

МИНОБРНАУКИ РОССИИ  
Федеральное государственное автономное образовательное  
учреждение высшего образования  
«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Филиал ЮФУ в г. Новошахтинске

**С. Ю. АВЕРЬЯНОВА, Н. В. РАСТЕРЯЕВ**

**ЛАБОРАТОРНЫЙ ПРАКТИКУМ  
ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ  
В СРЕДЕ ЭТ MS EXCEL**

Ростов-на-Дону - 2014

УДК [311:004.9](075.8)

ББК32.973.26-018.2я73

А 19

*Печатается по решению  
кафедры информатики и математики филиала ЮФУ в г. Новошахтинске  
протокол №1 от 01.09.2014 г.;  
учебно-методической комиссии филиала ЮФУ в г. Новошахтинске протокол  
№ 3 от 02.12.2014 г.*

Рецензенты:

доктор технических наук, профессор Галушкин Н.Е.;

кандидат технических наук, доцент Байдюк А.П.

А 19 Аверьянова С.Ю., Растеряев Н.В.

Лабораторный практикум по математической статистике в среде ЭТ MS Excel: учебное пособие; Южный федеральный университет. – Ростов-на-Дону: Издательство Южного федерального университета, 2014. – 64 с.

ISBN 978-5-9275-1459-5

Учебное пособие предназначено для проведения лабораторных работ, а также организации управляемой самостоятельной работы студентов.

Пособие содержит краткие основные теоретические положения и примеры решения типовых задач статистической обработки выборки. Рассмотрены вопросы графического представления выборки, вычисления ее числовых характеристик и проверки близости эмпирической и теоретической функций распределения. Задачи решаются в среде электронных таблиц MS Excel. Представлены варианты заданий для самостоятельной работы студентов.

Пособие предназначено для студентов, обучающихся по направлениям подготовки «Экономика», «Менеджмент», «Бизнес-информатика» и преподавателей высших учебных заведений.

Публикуется в авторской редакции

ISBN 978-5-9275-1459-5

УДК [311:004.9](075.8)

ББК32.973.26-018.2я73

© Южный федеральный университет, 2014

## СОДЕРЖАНИЕ

Предисловие	4
Лабораторная работа №1	5
ВАРИАЦИОННЫЕ РЯДЫ. ВЫБОРОЧНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ	
Лабораторная работа №2	25
ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ВЫБОРКИ	
Лабораторная работа №3	36
ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ВЫБОРКИ СТАНДАРТНЫМИ СРЕДСТВАМИ ЭТ MS EXCEL	
Лабораторная работа №4	41
РАСЧЕТ ЧИСЛОВЫХ ХАРАКТЕРИСТИК ВЫБОРКИ С ПОМОЩЬЮ ВСТРОЕННЫХ ФУНКЦИЙ ЭТ MSEXCEL	
Лабораторная работа №5	56
НАХОЖДЕНИЕ ЧИСЛОВЫХ ХАРАКТЕРИСТИК ВЫБОРКИ СТАНДАРТНЫМИ СРЕДСТВАМИ ЭТ MS EXCEL	
Список использованных источников	64

## Предисловие

Курс «Теории вероятностей и математической статистики» является одним из важнейших математических курсов для экономических специальностей. Весь комплекс социально-экономических наук построен и развивается на вероятностно-статистической базе, и без соответствующей подготовки невозможно полноценное изучение этих дисциплин. Математическая статистика использует математический аппарат и выводы теории вероятностей, изучает математические методы систематизации, обработки и использования статистических данных для научных и практических выводов. Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента), в ходе исследования (последовательный анализ) и решает многие другие задачи. Современную математическую статистику определяют как науку о принятии решений в условиях неопределенности.

Одной из важных сфер приложения теории вероятностей и математической статистики является экономика, так как при исследовании и прогнозировании экономических показателей используется эконометрика, опирающаяся на теорию вероятностей. Практическое значение вероятностных методов состоит в том, что они позволяют по известным характеристикам простых случайных явлений прогнозировать характеристики более сложных явлений. Знания и практические навыки, приобретенные в ходе изучения данного курса, могут найти применение при изучении дальнейшего цикла специальных финансовых дисциплин, а также в курсовом и дипломном проектировании.

Основная цель лабораторного практикума - дать краткие основные теоретические положения, рассмотреть примеры решения типовых задач статистической обработки выборки, вопросы графического представления выборки, вычисления ее числовых характеристик и проверки близости эмпирической и теоретической функций распределения. Рассмотрены способы решения задач в среде электронных таблиц MS Excel.

## Лабораторная работа №1 ВАРИАЦИОННЫЕ РЯДЫ.

### ВЫБОРОЧНАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

**Цель работы:** овладеть навыками составления дискретных и интервальных вариационных рядов выборки, построения выборочной (эмпирической) функции распределения в среде ЭТ MS.

#### Краткая теория

Для решения задач, связанных с анализом данных при наличии случайных непредсказуемых воздействий, разработан математический аппарат – математическая статистика, что позволяет выявлять закономерности на основе случайностей, делать на их основе обоснованные выводы и прогнозы.

Важнейшими понятиями математической статистики являются понятия генеральной совокупности и выборки.

**Генеральной совокупностью** наблюдаемого признака (случайной величины)  $X$  называют множество всевозможных значений, принимаемых наблюдаемым признаком  $X$ .

Часть отобранных объектов из генеральной совокупности называется **выборочной совокупностью**, или **выборкой**. Результаты измерений изучаемого признака  $n$  объектов выборочной совокупности порождают  $n$  значений  $x_1, x_2, \dots, x_n$  случайной величины  $X$ . Число  $n$  называется **объемом выборки**.

Выборку можно рассматривать двояко:

- а) как случайный вектор длины  $n$ , каждая компонента которого имеет такое же распределение, как и наблюдаемый признак;
- б) как на результаты измерений, т.е. набор  $n$  чисел.

Случайная величина  $X$  называется **дискретной случайной величиной**, если она принимает свое значение из некоторого конечного фиксированного набора, например, случайная величина  $X$  – число появления шестерки при двух бросках игрального кубика

$$X: 0, 1, 2 .$$

Случайная величина  $X$  называется **непрерывной случайной величиной**, если она принимает любое значение из некоторого интервала (в том числе  $-\infty$  и  $+\infty$ ), например, рост человека.

После получения выборки имеем данные, которые представляют собой множество чисел, расположенных в беспорядке. Анализ таких данных весьма затруднителен, и для изучения скрытых закономерностей их подвергают определенной обработке.

Простейшая операция – **ранжирование** опытных данных, результатом которого являются значения, расположенные в порядке не убывания. Если среди элементов встречаются одинаковые, то они объединяются в одну группу. Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется **вариантом**, а изменение этого значения – **варьированием**. Варианты будем обозначать строчными буквами с соответствующими порядковому номеру группы индексами  $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ , где  $N$  – число групп. При этом  $x^{(1)} < x^{(2)} < \dots < x^{(N)}$ .

Численность отдельной группы сгруппированного ряда данных называется частотой  $n_i$ , где  $i$  – индекс варианта, а отношение частоты данного варианта к общей сумме частот называется частностью (или относительной частотой) и обозначается  $\omega_i, i = 1, \dots, N$ , т.е.

$$\omega_i = \frac{n_i}{\sum_{j=1}^N n_j},$$

при этом  $\sum_{j=1}^N n_j = n$  – объему выборки.

**Дискретным вариационным рядом** называется ранжированная совокупность вариантов  $x^{(i)}$  с соответствующими им частотами  $n_i$  или частностями  $\omega_i$ .

Если число возможных значений дискретной случайной величины достаточно велико или наблюдаемая случайная величина является непрерывной, то строят **интервальный вариационный ряд**, под которым понимают упорядоченную совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частностями попаданий в каждый из них значений случайной величины.

Как правило, частичные интервалы, на которые разбивается весь интервал варьирования, имеют одинаковую длину  $\Delta$ , которая может быть вычислена по следующей формуле

$$\Delta = \frac{R}{N} = \frac{x_{max} - x_{min}}{N}.$$

где  $R$  – размах варьирования (изменения) случайной величины;

$x_{max}$ ,  $x_{min}$  – наибольшее и наименьшее значения исследуемой случайной величины;

$N$  – число частичных интервалов группировки.

Некоторые авторы рекомендуют пользоваться следующими эмпирическими формулами для определения числа интервалов:

$$N = \sqrt{n}, N = 5 \lg(n),$$

$N = 1 + 3,322 \lg(n)$  – формула Стерджеса.

В рекомендациях по стандартизации Р 50.1.033-2001 "Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа хи-квадрат" рекомендует следующие значения  $N$  в зависимости от объема выборки  $n$ :

Объем выборки $n$	Число интервалов группировки $N$
40 – 100	7 – 9
100 – 500	8 – 12
500 – 1000	10 – 16
1000 – 10000	12 – 22

В теории вероятностей для характеристики распределения случайной величины  $X$  служит функция распределения

$$F(x) = P(X < x),$$

определяющую для каждого значения  $x$  вероятность того, что случайная величина  $X$  примет значение, меньшее  $x$ , т.е. равная вероятности события  $A = \{X < x\}$ , где  $x$  – любое действительное число.

Одной из основных характеристик выборки является **выборочная (эмпирическая) функция распределения**

$$F_n^*(x) = \frac{n_x}{n},$$

где  $n_x$  – количество элементов выборки, меньших чем  $x$ . Другими словами,  $F_n^*(x)$  есть относительная частота появления события  $A = \{X < x\}$  в  $n$  независимых испытаниях. Главное различие между  $F(x)$  и  $F_n^*(x)$  состоит в том, что  $F(x)$  определяет вероятность события  $A$ , а

выборочная функция распределения  $F_n^*(x)$  – относительную частоту этого события.

Свойства функции  $F_n^*(x)$ :

1.  $0 \leq F_n^*(x) \leq 1$ .
2.  $F_n^*(x)$  – неубывающая функция.
3.  $F_n^*(-\infty) = 0$ ;  $F_n^*(+\infty) = 1$ .

Функция  $F_n^*(x)$  является "ступенчатой", имеются разрывы в точках, которым соответствуют наблюдаемые значения вариантов. Величина скачка равна относительной частоте варианта.

Аналитически  $F_n^*(x)$  задается следующим соотношением:

$$F_n^*(x) = \begin{cases} 0 & \text{при } x \leq x^{(1)}; \\ \sum_{j=1}^{i-1} \omega_j & \text{при } x^{(i-1)} < x \leq x^{(i)}, \quad i = 2, 3, \dots, N; \\ 1 & \text{при } x > x^{(N)}, \end{cases}$$

где  $\omega_i$  – соответствующие относительные частоты;

$x^{(i)}$  – элементы вариационного ряда (варианты).

*Замечание.* В случае интервального вариационного ряда под  $x^{(i)}$  понимается середина  $i$ -го частичного интервала. Эмпирическую функцию распределения непрерывной случайной величины так же называют «накопленная частота».

Перед вычислением  $F_n^*(x)$  полезно построить дискретный или интервальный вариационный ряд.

### Пример выполнения

**Постановка задачи 1.** На телефонной станции проводились наблюдения над числом неправильных соединений в минуту. Наблюдения в течение 30 минут дали следующие результаты (табл. 1).

Таблица 1.

3	0	1	5	1	2	4	5	3	4
2	4	2	0	2	3	1	3	2	1
4	3	0	2	1	0	4	2	3	2

**Требуется** найти дискретный вариационный ряд, выборочную (эмпирическую) функцию распределения данной выборки и построить ее график в среде ЭТ MS Excel.

**Решение.**

Очевидно, что число  $X$  является дискретной случайной величиной, а полученные данные есть значения этой случайной величины.

В результате выполнения операций ранжирования и группировки были получены шесть значений случайной величины (варианты): 0; 1; 2; 3; 4; 5. При этом значение 0 в этой группе встречается 4 раза, значение 1 – 5 раз, значение 2 – 8 раз, значение 3 – 6 раз, значение 4 – 5 раз, значение 5 – 2 раза. Вычисленные значения частот и частностей приведены в табл. 2.

Таблица 2.

