

Лабораторная работа № 7

ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Цель работы: приобрести навыки построения таблиц частот, гистограмм, полигона, графика эмпирической функции распределения вероятностей.

Основные теоретические сведения

Статистический ряд – это варианты, расположенные в порядке возрастания их значений с соответствующими им весами.

Статистический ряд называется дискретным, если он представляет собой выборку значений дискретной случайной величины.

Ряд называется непрерывным (интервальным), если он представляет выборку непрерывной случайной величины.

Частотой варианты x_i называется число (n_i) , показывающее сколько раз эта варианта встречается в выборке.

Частостью или относительной частотой или долей варианты называется число $w_i = n_i / n$.

Частоты и частости называют весами. Пусть x – некоторое число. Тогда количество вариант x_i , значения которых меньше x , называется накопленной частотой (n_x)

$$n_x = \sum_{x_i < x} n_i \quad \text{или} \quad n_x = \sum_{j=1}^i n_j, \quad (i = \overline{1, k}).$$

Отношение накопленной частоты к общему числу наблюдений n называется накопленной относительной частотой.

$$\frac{n_x}{n} = \sum_{j=1}^i n_j / n.$$

На некоторых этапах статистического анализа необходимо исходную выборку представить в группированном виде (группированный статистический ряд), что позволит получить общее представление о законе распределения случайной величины X .

Последовательность процедуры группировки неупорядоченной выборки из генеральной совокупности состоит из следующих этапов:

1. Формирование вариационного ряда.
2. Выделение \max и \min элементов выборки ($x_{\min} = x_1, x_{\max} = x_n$) для вычисления размаха:

$$R = x_{\max} - x_{\min}.$$

3. Нахождение числа интервалов (k) группировки осуществляется из соображений точности и устанавливается либо эмпирическим путем в зависимости от объема выборки, либо определяется природой явления или условия проведения эксперимента, либо по формуле Стерджерса

$$k = 1 + 1,4 \cdot \ln(n).$$

Округление при нахождении k осуществляется до ближайшего целого числа.

4. Определение ширины интервалов группирования (при равноточном группировании)

$$h = R / k = (x_{\max} - x_{\min}) / k.$$

Однако длины разрядов не обязательно брать равными друг другу, бывают ситуации, когда это даже неудобно.

Границы разрядов удобно принимать «круглыми» числами.

Если при вычислении h необходимо округлить результат, следует помнить, что последний интервал группирования будет меньше ширины h при округлении в большую сторону и больше h при округлении в меньшую сторону.

5. Формирование последовательности границ интервалов разбиения.

Полученный ряд будет иметь следующий вид:

$$x_1, (x_1 + h), x_1 + 2h, \dots, x_1 + (k - 1) \cdot h, x_n.$$

6. Определение количества элементов выборки (частоты), попавших в каждый интервал.

В итоге интервальный ряд можно представить таблицей, которая называется группированным статистическим рядом, в верхней строке таблицы указаны границы интервалов или середины интервалов группировки, а в нижней – соответствующие им частоты (n_i) или относительные частоты.

Для наглядности представления выборки используют графические изображения вариационных и статистических рядов в виде полигонов, гистограммы и кумуляты. Полигон частот, как правило, служит для изображения дискретного вариационного ряда и представляет собой ломаную, соединяющую точки плоскости с координатами (x_i, n_i) , $i = \overline{1, k}$.

Для интервального ряда также строится полигон, только его ломаная проходит через точки $(x_i + x_{i+1}) / 2, n_i$, $i = \overline{1, k}$.

Полигон относительных частот получается из полигона частот сжатием по оси «ОУ» в n раз.

Гистограмма служит только для представления интервальных вариационных рядов и имеет вид ступенчатой фигуры из прямоугольников с основаниями, равными длине интервалов ($i = \overline{1, n}$). Площадь ступенчатой фигуры в этом случае равна объему выборки (n).

Аналогично строится гистограмма относительных частот. Площадь соответствующей ступенчатой фигуры равна единице.

