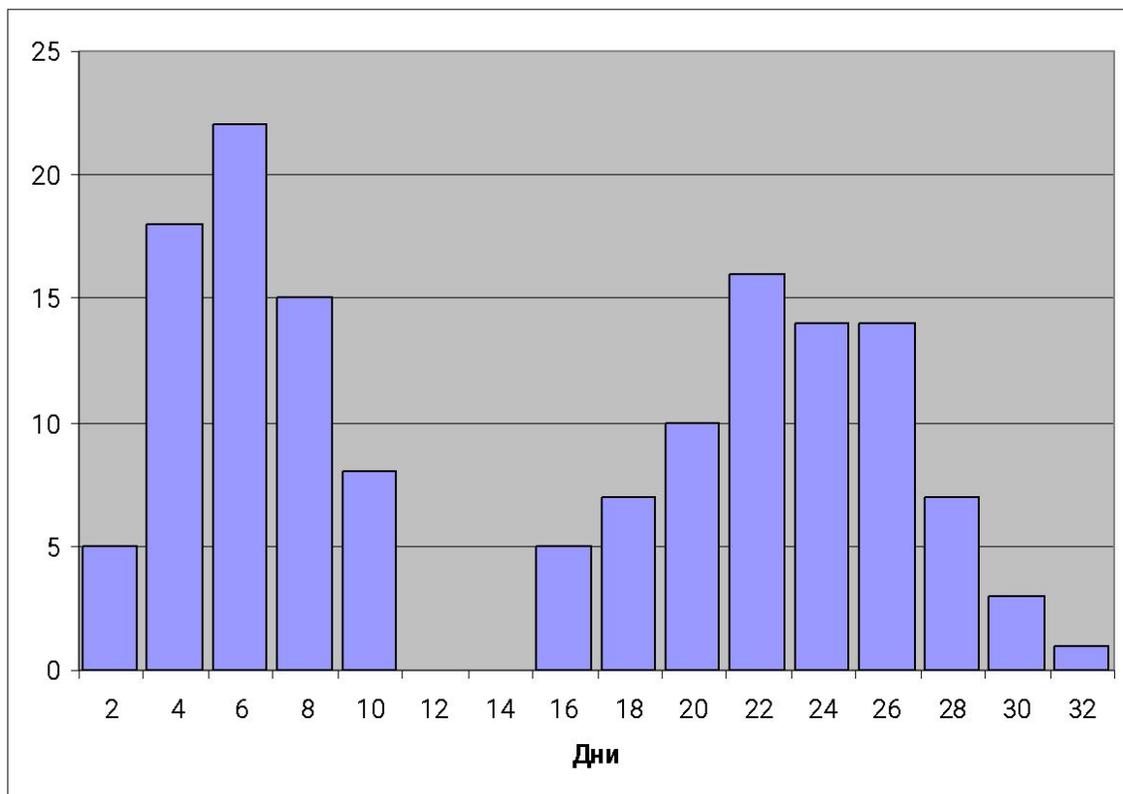
Если в гистограмме видны две отдельные группы данных, говорят о **бимодальном распределении данных**.

