

Статистика

Статистика — раздел математики, изучающий методы обработки и анализа данных с целью принятия обоснованных решений.

Генеральная совокупность и выборка

- **Генеральная совокупность** – вся интересующая исследователя совокупность изучаемых объектов.
- **Выборка** – некоторая, обычно небольшая, часть генеральной совокупности, исследуемая с целью получения выводов о свойствах всей генеральной совокупности.

Генеральная совокупность и выборка

- ▣ **Репрезентативная** выборка *хорошо* представляет генеральную совокупность. Это означает, что каждое свойство (или комбинация свойств) наблюдается в выборке с той же частотой, что и в генеральной совокупности.
- ▣ **Параметры** — это количественные показатели, характеризующий всю генеральную совокупность.
- ▣ **Оценки параметров** - количественные показатели, вычисленные по выборке .

Генеральная совокупность и выборка

Что рассматривается?	Генеральная совокупность	Выборка
Как называется числовая характеристика?	Параметры	Оценки параметров
Примеры	Все жители города (150 тыс.) Доля мужчин в городе Средний возраст жителей	Участники опроса (250 чел.) Доля мужчин в выборке Средний возраст в выборке

ИССЛЕДОВАНИЕ ВЫБОРКИ

Пусть X – дискретный признак.

Сделана выборка

$$x_1, x_2, \dots, x_n$$

где

n - объем выборки.

Составляется ряд распределения

**Для каждого элемента выборки x_i
вычисляется его частота m_i .**

**Строится таблица абсолютных
частот:**

x_1	x_2	...	x_k
m_1	m_2	...	m_k

рассчитываем относительные частоты :

$$p_i = \frac{m_i}{n}$$

Причем $\sum_i m_i = n$ $\sum_i p_i = 1$

Строится таблица относительных частот:

x_1	x_2	...	x_k
p_1	p_2	...	p_k

**Пример: 30 посетителей гипермаркета
попросили ответить на вопрос: сколько раз в
неделю в среднем они посещают
гипермаркет. Получили дискретную
выборку:**

**1,3,4,4,5,6,3,5,7,1,4,0,3,7,5,6,2,1,4,2,6,4,5,7,1,4,2,
3,5,3**

Пример: 30 посетителей гипермаркета попросили ответить на вопрос: сколько раз в неделю в среднем они посещают гипермаркет. Получили дискретную выборку:

1,3,4,4,5,6,3,5,7,1,4,0,3,7,5,6,2,1,4,2,6,4,5,7,1,4,2,3,5,3

Таблица абсолютных частот

X	0	1	2	3	4	5	6	7
m	1	4	3	5	6	5	3	3

Пример: 30 посетителей гипермаркета попросили ответить на вопрос: сколько раз в неделю в среднем они посещают гипермаркет. Получили дискретную выборку:

1,3,4,4,5,6,3,5,7,1,4,0,3,7,5,6,2,1,4,2,6,4,5,7,1,4,2,3,5,3

X	0	1	2	3	4	5	6	7
m	1	4	3	5	6	5	3	3
p	$\frac{1}{30}$	$\frac{4}{30}$	$\frac{3}{30}$	$\frac{5}{30}$	$\frac{6}{30}$	$\frac{5}{30}$	$\frac{3}{30}$	$\frac{3}{30}$

По таблице распределения строят *полигон относительных частот* – ломаную линию на плоскости, соединяющую точки с координатами

$$(x_i, p_i)$$



Полигон частот, построенный по выборке является оценкой распределения случайной величины X – числа посещений гипермаркета рядовым покупателем.



Задание 1

100 человек, посмотревших новый фильм, попросили оценить его. Результаты оценки представлены в файле film.