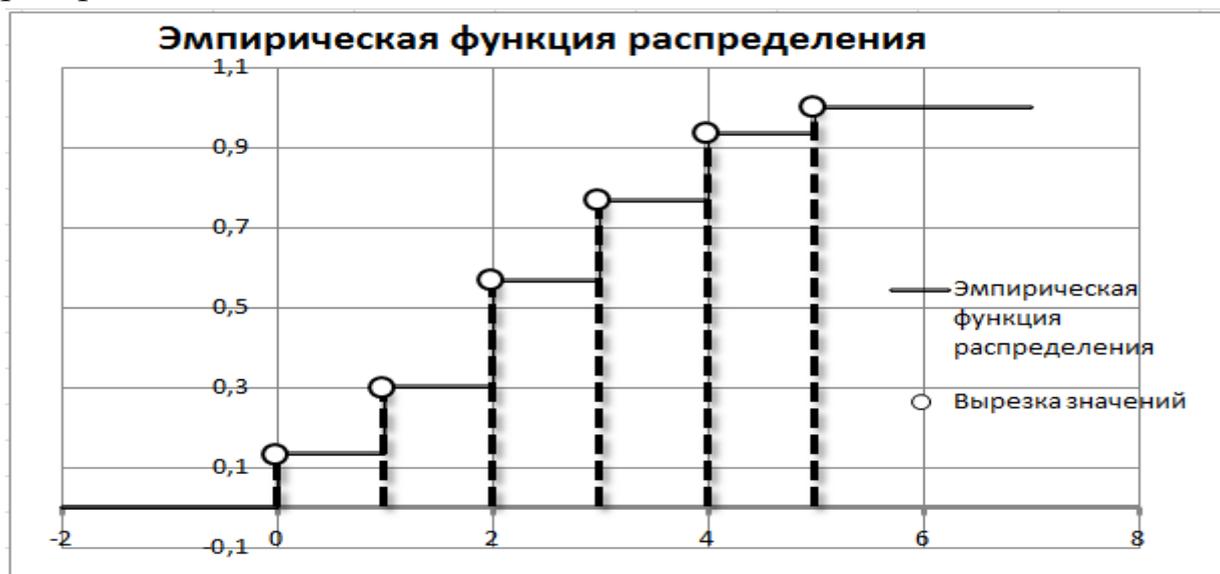
Индекс	$i$	1, 2, 3, 4, 5, 6
Вариант	$x^{(i)}$	0, 1, 2, 3, 4, 5
Частота	$n_i$	4, 5, 8, 6, 5, 2
Частность	$\omega_i$	$\frac{4}{30}, \frac{5}{30}, \frac{8}{30}, \frac{6}{30}, \frac{5}{30}, \frac{2}{30}$

Используя данный дискретный вариационный ряд (см. табл. 2), вычислим значения  $F_n^*(x)$  по формуле, приведенной выше, и занесем их в табл. 3.

Таблица 3.

$x$	$F_{30}^*(x)$
$x \leq 0$	0
$0 < x \leq 1$	$\omega_1 = \frac{4}{30}$
$1 < x \leq 2$	$\omega_1 + \omega_2 = \frac{4}{30} + \frac{5}{30} = \frac{9}{30}$
$2 < x \leq 3$	$\omega_1 + \omega_2 + \omega_3 = \frac{4}{30} + \frac{5}{30} + \frac{8}{30} = \frac{17}{30}$
$3 < x \leq 4$	$\omega_1 + \omega_2 + \omega_3 + \omega_4 = \frac{4}{30} + \frac{5}{30} + \frac{8}{30} + \frac{6}{30} = \frac{23}{30}$
$4 < x \leq 5$	$\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 = \frac{4}{30} + \frac{5}{30} + \frac{8}{30} + \frac{6}{30} + \frac{5}{30} = \frac{28}{30}$
$x > 5$	$\omega_1 + \omega_2 + \omega_3 + \omega_4 + \omega_5 + \omega_6 = \frac{28}{30} + \frac{2}{30} = \frac{30}{30} = 1$

По данным таблицы 3 построим график эмпирической функции распределения.



**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MS Excel необходимо выполнить следующие действия:

1. Идентифицируйте свою работу, переименовав Лист1 в Титульный лист и записав номер лабораторной работы, ее название, кто выполнил и проверил.

2. Переименуйте Лист 2 в Дискретный. Наберите массив 30 значений исходных данных выборки.

	A	B	C	D	E	F	G	H	I	J	K
1		Выборка X									
2		3	0	1	5	1	2	4	5	3	4
3		2	4	2	0	2	3	1	3	2	1
4		4	3	0	2	1	0	4	2	3	2

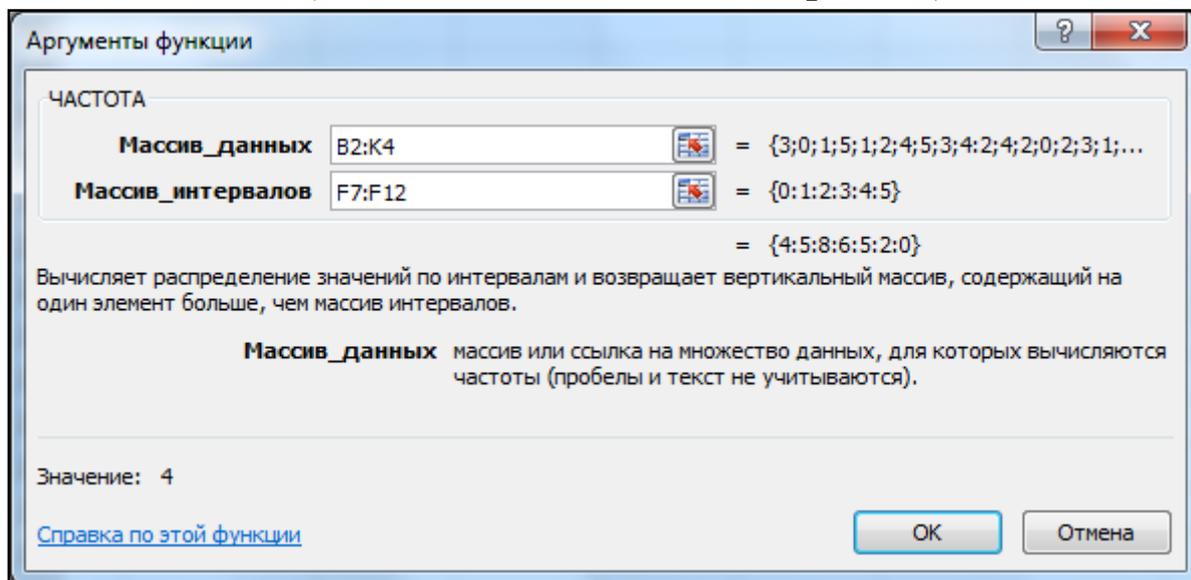
3. Найдите величины  $x_{max}$ ,  $x_{min}$ ,  $n$ , используя встроенные функции Excel **МАКС**, **МИН** и **СЧЕТ**.

	A	B	C	D	E
1		Выборка X			
2		3	0	1	5
3		2	4	2	0
4		4	3	0	2
5					
6		$x_{max} =$	5		
7		$x_{min} =$	0		
8		$n =$	30		

4. Сформируйте столбец вариант  $x^{(i)}$  от 0 до 5 и с помощью функции *ЧАСТОТА* найдите частоту появления значений случайной величины  $X$  в данном интервале.

Синтаксис функции:

**ЧАСТОТА(массивданных; массивинтервалов).**



Массив данных – массив или ссылка на множество данных, для которых вычисляются частоты. В нашем случае это диапазон B2:K2. Если массив данных не содержит значений, то функция *ЧАСТОТА* возвращает массив нулей.

Массив интервалов – массив или ссылка на множество интервалов, в которые группируются значения аргумента массив данных. В нашем случае это диапазон F7:F12. Если массив интервалов не содержит значений, то функция *ЧАСТОТА* возвращает количество элементов в аргументе Массив данных.

ЧАСТОТА									
A	B	C	D	E	F	G	H	I	
1	Выборка X								
2	3	0	1	5	1	2	4	5	
3	2	4	2	0	2	3	1	3	
4	4	3	0	2	1	0	4	2	
5									
6	$x_{\max} =$	5			Вариант	Частота	Частность	F(x)	
7	$x_{\min} =$	0			0	=ЧАСТОТА(B2:K4;F7:F12)	0,133333	0,133333	
8	n =	30			1	5	0,166667	0,3	
9					2	8	0,266667	0,566667	
10					3	6	0,2	0,766667	
11					4	5	0,166667	0,933333	
12					5	2	0,066667	1	
13						0			

Аргументы функции

ЧАСТОТА

Массив\_данных: B2:K4 = {3;0;1;5;1;2;4;5;3;4;2;4;2;0;2;3;1;...}

Массив\_интервалов: F7:F12 = {0;1;2;3;4;5}

= {4;5;8;6;5;2;0}

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

**Массив\_данных** массив или ссылка на множество данных, для которых вычисляются частоты (пробелы и текст не учитываются).

Значение: 4

[Справка по этой функции](#) OK Отмена

Функция *ЧАСТОТА* вводится как формула массива после выделения интервала смежных ячеек, в которые нужно вернуть полученный массив частот.

Количество элементов в возвращаемом массиве на единицу больше числа элементов в массиве интервалов. Дополнительный элемент в возвращаемом массиве содержит количество значений, больших, чем максимальное значение в интервалах, т.е. больше 5 в нашем случае.

Поскольку данная функция возвращает массив, она должна задаваться в качестве формулы массива и работа с ней завершается трехклавишной комбинацией CTRL+SHIFT+ENTER.

Функция *ЧАСТОТА* игнорирует пустые ячейки и тексты.

5. Сформируйте столбец частностей, вычислив значения  $\omega_i, i = 1, \dots, 6$  по формуле

$$\omega_i = \frac{n_i}{n} .$$

ЧАСТОТА									
	A	B	C	D	E	F	G	H	I
1		Выборка X							
2		3	0	1	5	1	2	4	5
3		2	4	2	0	2	3	1	3
4		4	3	0	2	1	0	4	2
5									
6		$x_{\max} =$	5			Вариант	Частота	Частность	F(x)
7		$x_{\min} =$	0			0	4	=G7/\$C\$8	0,133333
8		$n =$	30			1	5	0,166667	0,3
9						2	8	0,266667	0,566667
10						3	6	0,2	0,766667
11						4	5	0,166667	0,933333
12						5	2	0,066667	1
13							0		

6. Сформируйте столбец значений выборочной функции распределения  $F_n^*(x)$ . При этом первое значение в ячейке I7 просто копируется из ячейки H7.

Вариант	Частота	Частность	F(x)
0	4	0,133333	=H7
1	5	0,166667	0,3
2	8	0,266667	0,566667
3	6	0,2	0,766667
4	5	0,166667	0,933333
5	2	0,066667	1
	0		

Следующее значение вычисляется как накопленная сумма предыдущего значения  $\omega_1$  из ячейки I7 и текущего значения  $\omega_2$  из ячейки H8:

$$=I7+H8 .$$

Вариант	Частота	Частность	F(x)
0	4	0,133333	0,133333
1	5	0,166667	=I7+H8
2	8	0,266667	0,566667
3	6	0,2	0,766667
4	5	0,166667	0,933333
5	2	0,066667	1
	0		

Затем данная формула копируется автозаполнением в остальные ячейки диапазона, с выходом на значение, равное 1.

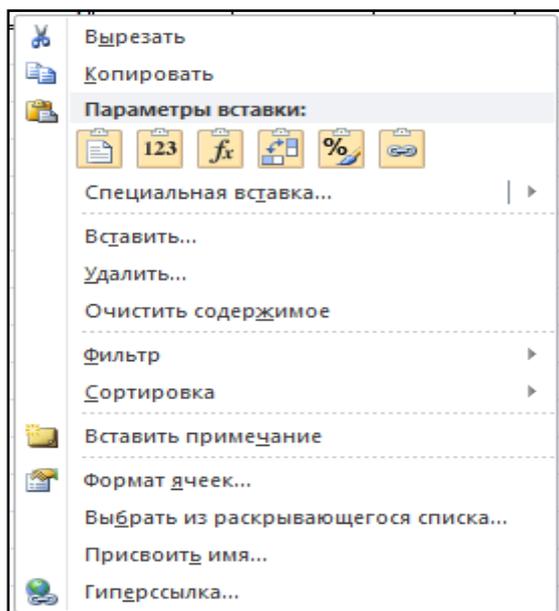
7. Построим график эмпирической функции распределения. С использованием штатных средств Мастера диаграмм ЭТ MS Excel построить ступенчатый график функции распределения дискретной случайной величины нельзя.

Покажем, как в MS Excel все-таки можно построить такой график.

7.1. Расположим данные полученного дискретного вариационного ряда так, как показано на рисунке ниже.