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

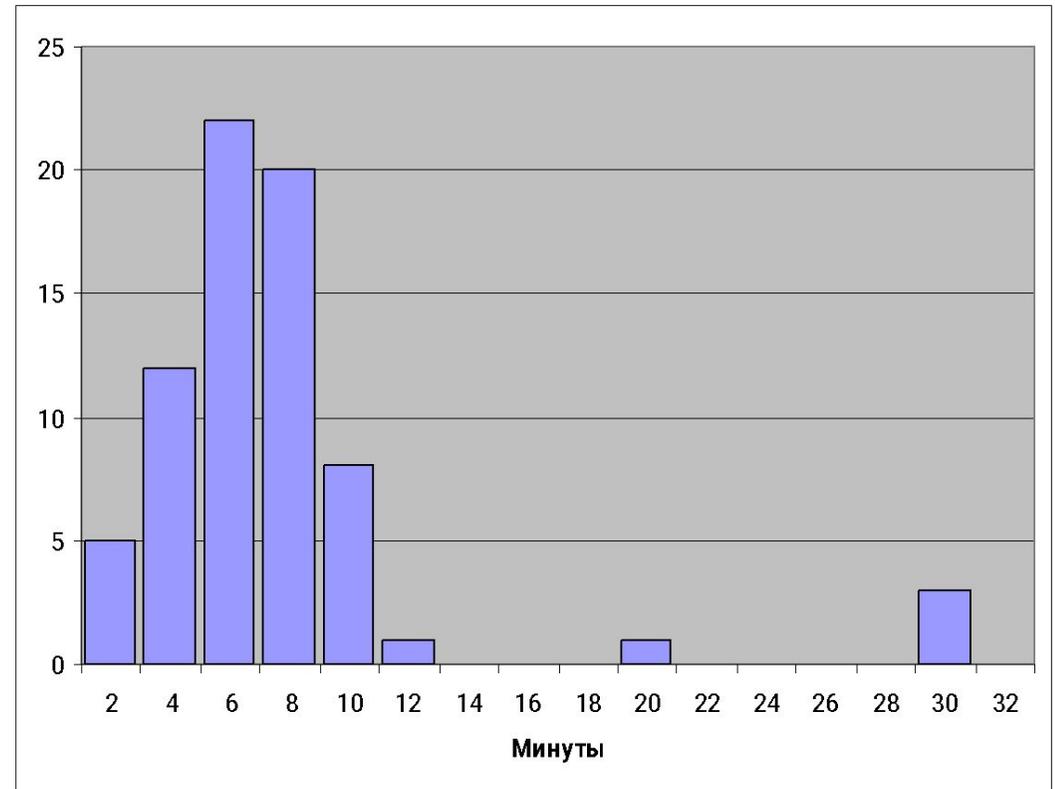


Выбросы

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.

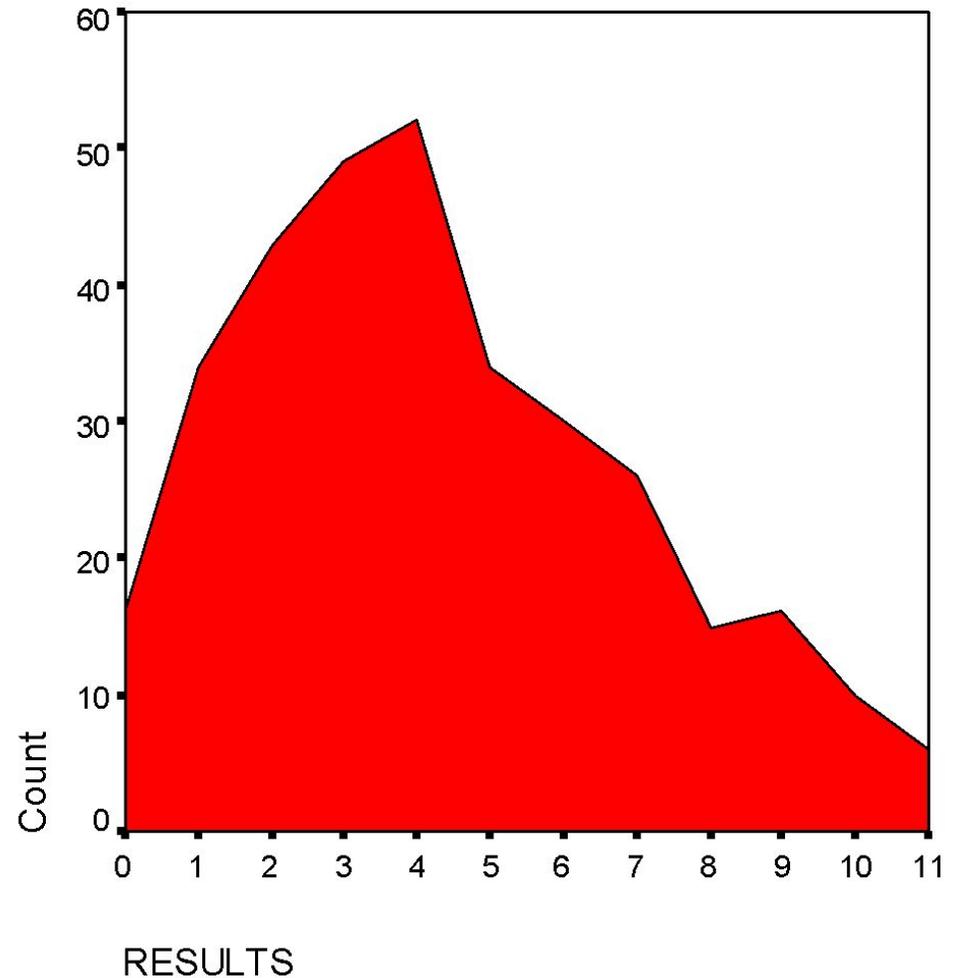
Пример. Время обслуживания одного клиента