	А	В
1	Не респондент	Оценка фильма
2		вариант 1
3	1	Очень понравился
4	2	Средний фильм
5	3	Понравился
6	4	Понравился
7	5	Средний фильм
8	6	Очень плохой
9	7	Понравился
10	8	Очень плохой
11	9	Понравился

Лабораторная 1. Задание 1

Составить таблицу частот для результатов оценки. Таблица должна иметь вид

КАТЕГОРИИ	<i>Частота</i>	<i>Относительная частота</i>
Очень понравился	24	24,00%
Понравился	28	28,00%
Фильм средний	32	32,00%
Не понравился	8	8,00%
Очень плохой	8	8,00%
Всего	100	100,00%

Лабораторная 1. Задание 1

Для вычисления абсолютных частот
можно использовать функцию
СЧЕТЕСЛИ

КАТЕГОРИИ	<i>Частота</i>	<i>Относительная частота</i>
Очень понравился	24	24,00%
Понравился	28	28,00%
Фильм средний	32	32,00%
Не понравился	8	8,00%
Очень плохой	8	8,00%
Всего	100	100,00%

Лабораторная 1. Задание 1

Для вычисления абсолютных частот
можно использовать функцию
СЧЕТЕСЛИ

	А	В
1	№ респон	Оценка фильма
2		вариант 1
3	1	Очень понравился
4	2	Средний фильм
5	3	Понравился
6	4	Понравился
7	5	Средний фильм
8	6	Очень плохой
9	7	Понравился
10	8	Очень плохой
11	9	Понравился



=счетесли(B3:B102;"понравился")

Лабораторная 1. Задание 1

Для вычисления абсолютных частот
можно использовать функцию
СЧЕТЕСЛИ

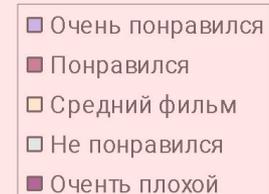
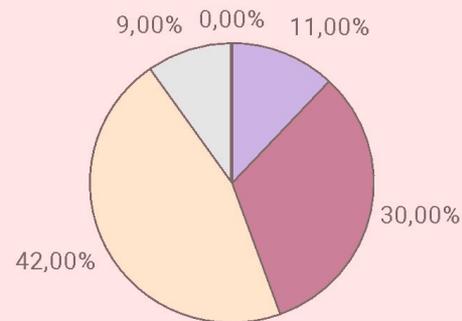
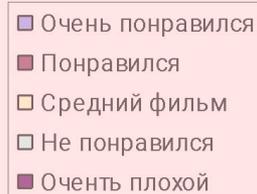
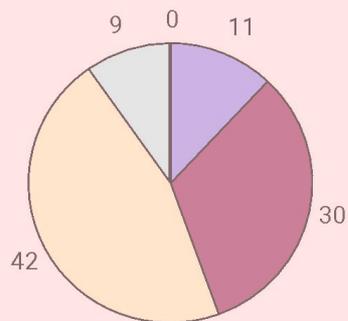
	А	В
1	№ респон	Оценка фильма
2		вариант 1
3	1	Очень понравился
4	2	Средний фильм
5	3	Понравился
6	4	Понравился
7	5	Средний фильм
8	6	Очень плохой
9	7	Понравился
10	8	Очень плохой
11	9	Понравился

Относительную частоту вычисляем разделив абсолютную частоту на объем выборки. Устанавливаем процентный формат ячеек

Представить частоты в виде гистограммы и круговой диаграммы.



Представить частоты в виде гистограммы и круговой диаграммы.



ИССЛЕДОВАНИЕ ВЫБОРКИ

Пусть X – непрерывный признак.

Сделана выборка x_1, x_2, \dots, x_n

составляется интервальный ряд
распределения

$$a = \min x_i$$

$$b = \max x_i$$

В Excel есть функции МИН и МАКС

Разбиваем весь диапазон $[a;b]$ на k равных по длине интервалов группирования. k дано в задании