Вариант	Частота	Частность	F(x)	x	F(x)
0	4	0,133333	0,133333		
1	5	0,166667	0,3		
2	8	0,266667	0,566667	0	0,133333
3	6	0,2	0,766667		
4	5	0,166667	0,933333	1	0,3
5	2	0,066667	1		
	0			2	0,566667
				3	0,766667
				4	0,933333
				5	1

При этом данные копируются из предыдущей таблицы. Используют контекстное меню команды Вставка: Параметры вставки → Значения 



7.2. В разреженную таким образом таблицу введем ряд дополнений. В ячейку K7 введем значение -2, а в ячейку K20 значение 7, это границы интервала  $[-2 ; 7]$  на котором будет построен наш график. В оставшиеся пустые ячейки введем значения, чуть меньшие значений полученных вариант (см. случай а) ниже).

x	F(x)
-2	
-0,000001	
0	0,133333
0,99999	
1	0,3
1,99999	
2	0,566667
2,99999	
3	0,766667
3,99999	
4	0,933333
4,99999	
5	1
7	

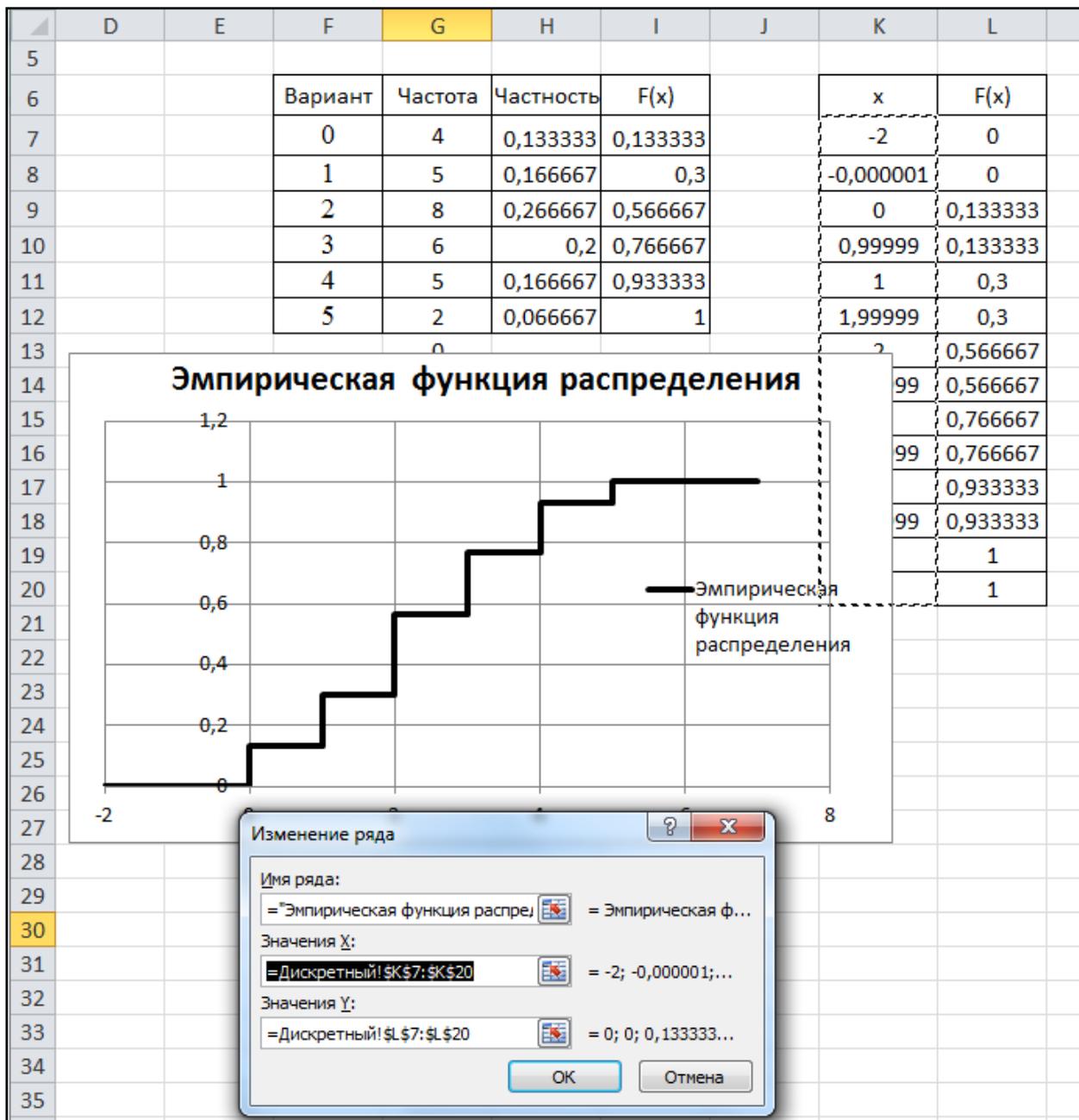
Случай а)

x	F(x)
-2	0
-0,000001	0
0	0,133333
0,99999	0,133333
1	0,3
1,99999	0,3
2	0,566667
2,99999	0,566667
3	0,766667
3,99999	0,766667
4	0,933333
4,99999	0,933333
5	1
7	1

Случай б)

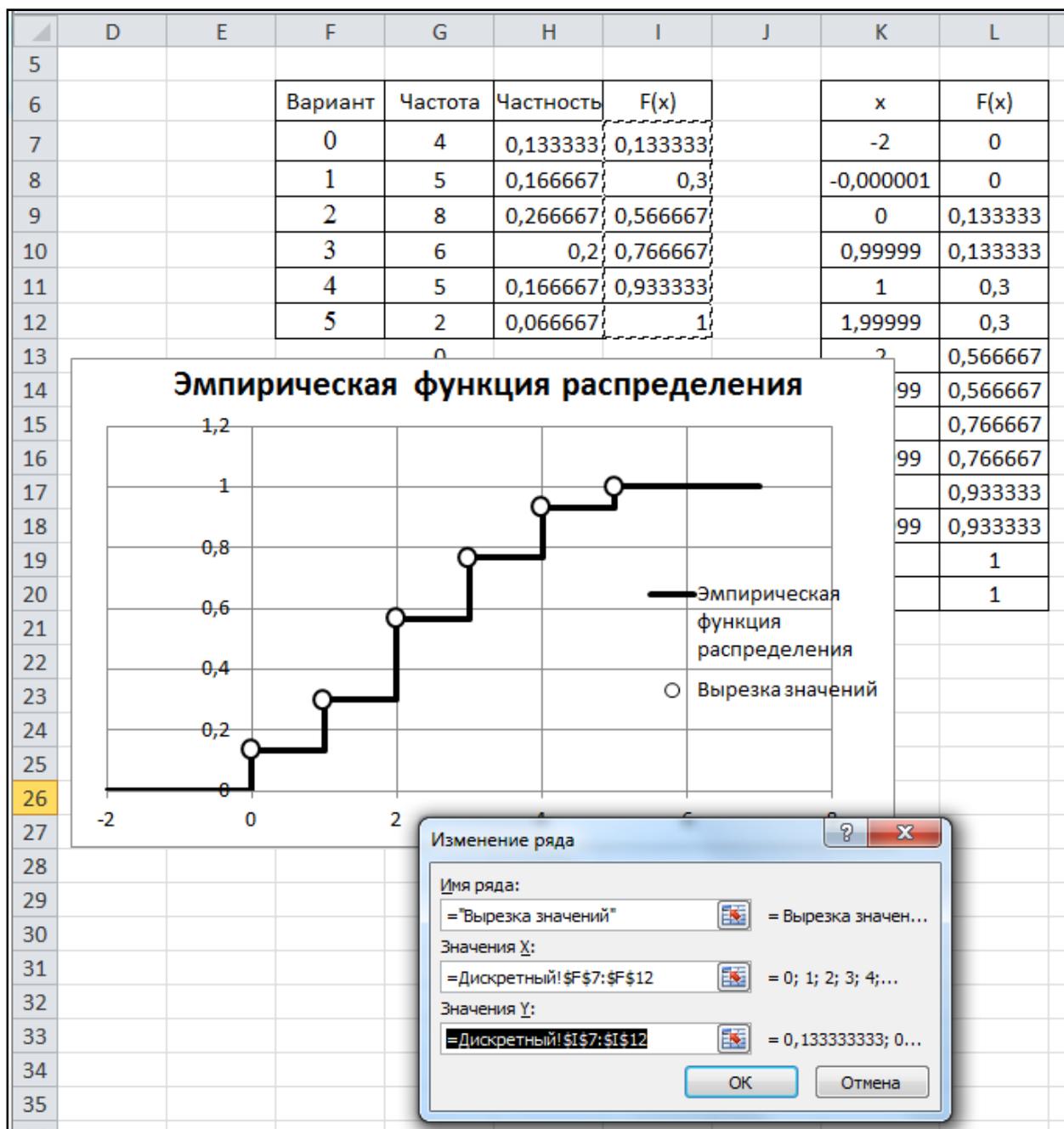
Два первых значения функции  $F(x)$  в ячейках L7 и L8 примем равным нулю, т.к.  $F_n^*(x) = 0$  при  $x \leq x^{(1)}$ . В оставшиеся пустые ячейки скопируем значения функции, расположенные выше (см. случай б) выше).

7.3. По данным, находящимся в диапазоне ячеек K7:L20, с помощью Мастера диаграмм, построим диаграмму типа Точечная без маркеров. Отформатируем диаграмму, убрав маркеры и задав линию, соединяющую табличные значения.

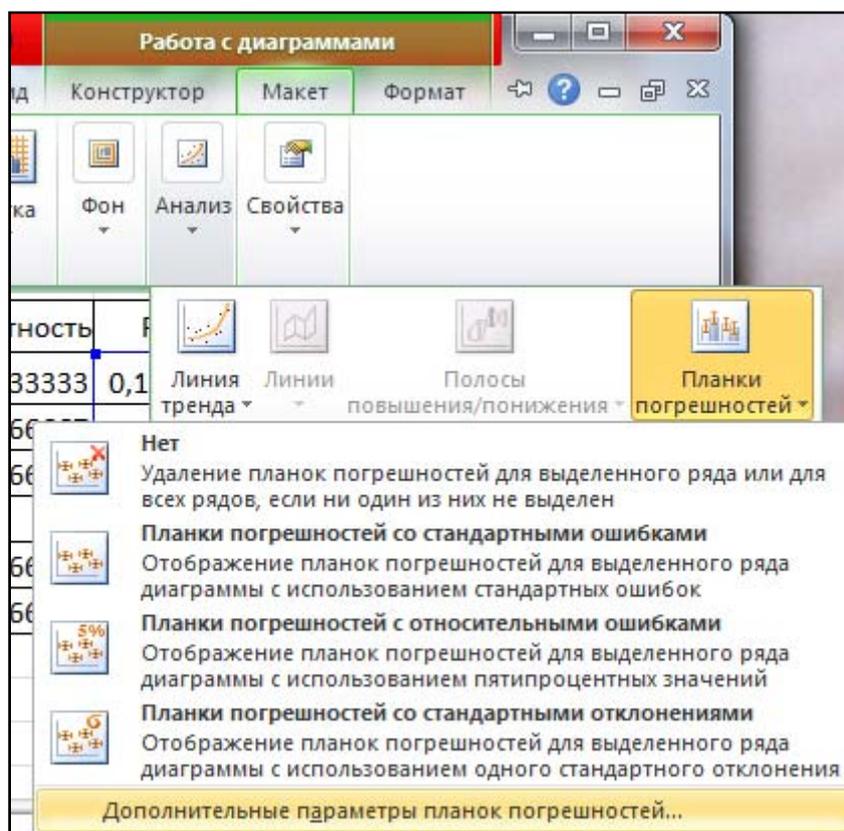


Т.к. функция  $F(x)$  – непрерывна слева в любой точке  $x$ , т.е.  $F(x-0) = F(x)$ , то устраним неоднозначность в точках разрыва,

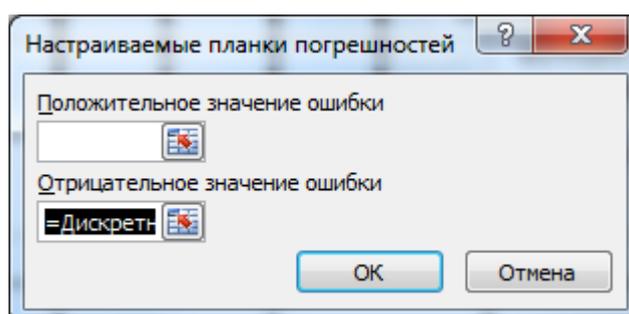
“вырезав” соответствующие значения. Для этого построим точечный график по данным первого и последнего столбца полученного дискретного вариационного ряда.



8. Постройте пунктирные линии в вырезанных точках графика. Для этого выделим точки графика и на вкладке Макет в группе Анализ нажмём кнопку Планки погрешностей, а затем выберем строку Дополнительные параметры планок погрешностей ... .



В диалоговом окне **Формат** планок погрешностей выполните установки, представленные ниже. Установите радиокнопку –  пользовательская и в появившемся окне, в поле ввода Отрицательное значение ошибки введите значения столбца  $F(x)$ .