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

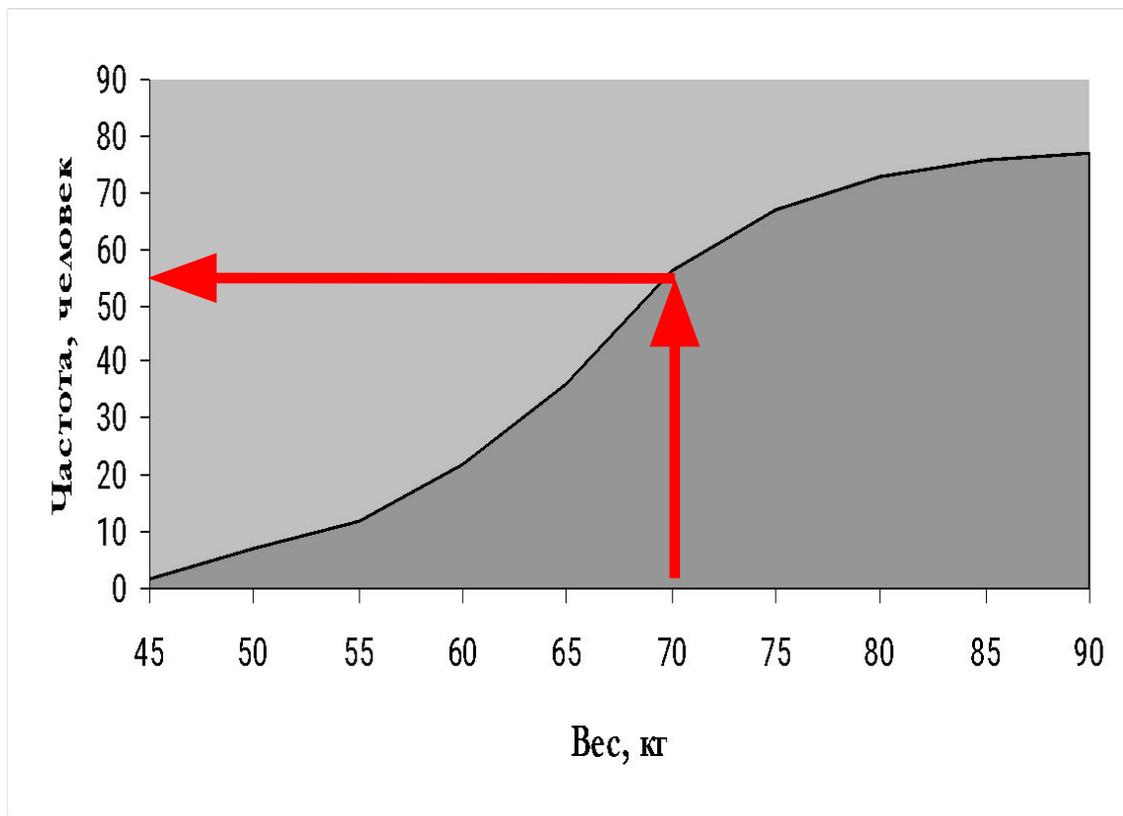
Кумулята позволяе при помощи графика ответить на вопрос, сколько человек имеет 4 и менее решенных задач.