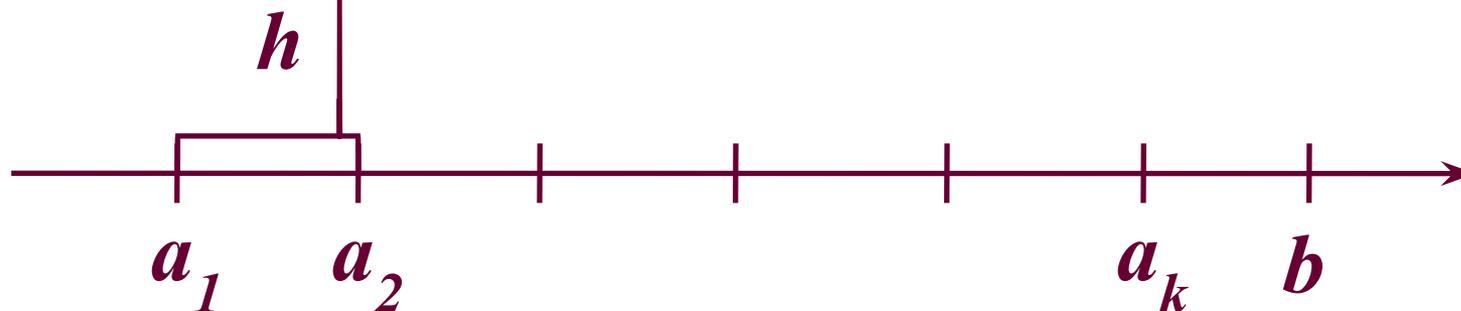
Для этого находим длину каждого интервала:

$$h = \frac{b - a}{k}$$



Устанавливается шкала интервалов

$$a_1 = a, \quad a_2 = a_1 + h, \quad a_3 = a_2 + h, \quad \dots \quad a_k + h = b$$



**Подсчитываются числа m_1, m_1, \dots, m_k
элементов выборки, попавших в каждый
интервал.**

**Находятся интервальные
относительные
частоты:**

$$p_i = \frac{m_i}{n}$$

Строится таблица:

$[a_1, a_2)$	$[a_2, a_3)$...	$[a_k, a_{k+1})$
p_1	p_2	...	p_k

Эта таблица и называется
ИНТЕРВАЛЬНЫМ РЯДОМ
РАСПРЕДЕЛЕНИЯ.

По ней можно построить гистограмму
частот.

Задание 2

В файле flat представлены данные о ценах на однокомнатные квартиры (тыс. USD), выставившихся на продажу в Москве.

	A	
1	Price	
2	54	
3	35	
4	59	
5	35	
6	33	
7	57	
8	43	
9	39	
10	70	
11	43	

Задание 2

В файле `flat` представлены данные о ценах на однокомнатные квартиры (тыс. USD), выставившихся на продажу в Москве.

1. Найдите минимальную и максимальную цену

Функции МИН и МАКС

Задание 2

В файле flat представлены данные о ценах на однокомнатные квартиры (тыс. USD), выставившихся на продажу в Москве.

2. Найдите ширину интервала при разбиении всего диапазона цен на 10 интервалов.

Минимум	27
Максимум	71
Ширина интервала	4,4

3. Вычислите левые и правые границы всех 10 интервалов.

Минимум	28,00	
Максимум	75,00	
h	4,70	
Левая граница	Правая граница	
28,00	=D9+\$E\$6	

3. Вычислите левые и правые границы всех 10 интервалов.

Минимум	28,00	
Максимум	75,00	
h	4,70	
Левая граница	Правая граница	
28,00	32,70	
=E9	32,70	

3. Вычислите левые и правые границы всех 10 интервалов.

Левая граница	Правая граница	
28,00	32,70	
32,70	37,40	
37,40	42,10	
42,10	46,80	
46,80	51,50	
51,50	56,20	
56,20	60,90	
60,90	65,60	
65,60	70,30	
70,30	75,00	

4. Найдите абсолютные частоты попадания в интервалы с помощью функции ЧАСТОТА.

Выделите ячейки, в которых будут храниться частоты и одну ячейку ниже

Левая граница	Правая граница	Абсолютная частота F
27,0	31,4	
31,4	35,8	
35,8	40,2	
40,2	44,6	
44,6	49,0	
49,0	53,4	
53,4	57,8	
57,8	62,2	
62,2	66,6	
66,6	71,0	

Левая граница	Правая граница	Абсолютная частота F
27,0	31,4	=ЧАСТОТА(
31,4	35,8	ЧАСТОТА(массив_данных; массив_интервалов)
35,8	40,2	
40,2	44,6	
44,6	49,0	
49,0	53,4	
53,4	57,8	
57,8	62,2	
62,2	66,6	
66,6	71,0	