6	$x_{\max} =$	5	Вариант	Частота	Частность	F(x)	x
7	$x_{\min} =$	0	0	4	0,133333	0,133333	-2
8	n =	30	1	5	0,166667	0,3	-0,000001
9			2	8	0,266667	0,566667	0
10			3	6	0,2	0,766667	0,99999
11			4	5	0,166667	0,933333	1
12			5	2	0,066667	1	1,99999
13				0			2
14							2,99999
15							3
16							3,99999
17							4
18							4,99999
19							5
20							7

**Формат планок погрешностей**

**Вертикальные планки погрешностей**

Цвет линии  
Тип линии  
Тень  
Свечение и сглаживание

**Вывод**

Направление  
 Все  
 Минус  
 Плюс

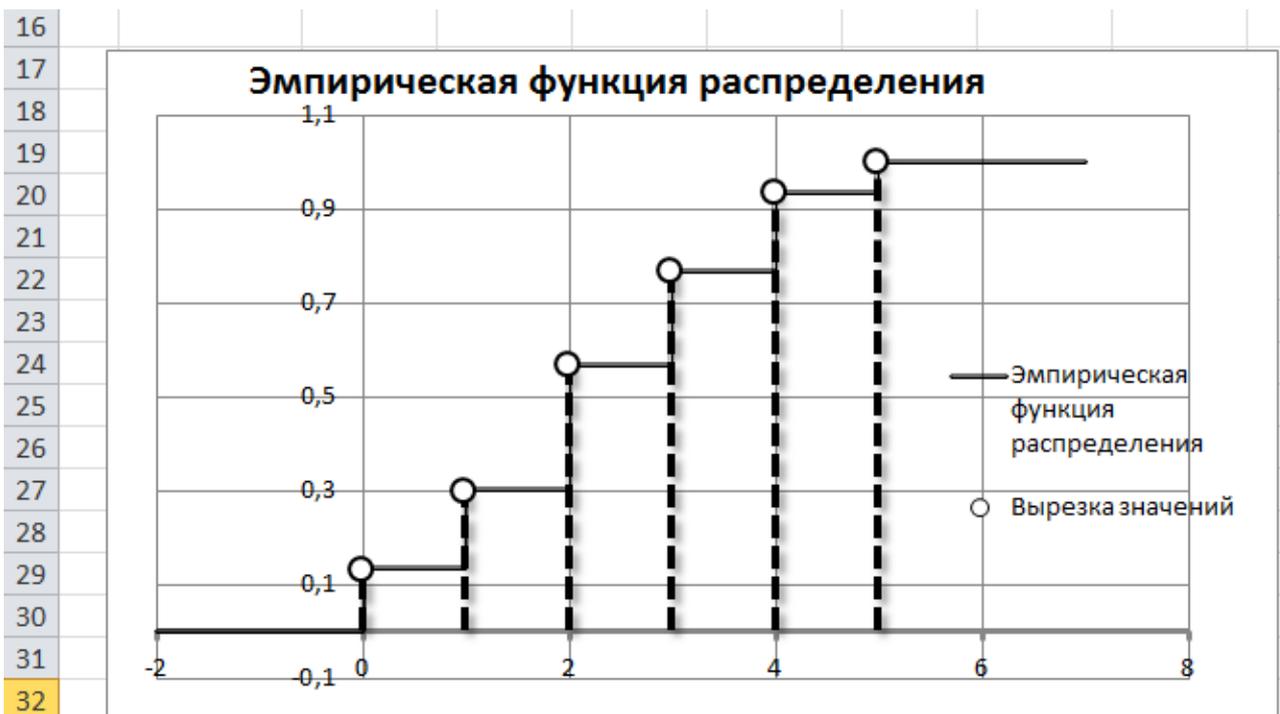
Конечный стиль  
 Без точки  
 Точка

Величина погрешности  
 фиксированное значение: 0,1  
 относительное значение: 5,0 %  
 стандартное отклонение: 1,0  
 стандартная погрешность  
 пользовательская:

Настраиваемые планки погрешностей

=Дискретный! \$I\$7:\$I\$12

Получили график функции распределения с пунктирными линиями.



9. Сделайте выводы и сохраните работу в вашем каталоге.

**Постановка задачи 2.** Исследуется рост учащихся (в сантиметрах) в студенческой группе из 25 человек. Получена выборка (см. табл. 4) из следующих 25 значений.

Таблица 4.

184	182	182	180	177
179	173	179	192	173
190	163	177	186	170
178	185	173	179	165
179	173	179	166	170

**Требуется:** найти интервальный вариационный ряд, выборочную (эмпирическую) функцию распределения данной выборки и построить ее график в среде ЭТ MS Excel.

### Решение.

Найдем максимальное и минимальное значения в исследуемой выборке

$$x_{max} = 192, \quad x_{min} = 163 \text{ см.}$$

Вычислим размах варьирования  $R$  исследуемого признака по формуле

$$R = x_{max} - x_{min} = 29.$$

Для нахождения числа интервалов группировки  $N$  воспользуемся формулой

$$N \approx \sqrt{n} = \sqrt{25} = 5.$$

Далее следует группировка выборки. При этом интервал варьирования признака  $[x_{min}, x_{max}]$  разбивается на  $N$  интервалов группировки одинаковой длины  $\Delta$ , а затем подсчитывается число попаданий признака в  $j$ -й интервал группировки –  $n_i, i=\overline{1, N}$ .

$$\Delta = \frac{R}{N} = \frac{x_{max} - x_{min}}{N} = 5,8 \approx 6.$$

При этом каждый интервал группировки  $\Delta_i=(a_i;b_i)$  характеризуется своим правым и левым концом, числом  $n_i$  – попаданием признака в этот интервал. Иногда интервал характеризуют не границами, а его средним значением.

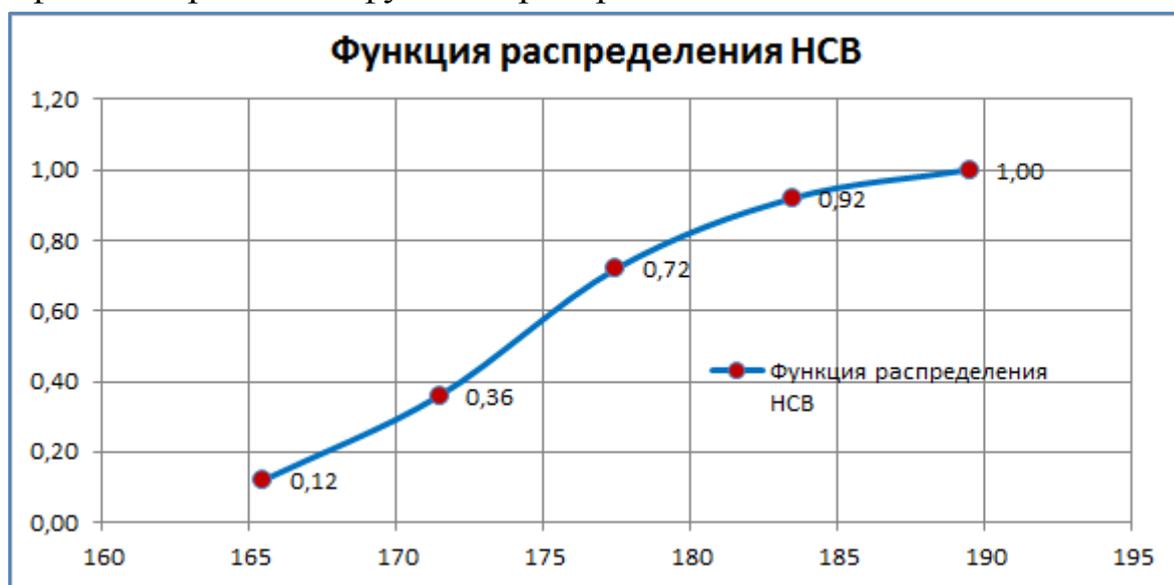
Дальнейшие вычисления удобно представить в табл. 5.

Таблица 5.

$i$	Интервал группировки и $\Delta_i$	Кол-во попаданий в интервал	Частоты $n_i$	Относительные частоты $\omega_i = \frac{n_i}{n}$	Накопленные частоты
1	162,5-168,5		3	3/25	3/25
2	168,5-174,5		6	6/25	9/25
3	174,5-180,5		9	9/25	18/25
4	180,5-186,5		5	5/25	23/25
5	186,5-192,5		2	2/25	25/25 = 1
$\Sigma$			25	1	

Чтобы значение исследуемого признака не попадало на границы интервала группировки, примем минимальное значение признака не 163, а 162,5 и от этого значения начнем строить интервалы длиной  $\Delta = 6$  (см. второй столбец табл. 5).

Откладывая по оси абсцисс средние значения интервалов группировки, а по оси ординат – значения накопленных частот, строим график эмпирической функции распределения.



**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MS Excel необходимо выполнить следующие действия:

1. Переименуйте Лист 3 в Непрерывный. Наберите массив 25 значений исходных данных выборки.

2. Найдите величины  $x_{max}$ ,  $x_{min}$ ,  $n$ ,  $N$ ,  $\Delta_{округл}$  используя встроенные функции Excel *МАКС*, *МИН*, *СЧЕТ*, *КОРЕНЬ* и *ОКРУГЛ*.

3. Сформируйте столбец интервалов варьирования от значения 162,5 с шагом  $\Delta = 6$ . Первое значение набираем с клавиатуры, а второе вычисляем с помощью формулы

$$=E9+\$C\$13 .$$

Остальные значения получим копированием с помощью Автозаполнения.

4. Сформируйте столбец Частота и с помощью функции *ЧАСТОТА* найдите частоту появления значений исследуемой случайной величины  $X$  в каждом из интервалов.

5. Заполните столбец относительных частот, рассчитав значение в ячейке G9 по формуле

$$=F9/\$C\$10 .$$

Остальные значения получим копированием формулы с помощью Автозаполнения.

6. Вычислите середины интервалов группировки, рассчитав значение в ячейке H9 по формуле

$$=(E9+E10)/2 .$$

Остальные значения в диапазоне H10:H13 получим копированием формулы с помощью Автозаполнения.

7. Заполните столбец накопленных частот. При этом, значение в ячейке I9 получим, копируя значение ячейки G10 по формуле

$$=G10 .$$

Значение в ячейке I10 получим по формуле

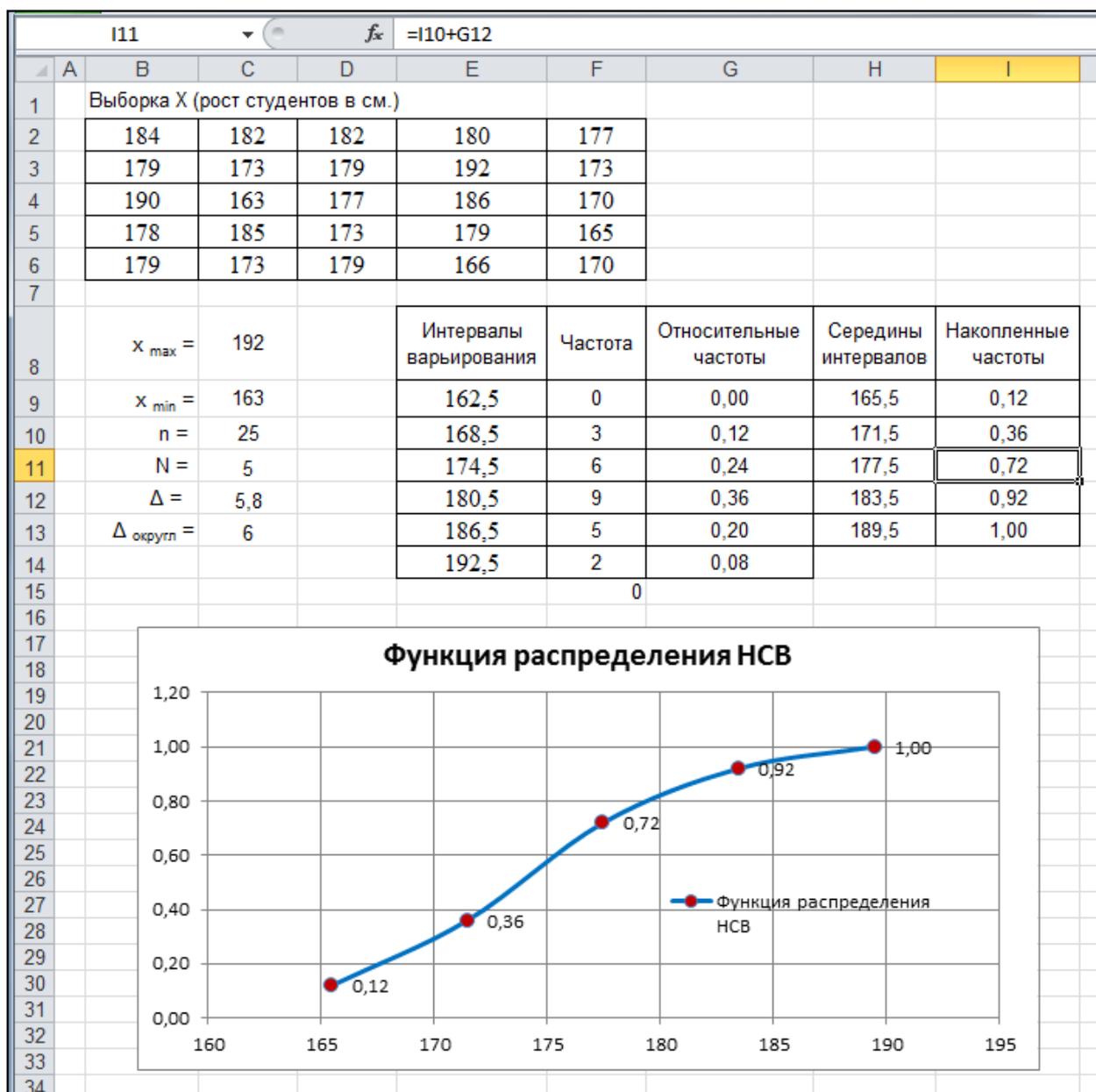
$$=I9+G11 .$$

Остальные значения в диапазоне I11:I13 получим, копируя формулу с помощью Автозаполнения.