ВЕЗНН 12

Ляйп	0	1	2	3	4	5	6	7	8	9	10	11	12
Бексеи	34	43	42	25	34	30	29	12	19	10	9	331	1000
Бексеи	103	130	148	121	103	81	72	42	48	30	18	1000	1000
Бексеи	121	158	176	151	103	81	62	42	48	30	18	1000	1000
Бексеи	176	219	234	176	103	81	62	42	48	30	18	1000	1000
Бексеи	219	262	277	201	103	81	62	42	48	30	18	1000	1000
Бексеи	262	305	317	226	103	81	62	42	48	30	18	1000	1000
Бексеи	305	348	360	251	103	81	62	42	48	30	18	1000	1000
Бексеи	348	391	403	276	103	81	62	42	48	30	18	1000	1000
Бексеи	391	434	446	301	103	81	62	42	48	30	18	1000	1000
Бексеи	434	477	489	326	103	81	62	42	48	30	18	1000	1000
Бексеи	477	520	532	351	103	81	62	42	48	30	18	1000	1000
Бексеи	520	563	575	376	103	81	62	42	48	30	18	1000	1000
Бексеи	563	606	618	401	103	81	62	42	48	30	18	1000	1000
Бексеи	606	649	661	426	103	81	62	42	48	30	18	1000	1000
Бексеи	649	692	704	451	103	81	62	42	48	30	18	1000	1000
Бексеи	692	735	747	476	103	81	62	42	48	30	18	1000	1000
Бексеи	735	778	790	501	103	81	62	42	48	30	18	1000	1000
Бексеи	778	821	833	526	103	81	62	42	48	30	18	1000	1000
Бексеи	821	864	876	551	103	81	62	42	48	30	18	1000	1000
Бексеи	864	907	919	576	103	81	62	42	48	30	18	1000	1000
Бексеи	907	950	962	601	103	81	62	42	48	30	18	1000	1000
Бексеи	950	993	1005	626	103	81	62	42	48	30	18	1000	1000
Бексеи	993	1036	1048	651	103	81	62	42	48	30	18	1000	1000
Бексеи	1036	1079	1091	676	103	81	62	42	48	30	18	1000	1000
Бексеи	1079	1122	1134	701	103	81	62	42	48	30	18	1000	1000
Бексеи	1122	1165	1177	726	103	81	62	42	48	30	18	1000	1000
Бексеи	1165	1208	1220	751	103	81	62	42	48	30	18	1000	1000
Бексеи	1208	1251	1263	776	103	81	62	42	48	30	18	1000	1000
Бексеи	1251	1294	1306	801	103	81	62	42	48	30	18	1000	1000
Бексеи	1294	1337	1349	826	103	81	62	42	48	30	18	1000	1000
Бексеи	1337	1380	1392	851	103	81	62	42	48	30	18	1000	1000
Бексеи	1380	1423	1435	876	103	81	62	42	48	30	18	1000	1000
Бексеи	1423	1466	1478	901	103	81	62	42	48	30	18	1000	1000
Бексеи	1466	1509	1521	926	103	81	62	42	48	30	18	1000	1000
Бексеи	1509	1552	1564	951	103	81	62	42	48	30	18	1000	1000
Бексеи	1552	1595	1607	976	103	81	62	42	48	30	18	1000	1000
Бексеи	1595	1638	1650	1001	103	81	62	42	48	30	18	1000	1000