При увеличении объема выборки и уменьшении интервала группировки гистограмма относительных частот является статистическим аналогом плотности распределения $f(x)$ генеральной совокупности, что соответствует условию
$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Если на гистограмме соединить середины верхних сторон прямоугольников, то полученная ломаная образует полигон относительных частот. Гистограмму можно рассматривать как график эмпирической (выборочной) плотности распределения $f(x)$. Если у теоретического распределения существует конечная плотность, то эмпирическая плотность является некоторым приближением для теоретической.

При анализе характера распределения иногда полезна стратификация данных. Если одно и то же изделие изготавливается разными рабочими, часто имеет смысл проанализировать работу каждого из них отдельно: провести стратификацию, или расслоение, по квалификации рабочих. При использовании материала из разных партий иногда уточнить природу дефекта можно, если анализировать эти партии отдельно.

В производстве для стратификации удобен метод, называемый 5М (по первым буквам английских наименований): необходимо провести стратификацию данных по квалификации работников (*Men*), используемому оборудованию (*Machine*), по материалам (*Material*), технологии изготовления (*Method*), методам и средствам измерения (*Measure*).

Задания на выполнение лабораторной работы

1. Изучите алгоритм работы с функциями MS Excel.
2. По данным выборки постройте интервальный статистический ряд
3. Постройте графики: накопительных частот— кумуляту, полигон и гистограмму
4. Ответьте на контрольные вопросы.

Методика выполнения работы

1. Введите в один столбец MS Excel результаты измерений, выполненных на двух станках А и Б из табл. 7.1 для своего варианта.
2. Найдите максимальное и минимальное значения, используя встроенные статистические функции МАКС и МИН (рис.7.1).
3. Вычислите размах выборки (R).
4. Определите количество интервалов (k) и найдите их ширину (h).

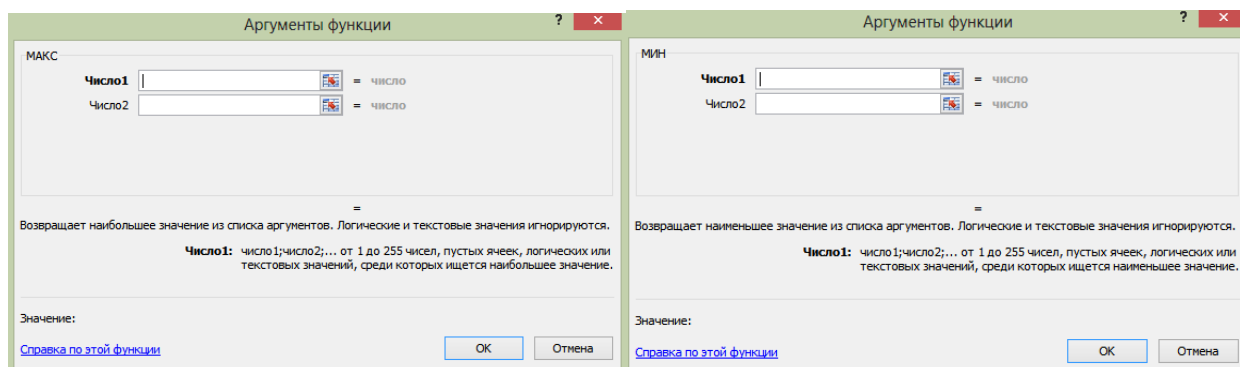


Рис. 7.1 . Встроенные функция массива МАКС и МИН

5. Подготовьте массив классов: в качестве первого значения введите найденное минимальное значение, последующие значения – с шагом, равным ширине интервала.
6. Для подсчета частот используйте функцию массива ЧАСТОТА из списка Статистических функций (рис.7.2), которая возвращает частотное распределение в виде массива из одного столбца.