8. По данным двух последних столбцов построим график эмпирической функции распределения.

9. Сделайте выводы и сохраните работу в вашем каталоге.

Лист Excel лабораторной работы имеет вид, представленный на рисунке.



### Исходные данные для самостоятельного решения

**Задание 1.** Имеется выборка непрерывной случайной величины объема  $n = 26$  (табл. 6).

**Задание 2.** Имеется выборка дискретной случайной величины объема  $n = 30$  (табл. 7).

**Требуется:** найти дискретный и интервальный вариационные ряды, выборочную (эмпирическую) функцию распределения данных выборок и построить их графики в среде ЭТ MS Excel.

Таблица 6.

вариант	Выборка													
	1	11,7	9,83	5,49	7,43	9,92	3,41	6,83	8,22	8,30	8,14	9,29	9,27	7,43
7,41		3,56	7,72	12,1	6,06	10,6	6,76	8,21	9,86	8,13	9,04	4,75	9,33	
2	4,49	9,25	7,94	9,10	6,27	6,77	3,47	8,84	6,48	4,92	6,98	10,1	6,32	
	6,36	5,16	7,92	12,0	7,46	7,01	13,0	7,34	6,71	5,48	9,95	11,9	8,89	
3	6,13	8,56	9,77	9,17	8,89	6,19	7,70	6,96	6,72	6,08	4,41	5,52	9,59	
	9,02	6,22	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98	
4	6,52	9,27	7,91	5,77	8,02	3,07	2,22	5,76	11,6	6,62	7,07	12,5	1,65	
	10,5	3,67	7,62	4,94	5,39	3,64	4,62	8,88	6,75	5,77	6,38	10,3	5,74	
5	8,18	9,56	6,06	5,85	6,78	5,60	10,8	7,70	6,44	8,64	6,95	5,66	4,84	
	4,96	4,62	5,57	6,47	5,97	8,02	3,66	9,24	4,13	6,58	7,51	5,67	7,89	
6	10,2	9,23	8,77	10,4	9,44	9,09	6,30	9,42	6,12	9,69	8,59	8,68	7,97	
	8,64	6,45	5,29	5,00	8,42	8,84	8,26	6,66	6,96	6,51	6,72	6,00	5,36	
7	7,13	9,12	9,77	9,17	8,89	6,19	7,71	6,96	6,72	6,08	4,41	5,52	9,59	
	8,06	6,26	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98	
8	3,53	9,56	7,03	9,18	7,45	5,59	6,85	11,3	7,90	6,00	6,68	5,66	8,64	
	8,87	4,58	11,3	5,02	4,33	9,31	10,3	5,99	6,98	5,23	8,75	7,73	9,16	
9	3,38	7,87	4,04	8,21	4,08	3,46	4,37	6,66	1,46	5,59	3,78	8,73	5,57	
	8,22	3,25	3,38	4,20	2,49	6,11	4,54	6,53	5,20	3,84	5,35	9,72	4,63	
10	4,21	5,68	3,45	6,79	3,39	2,99	3,88	3,77	1,43	5,96	4,94	6,55	5,92	
	4,20	4,25	5,64	5,58	5,87	5,05	3,55	7,95	4,45	5,85	6,68	1,24	7,09	

Таблица 7.

№ варианта	Выборка															
	1	4	0	2	5	1	2	4	5	3	4	4	0	1	5	1
2		3	2	0	2	3	1	3	2	1	2	4	2	0	2	
2	2	0	3	5	1	2	4	5	3	4	4	0	1	5	1	
	1	3	4	0	2	3	1	3	2	1	2	4	2	0	2	
3	2	3	2	0	2	3	1	1	2	3	2	4	2	0	2	
	2	0	3	5	1	2	4	5	3	4	4	0	1	5	1	
4	4	2	1	5	1	2	4	5	3	4	4	0	1	5	1	
	3	3	4	0	2	3	1	3	2	1	2	4	2	0	2	
5	2	3	4	0	2	1	2	3	2	1	2	2	4	0	2	
	4	3	2	2	1	3	1	3	2	2	1	4	2	2	0	
6	2	3	2	1	2	2	4	0	2	4	4	0	1	5	1	
	1	3	2	2	1	4	2	2	0	1	2	4	2	0	2	
7	4	3	2	2	5	3	1	3	2	2	1	4	2	2	0	
	2	3	2	1	2	2	4	0	2	4	4	0	1	5	1	
8	5	3	4	4	0	1	5	1	2	2	1	4	2	2	0	
	3	2	1	2	4	2	0	2	2	0	4	0	1	5	1	
9	2	3	2	1	2	2	4	0	2	4	4	0	1	5	1	
	5	3	4	4	0	1	5	1	2	2	1	4	2	2	0	
10	0	2	3	5	1	2	4	5	3	4	4	0	1	5	1	
	4	2	1	5	1	2	4	5	3	4	4	0	1	5	1	

## Лабораторная работа №2

### ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ВЫБОРКИ

**Цель работы:** овладеть навыками графического представления выборки в виде гистограммы, полигона и огивы в среде ЭТ MS Excel.

#### Краткая теория

Существует три основных метода графического представления выборочных данных – гистограмма (столбчатая диаграмма), полигон частот и сглаженная кривая (огива).

*Гистограммой частот* называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной  $\Delta$ , а высоты равны отношению  $n_i / \Delta$  (плотность частоты).

Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс, на расстоянии  $n_i / \Delta$ .

Площадь  $i$ -го частичного прямоугольника равна  $\Delta \cdot (n_i / \Delta) = n_i$  – сумме частот вариант  $i$ -го интервала; следовательно, площадь гистограммы частот равна сумме всех частот, то есть объему выборки  $n$ .

*Гистограммой относительных частот* называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной  $\Delta$ , а высоты равны отношению  $\omega_i / \Delta$  (плотность относительной частоты).

Для построения гистограммы относительных частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии  $\omega_i / \Delta$ . Площадь  $i$ -го частичного прямоугольника равна  $\Delta \cdot (\omega_i / \Delta) = \omega_i$  – относительной частоте вариант, попавших в  $i$ -й интервал. Следовательно, площадь гистограммы относительных частот равна сумме всех относительных частот, то есть единице.

*Полигоном частот* называют ломаную, отрезки которой соединяют точки  $(x_1, n_1), (x_2, n_2), \dots, (x_N, n_N)$ .

Для построения полигона частот на оси абсцисс откладывают варианты  $x_i$ , а на оси ординат – соответствующие им частоты  $n_i$ . Точки  $(x_i, n_i)$  соединяют отрезками прямых и получают полигон частот.

*Полигоном относительных частот* называют ломаную, отрезки которой соединяют точки  $(x_1, \omega_1), (x_2, \omega_2), \dots, (x_N, \omega_N)$ .

Для построения полигона частот на оси абсцисс откладывают варианты  $x_i$ , а на оси ординат  $\omega_i$ . Точки  $(x_i, \omega_i)$  соединяют отрезками прямых и получают полигон относительных частот.

*Замечание.* В случае интервального вариационного ряда под  $x_i$  понимается середина  $i$ -го частичного интервала.

*Сглаженная кривая* или *огива*. Иногда вместо гистограммы или полигона частот строят сглаженную кривую. Основное отличие в том, что она проводится по точкам таким образом, чтобы график не имел острых углов или зубцов. Для ее построения по вертикальной оси всегда откладываются значения от 0 до 100 (они соответствуют процентам). По горизонтальной оси откладываются границы интервалов группирования данных. После этого на координатной плоскости наносятся точки, первая координата которой соответствует середине интервала или значению варианты  $x_i$ , а вторая координата – накопленной частоте попадания, выраженной в процентах. Для окончательного построения нанесенные точки соединяются гладкой кривой.

В качестве исходных данных для построения огивы используется таблица, полученная после табулирования данных, но при этом второй столбец этой таблицы (частоты) мы должны преобразовать в накопленные частоты, а затем в проценты.

### Пример выполнения

**Постановка задачи 1.** На телефонной станции проводились наблюдения над числом неправильных соединений в минуту. Наблюдения в течение 30 минут дали следующие результаты (табл. 1).

Таблица 1.

3	0	1	5	1	2	4	5	3	4
2	4	2	0	2	3	1	3	2	1
4	3	0	2	1	0	4	2	3	2

**Требуется** построить гистограмму, полигон и огиву в среде ЭТ MS Excel.

**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MS Excel необходимо выполнить следующие действия:

1. Идентифицируйте свою работу, переименовав Лист1 в Титульный лист и записав номер лабораторной работы, ее название, кто выполнил и проверил.

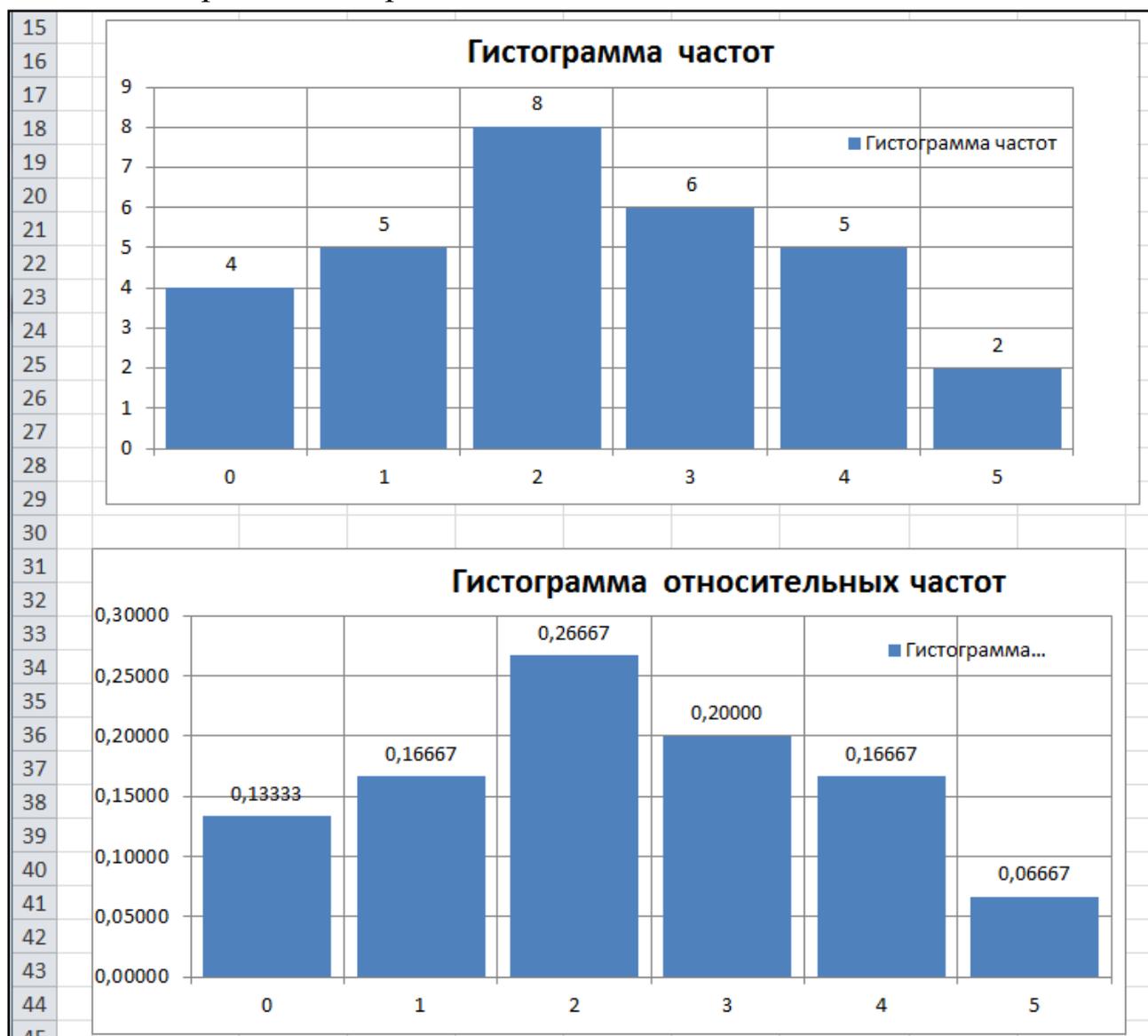
2. Переименуйте Лист 2 в Дискретный. Наберите массив 30 значений исходных данных выборки. Найдите величины  $x_{max}$ ,  $x_{min}$ ,  $n$ ,  $\Delta$  используя встроенные функции Excel *МАКС*, *МИН* и *СЧЕТ*.