Бексеи	1638	1681	1693	1026	103	81	62	42	48	30	18	1000	1000
Бексеи	1681	1724	1736	1051	103	81	62	42	48	30	18	1000	1000
Бексеи	1724	1767	1779	1076	103	81	62	42	48	30	18	1000	1000
Бексеи	1767	1810	1822	1101	103	81	62	42	48	30	18	1000	1000
Бексеи	1810	1853	1865	1126	103	81	62	42	48	30	18	1000	1000
Бексеи	1853	1896	1908	1151	103	81	62	42	48	30	18	1000	1000
Бексеи	1896	1939	1951	1176	103	81	62	42	48	30	18	1000	1000
Бексеи	1939	1982	1994	1201	103	81	62	42	48	30	18	1000	1000
Бексеи	1982	2025	2037	1226	103	81	62	42	48	30	18	1000	1000
Бексеи	2025	2068	2080	1251	103	81	62	42	48	30	18	1000	1000
Бексеи	2068	2111	2123	1276	103	81	62	42	48	30	18	1000	1000
Бексеи	2111	2154	2166	1301	103	81	62	42	48	30	18	1000	1000
Бексеи	2154	2197	2209	1326	103	81	62	42	48	30	18	1000	1000
Бексеи	2197	2240	2252	1351	103	81	62	42	48	30	18	1000	1000
Бексеи	2240	2283	2295	1376	103	81	62	42	48	30	18	1000	1000
Бексеи	2283	2326	2338	1401	103	81	62	42	48	30	18	1000	1000
Бексеи	2326	2369	2381	1426	103	81	62	42	48	30	18	1000	1000
Бексеи	2369	2412	2424	1451	103	81	62	42	48	30	18	1000	1000
Бексеи	2412	2455	2467	1476	103	81	62	42	48	30	18	1000	1000
Бексеи	2455	2498	2510	1501	103	81	62	42	48	30	18	1000	1000
Бексеи	2498	2541	2553	1526	103	81	62	42	48	30	18	1000	1000
Бексеи	2541	2584	2596	1551	103	81	62	42	48	30	18	1000	1000
Бексеи	2584	2627	2639	1576	103	81	62	42	48	30	18	1000	1000
Бексеи	2627	2670	2682	1601	103	81	62	42	48	30	18	1000	1000
Бексеи	2670	2713	2725	1626	103	81	62	42	48	30	18	1000	1000
Бексеи	2713	2756	2768	1651	103	81	62	42	48	30	18	1000	1000
Бексеи	2756	2799	2811	1676	103	81	62	42	48	30	18	1000	1000
Бексеи	2799	2842	2854	1701	103	81	62	42	48	30	18	1000	1000
Бексеи	2842	2885	2897	1726	103	81	62	42	48	30	18	1000	1000
Бексеи	2885	2928	2940	1751	103	81	62	42	48	30	18	1000	1000
Бексеи	2928	2971	2983	1776	103	81	62	42	48	30	18	1000	1000
Бексеи	2971	3014	3026	1801	103	81	62	42	48	30	18	1000	1000
Бексеи	3014	3057	3069	1826	103	81	62	42	48	30	18	1000	1000
Бексеи	3057	3100	3112	1851	103	81	62	42	48	30	18	1000	1000
Бексеи	3100	3143	3155	1876	103	81	62	42	48	30	18	1000	1000
Бексеи	3143	3186	3198	1901	103	81	62	42	48	30	18	1000	1000
Бексеи	3186	3229	3241	1926	103	81	62	42	48	30	18	1000	1000