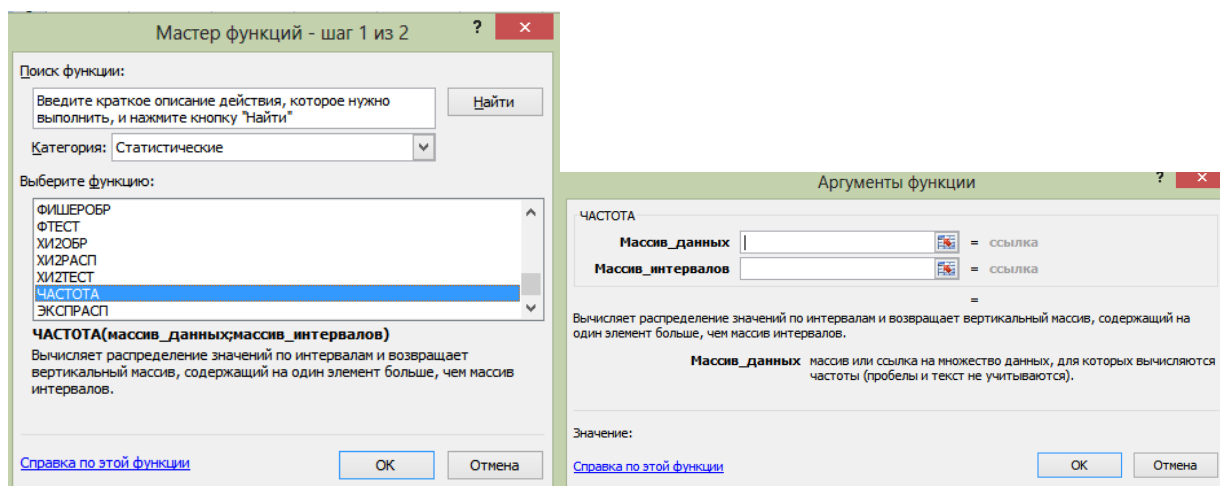


Рис. 7.2. Встроенная функция «Частота»

Функция служит для подсчета количества значений в интервале. Поскольку данная функция возвращает массив, она вводится как формула массива.

Синтаксис функции:

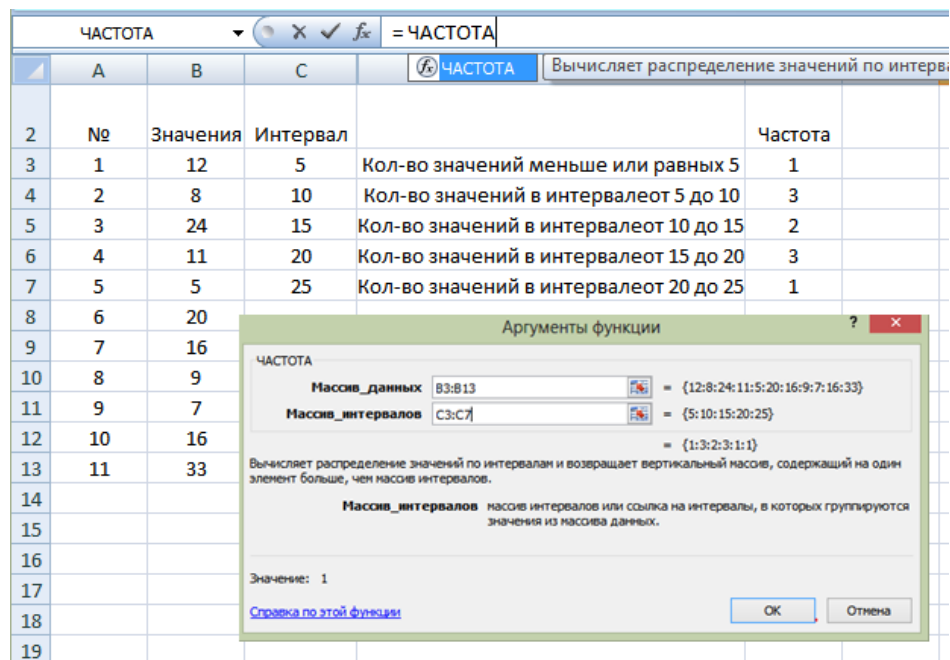
= ЧАСТОТА (массив данных, массив интервалов).

Массив данных – обязательный аргумент. Массив или ссылка на множество значений, для которых вычисляются частоты. Если аргумент «массив данных» не содержит значений, функция ЧАСТОТА возвращает массив нулей.

Массив интервалов – обязательный аргумент. Массив или ссылка на множество интервалов, в которые группируются значения аргумента «массив данных». Если аргумент «массив интервалов» не содержит значений, функция ЧАСТОТА возвращает количество элементов в аргументе «массив данных».