Массив с правыми границами

Левая граница	Правая граница	Абсолютная частота F
27,0	31,4	=ЧАСТОТА(J2:J70;N23:N32)
31,4	35,8	ЧАСТОТА(массив_данных; массив_интервалов)
35,8	40,2	
40,2	44,6	
44,6	49,0	
49,0	53,4	
53,4	57,8	
57,8	62,2	
62,2	66,6	
66,6	71,0	

10R x 1C



*Завершаем формулу нажатием
Ctrl+Shift+Enter (одновременно)*

Левая граница	Правая граница	Абсолютная частота F
27,0	31,4	=ЧАСТОТА(J2:J70;N23:N32)
31,4	35,8	
35,8	40,2	
40,2	44,6	
44,6	49,0	
49,0	53,4	
53,4	57,8	
57,8	62,2	
62,2	66,6	
66,6	71,0	

Левая граница	Правая граница	Абсолютная частота F
27,0	31,4	9
31,4	35,8	18
35,8	40,2	17
40,2	44,6	12
44,6	49,0	5
49,0	53,4	3
53,4	57,8	0
57,8	62,2	1
62,2	66,6	3
66,6	71,0	1
		0

Цену от 27 тыс. USD до 31,4 тыс. USD имеют 9 квартир в выборке.

**5. Найдите относительные частоты.
Дайте экономическую интерпретацию
полученным значениям**

Левая граница	Правая граница	Абсолютная частота F	Относительная частота P
27,0	31,4	9	0,13
31,4	35,8	18	0,26
35,8	40,2	17	0,25
40,2	44,6	12	0,17
44,6	49,0	5	0,07
49,0	53,4	3	0,04
53,4	57,8	0	0,00
57,8	62,2	1	0,01
62,2	66,6	3	0,04
66,6	71,0	1	0,01
		0	
	Сумма	69	1

**Цену от 27 тыс. USD до 31,4 тыс. USD
Имеют 13% квартир в выборке.**

6. Найдите накопленные относительные частоты. Дайте экономическую интерпретацию полученным значениям (достаточно интерпретировать одну строку, только не первую)