3. Сформируйте столбец вариант  $x^{(i)}$  от 0 до 5 и с помощью функции *ЧАСТОТА* найдите частоту появления значений случайной величины  $X$  в данном интервале.

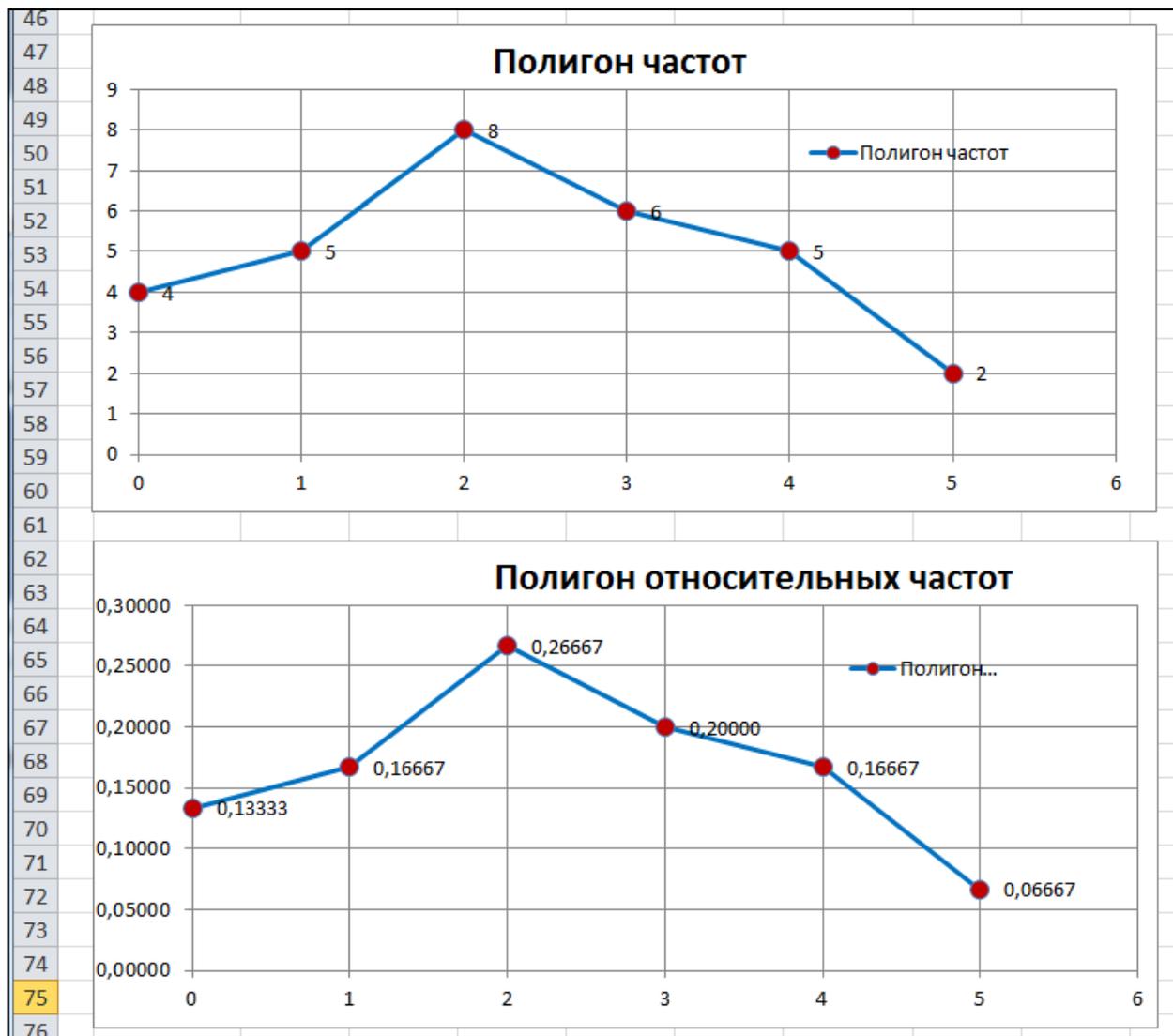
4. Вычислите столбцы значений  $n_i / \Delta$  (плотность частоты) и  $\omega_i / \Delta$  (плотность относительной частоты).

17		fx		=G7/(\$C\$9*\$C\$8)						
A	B	C	D	E	F	G	H	I	J	K
1	Выборка X									
2	3	0	1	5	1	2	4	5	3	4
3	2	4	2	0	2	3	1	3	2	1
4	4	3	0	2	1	0	4	2	3	2
5										
6	$x_{max} =$	5			Вариант	Частота	Плотность $n_i / \Delta$	$n_i / (\Delta \cdot n)$		
7	$x_{min} =$	0			0	4	4	0,13333		
8	$n =$	30			1	5	5	0,16667		
9	$\Delta =$	1			2	8	8	0,26667		
10					3	6	6	0,20000		
11					4	5	5	0,16667		
12					5	2	2	0,06667		
13						0	<b>30</b>	<b>1</b>		

5. Построим гистограммы частот и относительных частот.



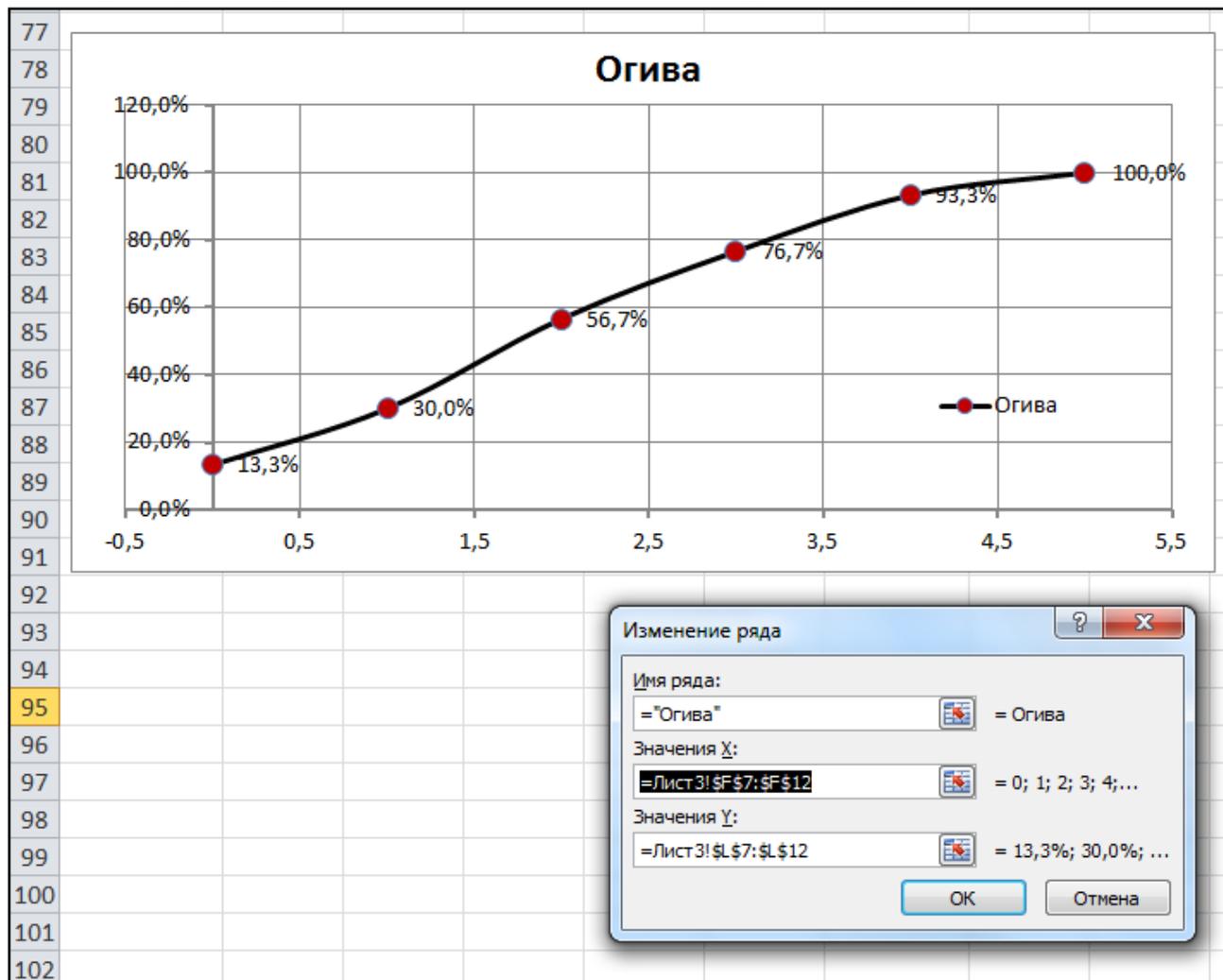
6. Вычислите столбец значений  $\omega_i$  – относительных частот выборки и по данным столбцов 1, 2 и 5 постройте графики полигона частот и полигона относительных частот.



7. Для построения оживы сформируйте столбцы накопленных частот и накопленных частот в процентах.

Вариант	Частота	Плотность $n_i / \Delta$	$n_i / (\Delta \cdot n)$	$\omega_i$	Накопленные частоты	Накопленные частоты в %
0	4	4	0,13333	0,13333	4	13,3%
1	5	5	0,16667	0,16667	9	30,0%
2	8	8	0,26667	0,26667	17	56,7%
3	6	6	0,20000	0,20000	23	76,7%
4	5	5	0,16667	0,16667	28	93,3%
5	2	2	0,06667	0,06667	30	100,0%
	0	<b>30</b>	<b>1</b>			

По первому и двум последним столбцам построим график огивы.



8. Сделайте выводы и сохраните работу в вашем каталоге.

**Постановка задачи2.** Исследуется рост учащихся (в сантиметрах) в студенческой группе из 25 человек. Получена выборка (см. табл. 4) из следующих 25 значений.

Таблица 2.

184	182	182	180	177
179	173	179	192	173
190	163	177	186	170
178	185	173	179	165
179	173	179	166	170

**Требуется** построить гистограмму, полигон и огиву в среде ЭТ MS Excel.

**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MSExcel необходимо выполнить следующие действия:

1. Переименуйте Лист 3 в Непрерывный. Наберите массив 25 значений исходных данных выборки.

2. Выполните пункты 3 – 6 предыдущей лабораторной работы.