7. Вычислите частоты, используя функцию массива ЧАСТОТА. Для этого выделите ячейки, в которых требуется рассчитать частоты. Выделите знак равно или знак $f(x)$ и выберите функцию ЧАСТОТА. Введите необходимые диапазоны и нажмите сочетание клавиш CTRL + SHIFT + ENTER. Если нажать просто «ОК», то формула не будет введена как формула массива и отобразится только один результат в самой первой ячейке (рис.7.3).

8. С помощью мастера диаграмм постройте гистограмму частот. Для построения гистограммы в MS Excel введите в таблицу результаты наблюдений и подготовьте столбец рассчитанных значений границ интервалов (рис.7.3).



	A	B	C		Частота
2	№	Значения	Интервал		
3	1	12	5	Кол-во значений меньше или равных 5	1
4	2	8	10	Кол-во значений в интервале от 5 до 10	3
5	3	24	15	Кол-во значений в интервале от 10 до 15	2
6	4	11	20	Кол-во значений в интервале от 15 до 20	3
7	5	5	25	Кол-во значений в интервале от 20 до 25	1
8	6	20			
9	7	16			
10	8	9			
11	9	7			
12	10	16			
13	11	33			
14					
15					
16					
17					
18					
19					

Аргументы функции

ЧАСТОТА

Массив_данных: B3:B13 = {12;8;24;11;5;20;16;9;7;16;33}

Массив_интервалов: C3:C7 = {5;10;15;20;25}

= {1;3;2;3;1;1}

Вычисляет распределение значений по интервалам и возвращает вертикальный массив, содержащий на один элемент больше, чем массив интервалов.

Массив_интервалов: массив интервалов или ссылка на интервалы, в которых группируются значения из массива данных.

Значение: 1

[Справка по этой функции](#)

OK Отмена

Рис. 7.3. Встроенная функция массива ЧАСТОТА

9. Стратифицируйте гистограмму по станкам: проведите расчеты по п. 2–8 для станка А и для станка Б, а также отдельно для каждого и представьте результаты (рис.7.4).

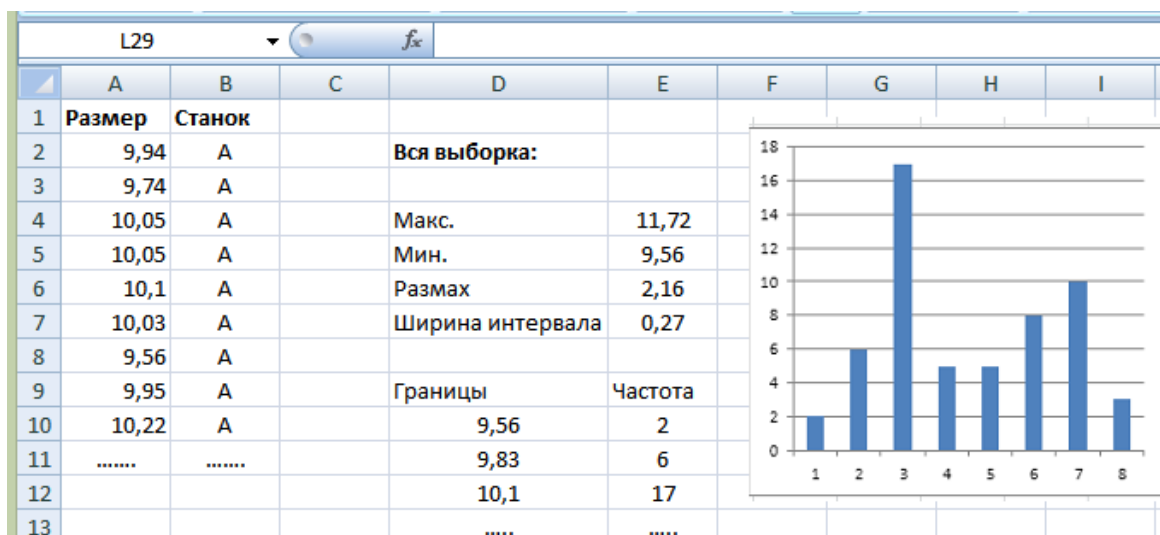


Рис. 7.4. Пример фрагмента представления результатов

10. Для данных по станку А постройте полигон и гистограмму частот в одной системе координат (пример на рис. 7.5.а), для станка Б – кумуляту (кумулятивную кривую – график накопленных частот) и график. Для построения графика накопленных частот (пример на рис. 7.5.б) предварительно составьте расчетную таблицу с полями: варианты (x_i), относительные частоты ($w_i = n_i / n$), накопительные относительные частоты ($W_i = W_{i-1} + w_i$).