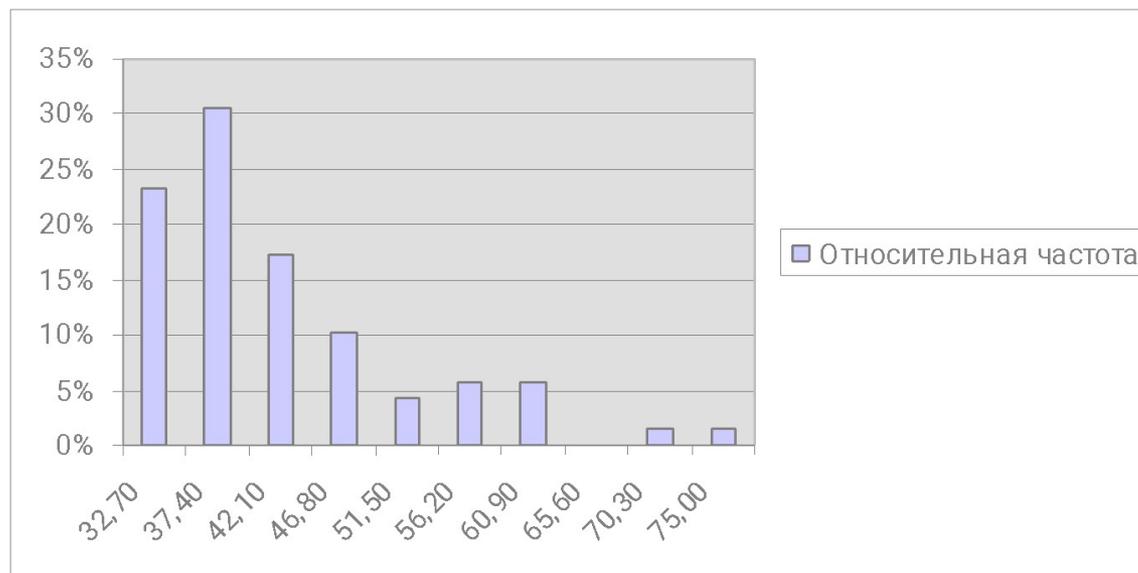
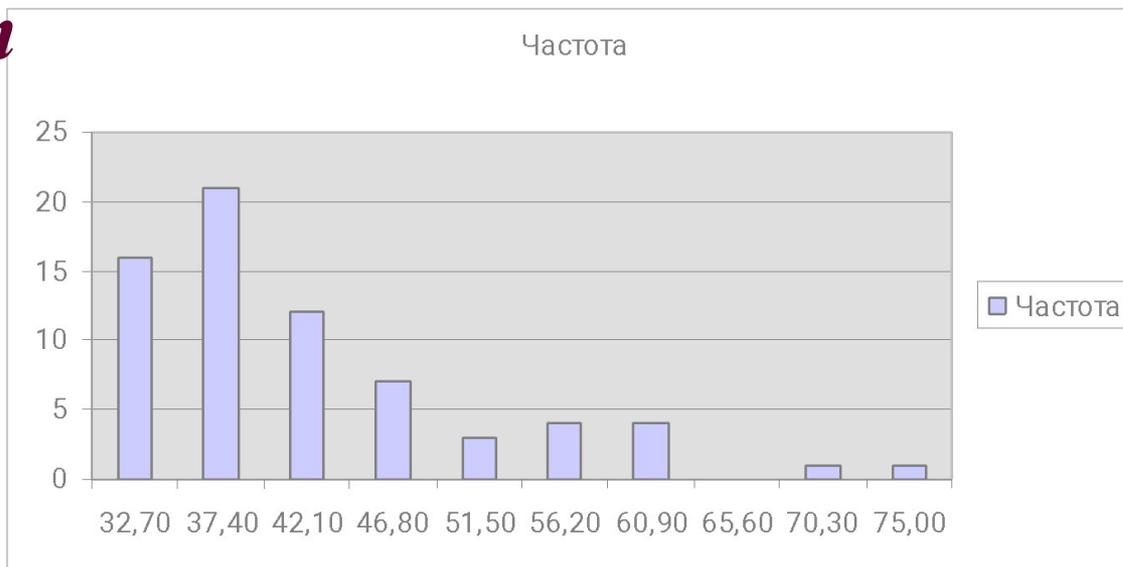
Левая граница	Правая граница	Абсолютная частота F	Относительная частота P	Накопленная относительная частота
27,0	31,4	9	0,13	0,13
31,4	35,8	18	0,26	0,39
35,8	40,2	17	0,25	0,64
40,2	44,6	12	0,17	0,81
44,6	49,0	5	0,07	0,88
49,0	53,4	3	0,04	0,93
53,4	57,8	0	0,00	0,93
57,8	62,2	1	0,01	0,94
62,2	66,6	3	0,04	0,99
66,6	71,0	1	0,01	1,00
		0		
	Сумма	69	1	

64% квартир в выборке имеет цену от 27 тыс. USD до 40,2 тыс. USD

Левая граница	Правая граница	Абсолютная частота F	Относительная частота P	Накопленная относительная частота
27,0	31,4	9	0,13	0,13
31,4	35,8	18	0,26	0,39
35,8	40,2	17	0,25	0,64
40,2	44,6	12	0,17	0,81
44,6	49,0	5	0,07	0,88
49,0	53,4	3	0,04	0,93
53,4	57,8	0	0,00	0,93
57,8	62,2	1	0,01	0,94
62,2	66,6	3	0,04	0,99
66,6	71,0	1	0,01	1,00
		0		
	Сумма	69	1	

7. Постройте гистограммы абсолютных и относительных частот. Для построения гистограммы воспользуйтесь мастером диаграмм. На корешке вкладки Стандартные выберите в списке диаграмм пункт Гистограмма - обычная. В качестве Диапазона данных задайте массив с частотами. В качестве подписей по оси X укажите массив с правыми границами.

Гистограммы абсолютных и относительных частот



8. Постройте полигон относительных частот. Чтобы построить полигон, воспользуйтесь мастером диаграмм. На корешке вкладки Стандартные выберите в списке диаграмм пункт График с маркерами. В качестве Диапазона данных задайте массив с относительными частотами. В качестве подписей по оси X укажите массив с правыми границами.



Полигон относительных частот

9. Постройте кумуляту (график накопленных относительных частот). По оси X должны быть правые границы, по оси Y накопленные относительные частоты.