Н9		fx		=(E9+E10)/2			
A	B	C	D	E	F	G	H
1	Выборка X (рост студентов в см.)						
2	184	182	182	180	177		
3	179	173	179	192	173		
4	190	163	177	186	170		
5	178	185	173	179	165		
6	179	173	179	166	170		
7							
8	$x_{\max} =$	192		Интервалы варьирования	Частота	Относительные частоты	Середины интервалов
9	$x_{\min} =$	163		162,5	0	0,00	165,5
10	$n =$	25		168,5	3	0,12	171,5
11	$N =$	5		174,5	6	0,24	177,5
12	$\Delta =$	5,8		180,5	9	0,36	183,5
13	$\Delta_{\text{округл}} =$	6		186,5	5	0,20	189,5
14				192,5	2	0,08	
15					0		

3. Сформируйте столбец частот  $n_i$  и скопируйте в него не нулевые данные столбца частот, полученные с помощью встроенной функции *ЧАСТОТА*. Используйте контекстное меню команды Вставка:

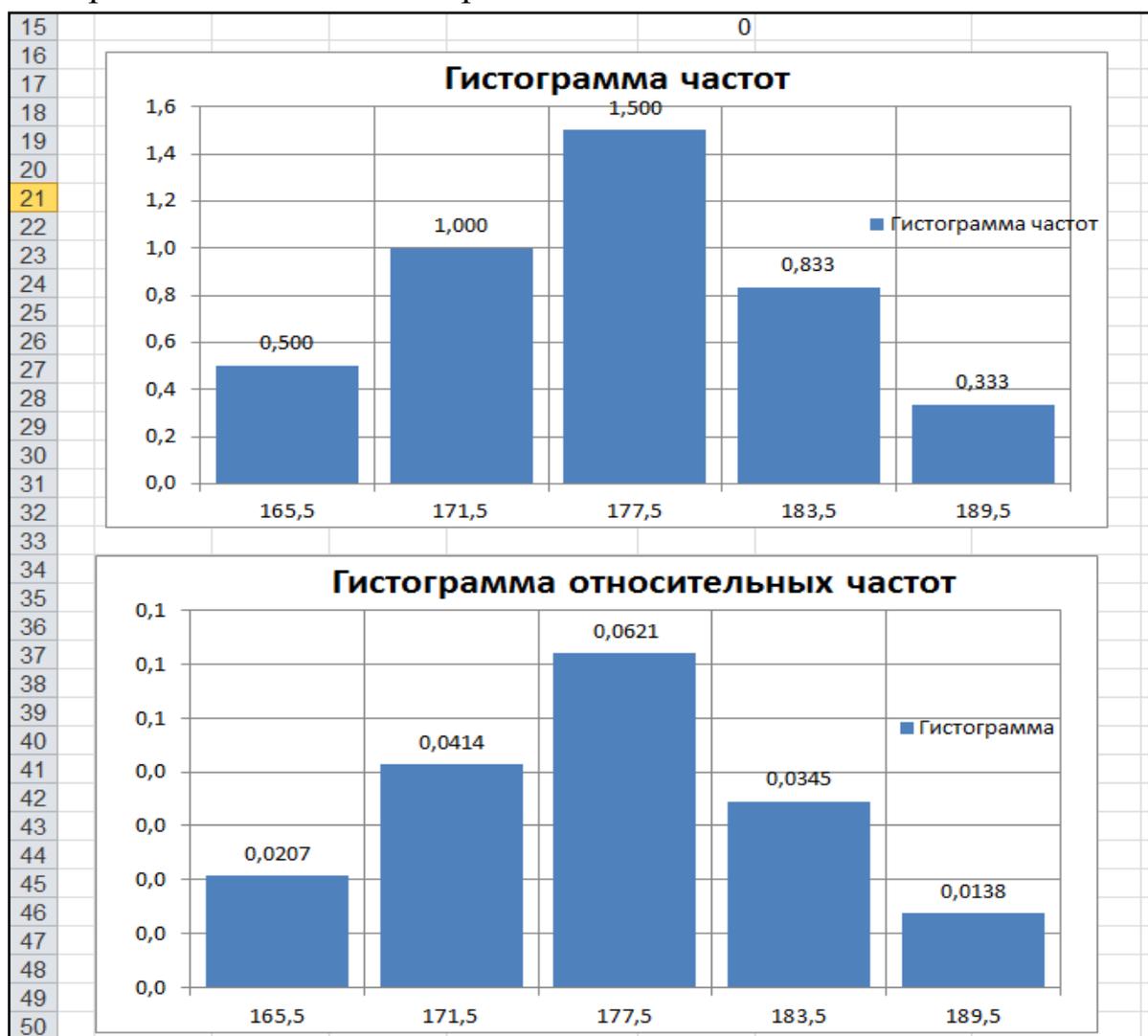
Параметры вставки → Значения 

4. Вычислите плотности частот по формуле  $=I9/SC\$13$  в ячейке J9 . Остальные значения получим копированием с помощью Автозаполнения.

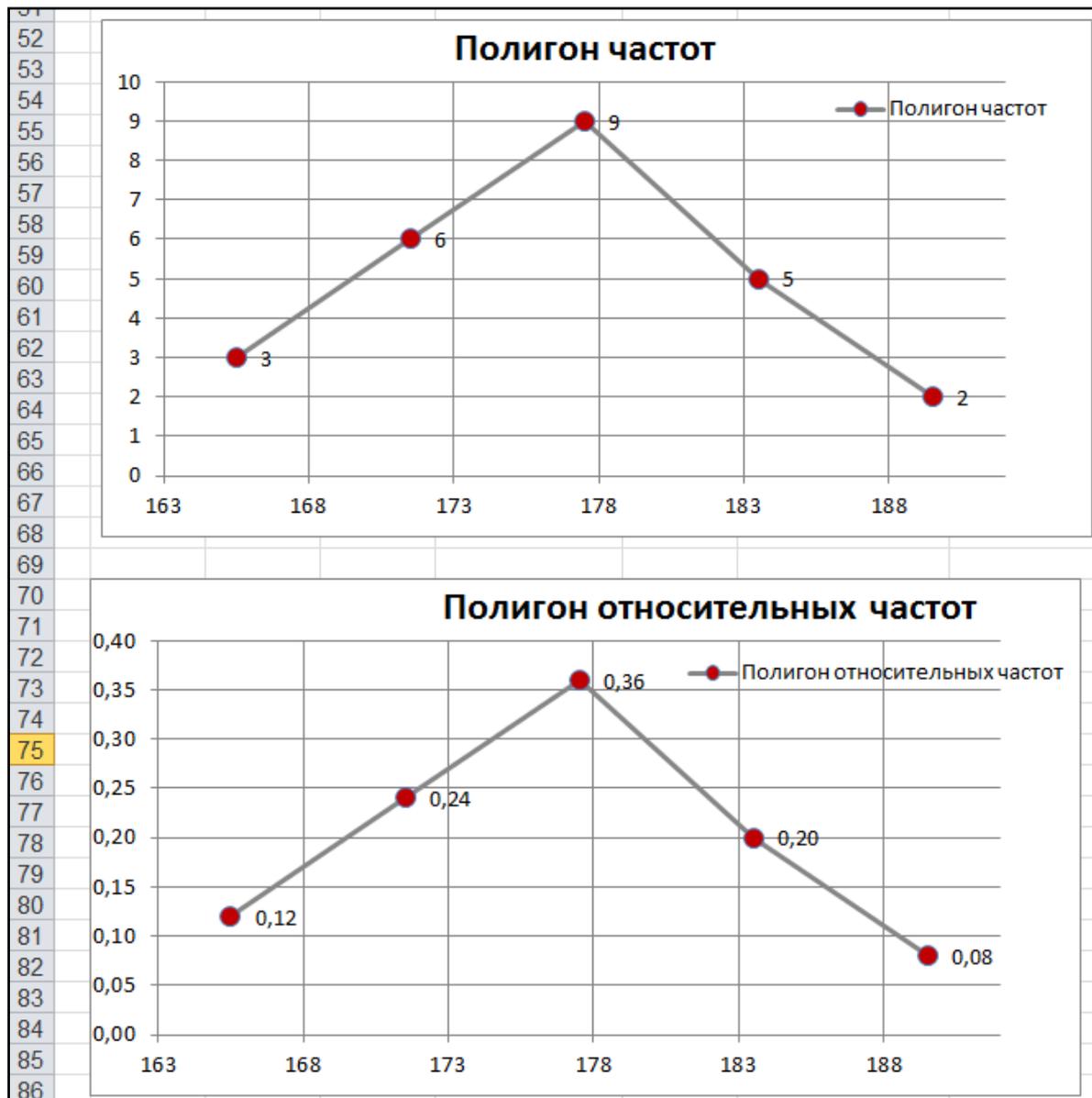
5. Вычислите плотности относительных частот по формуле  $=I9/(SC\$12*SC\$10)$  в ячейке K9 . Остальные значения получим копированием с помощью Автозаполнения. Полученная таблица имеет вид:

ОКРУГЛ								
	D	E	F	G	H	I	J	K
1	дентов в см.)							
2	182	180	177					
3	179	192	173					
4	177	186	170					
5	173	179	165					
6	179	166	170					
7								
8		Интервалы варьирования	Частота	Относительные частоты	Середины интервалов	Частоты $n_i$	Плотности частот $n_i/\Delta$	Плотности относительных частот ( $n_i/\Delta \cdot n$ )
9		162,5	0	0,00	165,5	3	0,500	$=19/(\$C\$12*\$C\$10)$
10		168,5	3	0,12	171,5	6	1,000	0,0414
11		174,5	6	0,24	177,5	9	1,500	0,0621
12		180,5	9	0,36	183,5	5	0,833	0,0345
13		186,5	5	0,20	189,5	2	0,333	0,0138
14		192,5	2	0,08				
15			0					

6. По данным двух последних столбцов построим графики гистограммы частот и гистограммы относительных частот.



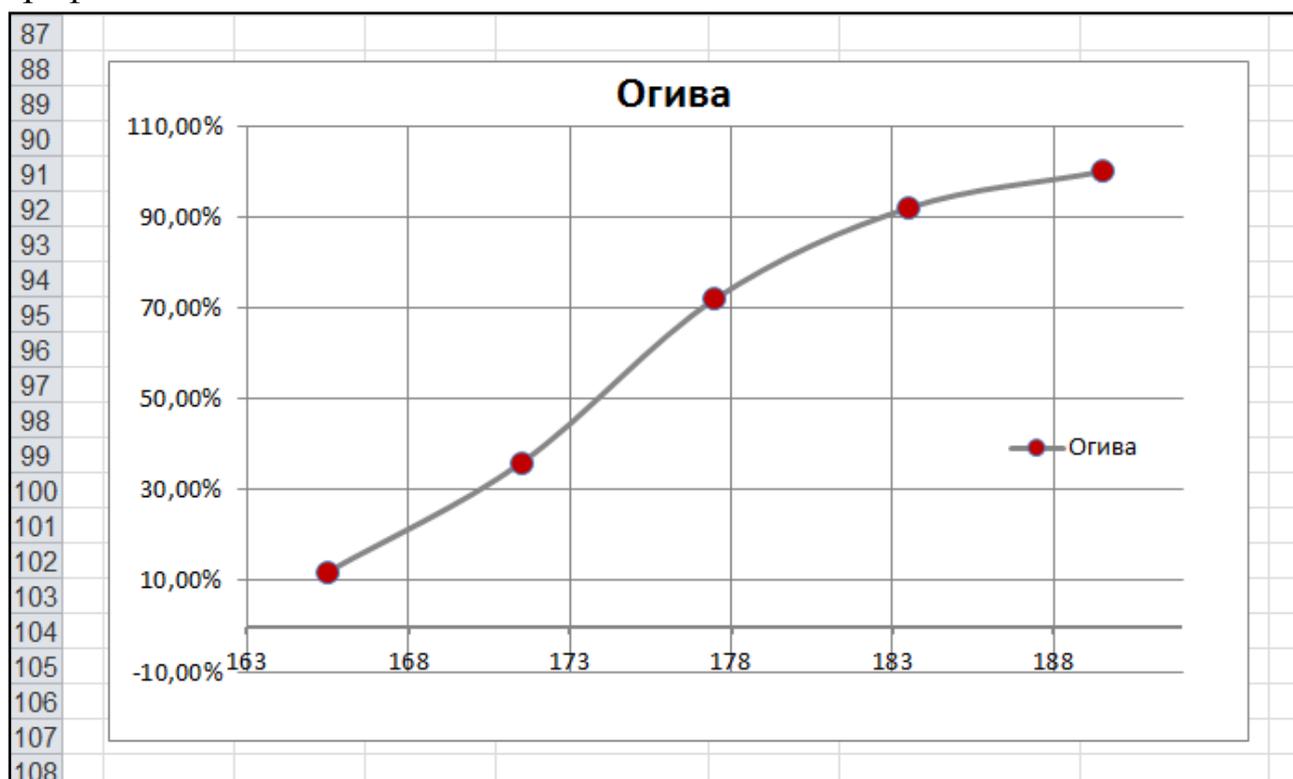
7. Постройте графики полигона частот и полигона относительных частот.



8. Для построения оживы сформируйте столбцы накопленных частот и накопленных частот в процентах.