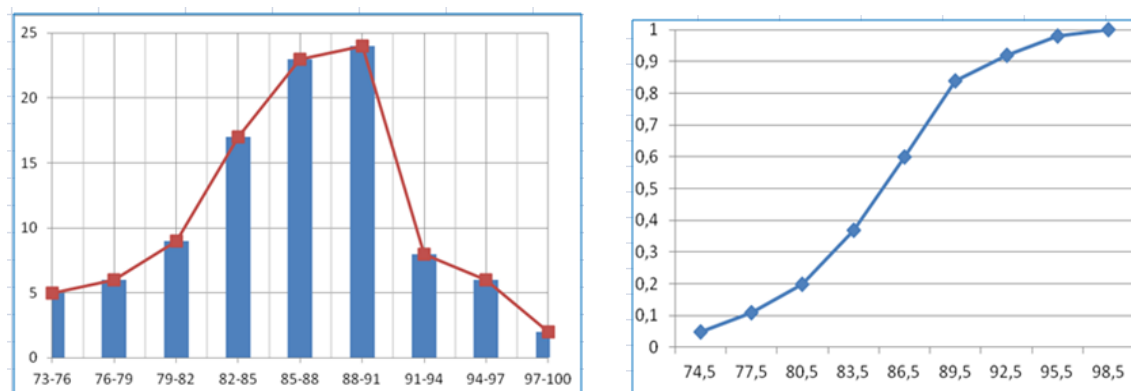


Рис. 7. а) полигон и гистограмма б) кумулятивная кривая

11. Сформулируйте выводы по результатам рассмотрения всех графических методов представления статистических данных.

Задания для индивидуального выполнения

Результаты измерений, выполненных на двух станках А и Б (к значениям А и В прибавьте N – свой номер в списке группы) представлены в табл. 7.1

Таблица 7.1

№	1	2	3	4	5	6	7	8	9	10	11
Значение	9,94	9,74	10,05	10,12	10,1	10,1	9,56	9,95	10,22	9,78	9,86
Станок	А	А	А	А	А	А	А	А	А	А	А
№	12	13	14	15	16	17	18	19	20	21	22
Значение	9,66	9,63	9,8	9,85	9,58	9,89	9,92	10,03	9,93	9,93	9,93
Станок	А	А	А	А	А	А	А	А	А	А	А
№	23	24	25	26	27	28	29	30	31	32	33
Значение	10,21	9,98	9,96	9,9	10,22	10,17	10,48	9,87	11,04	10,89	11,72
Станок	А	А	А	А	А	А	А	А	Б	Б	Б
№	34	35	36	37	38	39	40	41	42	43	44
Значение	10,76	11,35	11	11	10,63	10,88	10,92	11,11	10,76	11,01	10,84
Станок	Б	Б	Б	Б	Б	Б	Б	Б	Б	Б	Б
№	45	46	47	48	49	50	51	52	53	54	55
Значение	10,84	10,93	11,13	10,5	11,34	10,51	10,42	10,67	10,98	10,65	11,07
Станок	А	А	А	А	А	А	Б	Б	Б	Б	Б

Требования к оформлению отчета

Отчет должен содержать следующие составляющие:

1. Тема лабораторной работы.
2. Цель работы.
3. Основные формулы, методика выполнения и алгоритм расчетов.
4. Полученные результаты в виде таблиц, графиков.
5. Анализ полученных результатов.
6. Вывод по работе.

Контрольные вопросы

1. Сформулируйте алгоритм построения эмпирической функции распределения.
2. Какие статистики относятся к показателям формы?
3. Назовите показатели, описывающие закон распределения.
4. Как формируется таблица частот?
5. Что показывает график накопленных относительных частот?
6. Чему равна площадь гистограммы относительных частот?
7. Поясните, в каких случаях строится полигон частот и что он характеризует.