Бексеи	3229	3272	3284	1951	103	81	62	42	48	30	18	1000	1000
Бексеи	3272	3315	3327	1976	103	81	62	42	48	30	18	1000	1000
Бексеи	3315	3358	3370	2001	103	81	62	42	48	30	18	1000	1000
Бексеи	3358	3401	3413	2026	103	81	62	42	48	30	18	1000	1000
Бексеи	3401	3444	3456	2051	103	81	62	42	48	30	18	1000	1000
Бексеи	3444	3487	3499	2076	103	81	62	42	48	30	18	1000	1000
Бексеи	3487	3530	3542	2101	103	81	62	42	48	30	18	1000	1000
Бексеи	3530	3573	3585	2126	103	81	62	42	48	30	18	1000	1000
Бексеи	3573	3616	3628	2151	103	81	62	42	48	30	18	1000	1000
Бексеи	3616	3659	3671	2176	103	81	62	42	48	30	18	1000	1000
Бексеи	3659	3702	3714	2201	103	81	62	42	48	30	18	1000	1000
Бексеи	3702	3745	3757	2226	103	81	62	42	48	30	18	1000	1000
Бексеи	3745	3788	3800	2251	103	81	62	42	48	30	18	1000	1000
Бексеи	3788	3831	3843	2276	103	81	62	42	48	30	18	1000	1000
Бексеи	3831	3874	3886	2301	103	81	62	42	48	30	18	1000	1000
Бексеи	3874	3917	3929	2326	103	81	62	42	48	30	18	1000	1000
Бексеи	3917	3960	3972	2351	103	81	62	42	48	30	18	1000	1000
Бексеи	3960	4003	4015	2376	103	81	62	42	48	30	18	1000	1000
Бексеи	4003	4046	4058	2401	103	81	62	42	48	30	18	1000	1000
Бексеи	4046	4089	4101	2426	103	81	62	42	48	30	18	1000	1000
Бексеи	4089	4132	4144	2451	103	81	62	42	48	30	18	1000	1000
Бексеи	4132	4175	4187	2476	103	81	62	42	48	30	18	1000	1000
Бексеи	4175	4218	4230	2501	103	81	62	42	48	30	18	1000	1000
Бексеи	4218	4261	4273	2526	103	81	62	42	48	30	18	1000	1000
Бексеи	4261	4304	4316	2551	103	81	62	42	48	30	18	1000	1000
Бексеи	4304	4347	4359	2576	103	81	62	42	48	30	18	1000	1000
Бексеи	4347	4390	4402	2601	103	81	62	42	48	30	18	1000	1000
Бексеи	4390	4433	4445	2626	103	81	62	42	48	30	18	1000	1000
Бексеи	4433	4476	4488	2651	103	81	62	42	48	30	18	1000	1000
Бексеи	4476	4519	4531	2676	103	81	62	42	48	30	18	1000	1000
Бексеи	4519	4562	4574	2701	103</								

Кумулята

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

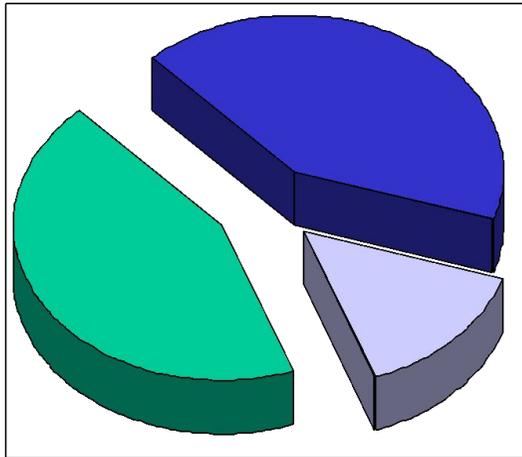


ИНТЕРВАЛЫ	f
45-49	2
50-54	5
55-59	5
60-64	10
65-69	14
70-74	20
75-79	11
80-84	6
85-89	3
90-94	1
Итого	77

Круговая диаграмма

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

Объемная



Кольцевая