	$x_{\max} = 192$	Интервалы варьирования	Частота	Относительные частоты	Средины интервалов	Частоты $n_i$	Плотности частот $n_i/\Delta$	Плотности относительных частот ( $n_i/\Delta n$ )	Накопленные частоты	Накопленные частоты в %
9	$x_{\min} = 163$	162,5	0	0,00	165,5	3	0,500	0,0207	3	=L9/\$C\$10
10	$n = 25$	168,5	3	0,12	171,5	6	1,000	0,0414	9	36,00%
11	$N = 5$	174,5	6	0,24	177,5	9	1,500	0,0621	18	72,00%
12	$\Delta = 5,8$	180,5	9	0,36	183,5	5	0,833	0,0345	23	92,00%
13	$\Delta_{\text{округл}} = 6$	186,5	5	0,20	189,5	2	0,333	0,0138	25	100,00%
14		192,5	2	0,08						
15			0							

По столбцу середин интервалов и двум последним столбцам построим график огивы.



9. Сделайте выводы и сохраните работу в вашем каталоге.

### Исходные данные для самостоятельного решения

**Задание 1.** Имеется выборка непрерывной случайной величины объема  $n = 26$  (табл. 3).

**Задание 2.** Имеется выборка дискретной случайной величины объема  $n = 30$  (табл. 4).

**Требуется:** построить гистограмму, полигон и огиву в среде ЭТ MS Excel.

Таблица 3.

№ варианта	Выборка													
	1	3,38	7,87	4,04	8,21	4,08	3,46	4,37	6,66	1,46	5,59	3,78	8,73	5,57
8,22		3,25	3,38	4,20	2,49	6,11	4,54	6,53	5,20	3,84	5,35	9,72	4,63	
2	4,21	5,68	3,45	6,79	3,39	2,99	3,88	3,77	1,43	5,96	4,94	6,55	5,92	
	4,20	4,25	5,64	5,58	5,87	5,05	3,55	7,95	4,45	5,85	6,68	1,24	7,09	
3	11,7	9,83	5,49	7,43	9,92	3,41	6,83	8,22	8,30	8,14	9,29	9,27	7,43	
	7,41	3,56	7,72	12,1	6,06	10,6	6,76	8,21	9,86	8,13	9,04	4,75	9,33	
4	4,49	9,25	7,94	9,10	6,27	6,77	3,47	8,84	6,48	4,92	6,98	10,1	6,32	
	6,36	5,16	7,92	12,0	7,46	7,01	13,0	7,34	6,71	5,48	9,95	11,9	8,89	
5	6,13	8,56	9,77	9,17	8,89	6,19	7,70	6,96	6,72	6,08	4,41	5,52	9,59	
	9,02	6,22	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98	
6	6,52	9,27	7,91	5,77	8,02	3,07	2,22	5,76	11,6	6,62	7,07	12,5	1,65	
	10,5	3,67	7,62	4,94	5,39	3,64	4,62	8,88	6,75	5,77	6,38	10,3	5,74	
7	8,18	9,56	6,06	5,85	6,78	5,60	10,8	7,70	6,44	8,64	6,95	5,66	4,84	
	4,96	4,62	5,57	6,47	5,97	8,02	3,66	9,24	4,13	6,58	7,51	5,67	7,89	
8	10,2	9,23	8,77	10,4	9,44	9,09	6,30	9,42	6,12	9,69	8,59	8,68	7,97	
	8,64	6,45	5,29	5,00	8,42	8,84	8,26	6,66	6,96	6,51	6,72	6,00	5,36	
9	7,13	9,12	9,77	9,17	8,89	6,19	7,71	6,96	6,72	6,08	4,41	5,52	9,59	
	8,06	6,26	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98	
10	3,53	9,56	7,03	9,18	7,45	5,59	6,85	11,3	7,90	6,00	6,68	5,66	8,64	
	8,87	4,58	11,3	5,02	4,33	9,31	10,3	5,99	6,98	5,23	8,75	7,73	9,16	

Таблица 4.

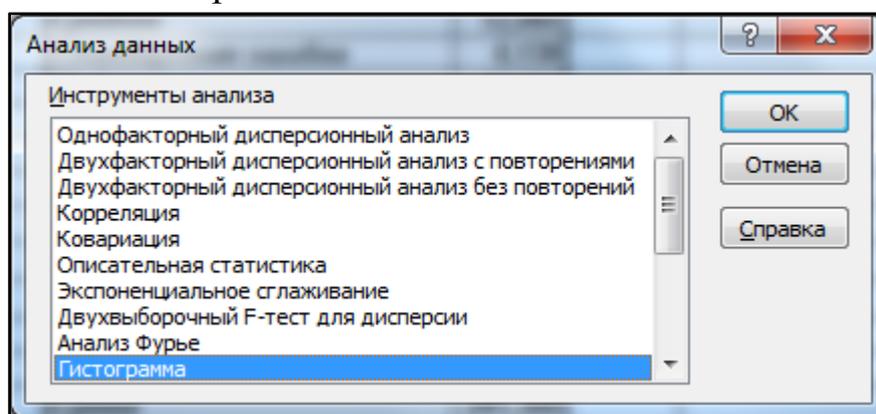
№ варианта	Выборка															
	1	2	3	2	1	2	2	4	0	2	4	4	0	1	5	1
5		3	4	4	0	1	5	1	2	2	1	4	2	2	0	
2	0	2	3	5	1	2	4	5	3	4	4	0	1	5	1	
	4	2	1	5	1	2	4	5	3	4	4	0	1	5	1	
3	4	0	2	5	1	2	4	5	3	4	4	0	1	5	1	
	2	3	2	0	2	3	1	3	2	1	2	4	2	0	2	
4	2	0	3	5	1	2	4	5	3	4	4	0	1	5	1	
	1	3	4	0	2	3	1	3	2	1	2	4	2	0	2	
5	2	3	2	0	2	3	1	1	2	3	2	4	2	0	2	
	2	0	3	5	1	2	4	5	3	4	4	0	1	5	1	
6	4	2	1	5	1	2	4	5	3	4	4	0	1	5	1	
	3	3	4	0	2	3	1	3	2	1	2	4	2	0	2	
7	2	3	4	0	2	1	2	3	2	1	2	2	4	0	2	
	4	3	2	2	1	3	1	3	2	2	1	4	2	2	0	
8	2	3	2	1	2	2	4	0	2	4	4	0	1	5	1	
	1	3	2	2	1	4	2	2	0	1	2	4	2	0	2	
9	4	3	2	2	5	3	1	3	2	2	1	4	2	2	0	
	2	3	2	1	2	2	4	0	2	4	4	0	1	5	1	
10	5	3	4	4	0	1	5	1	2	2	1	4	2	2	0	
	3	2	1	2	4	2	0	2	2	0	4	0	1	5	1	

**Лабораторная работа №3**  
**ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ**  
**ВЫБОРКИ СТАНДАРТНЫМИ СРЕДСТВАМИ ЭТ MS EXCEL**

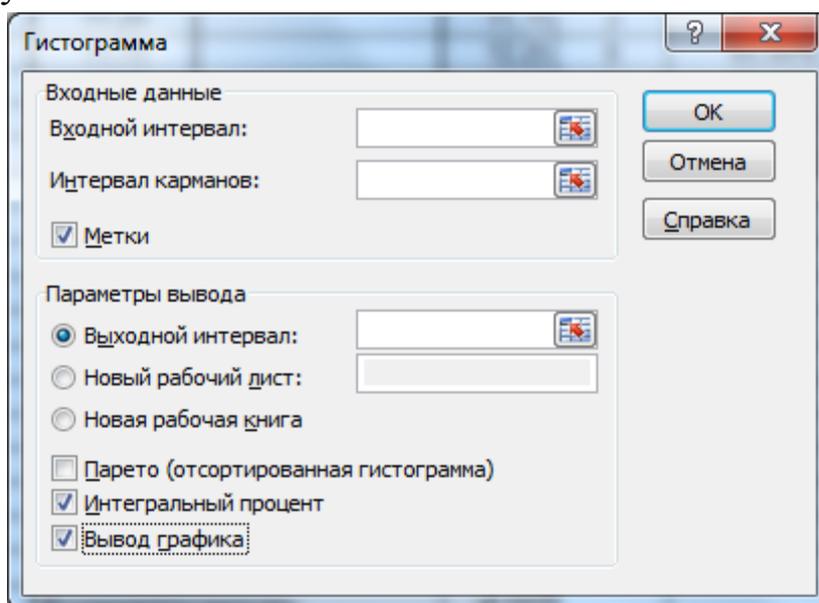
**Цель работы:** овладеть навыками графического представления выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

**Краткая теория**

В среде ЭТ MS Excel построение гистограммы анализируемой выборки можно выполнить с помощью Надстройки Пакет анализа. Для этого необходимо выполнить команду главного меню Данные → Анализ данных, и в появившемся диалоговом окне Анализ данных выбрать режим Гистограмма.



Появится окно Гистограмма, в котором необходимо задать следующие установки:



*Входной интервал:* – адреса ячеек, содержащие значения выборки.

*Интервал карманов:* (необязательный параметр) – адреса ячеек, содержащие границы интервалов (карманы). Эти значения должны быть введены в возрастающем порядке.

*Метки* – флажок, включаемый, если первая строка во входных данных содержит заголовки. Если заголовки отсутствуют, то флажок следует выключить.

*Выходной интервал:* / *Новый рабочий лист:* / *Новая рабочая книга.* Включенная радиокнопка *Выходной интервал* требует ввода адреса верхней ячейки, начиная с которой будут размещаться вычисленные относительные частоты  $\omega_j$ . В положении радиокнопки *Новый рабочий лист:* открывается новый лист, в котором начиная с ячейки A1 размещаются частности  $\omega_j$ . В положении радиокнопки *Новая рабочая книга* открывается новая книга, на первом листе которой начиная с ячейки A1 размещаются частности  $\omega_j$ .

*Парето (отсортированная гистограмма)* – устанавливается в активное состояние, чтобы представить  $\omega_j$  в порядке их убывания. Если параметр выключен, то  $\omega_j$  приводятся в порядке следования интервалов.

*Интегральный процент* – устанавливается в активное состояние для расчета выраженных в процентах накопленных относительных частот (процентный аналог значений выборочной функции распределения).

*Вывод графика* – устанавливается в активное состояние для автоматического создания встроенной диаграммы на листе, содержащем относительные частоты  $\omega_j$ .

*Замечание.* При использовании надстройки *Анализ данных* необходимо помнить: если границы интервалов не заданы, то автоматически будет создан набор интервалов одинаковой длиной

$$\Delta = \frac{x_{max} - x_{min}}{[N] - 1},$$

где  $[N]$  – целая часть величины  $N = 1 + 3.322 \cdot \lg n$ ,  
 $n$  – объем выборки.

### Пример выполнения

**Постановка задачи.** Исследуются значения питающего напряжения некоторого устройства. Замер осуществляется с интервалом в 1 час. Выполнено 21 измерение напряжения в вольтах, в результате которых получена следующая выборка значений величины  $U(B)$  (табл. 1).

12.2	12.3	12.9	13.0	13.2	11.8	11.9
11.3	11.5	10.8	11.1	12.0	12.1	12.1
11.6	12.0	11.9	12.5	12.6	12.5	11.9

**Требуется:** построить гистограмму выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MS Excel необходимо выполнить следующие действия:

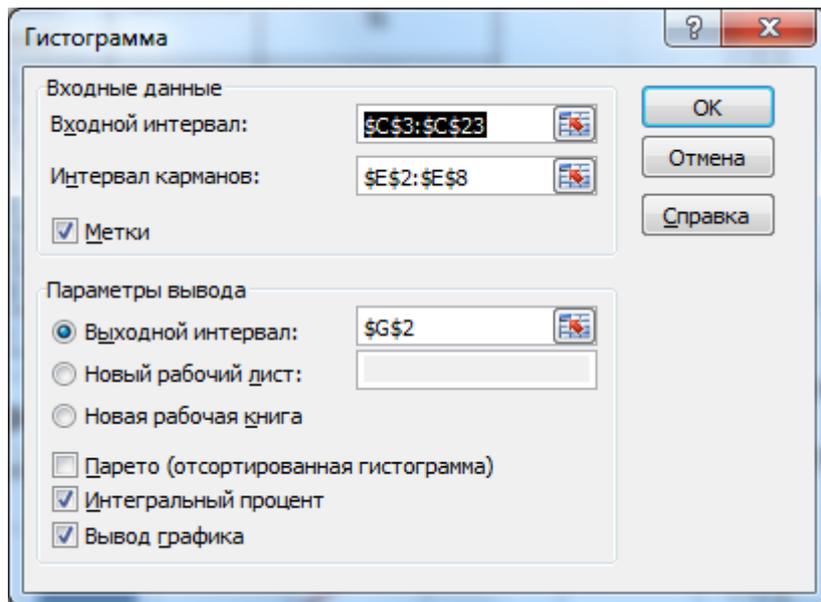
1. Идентифицируйте свою работу, переименовав Лист1 в Титульный лист и записав номер лабораторной работы, ее название, кто выполнил и проверил.

2. Переименуйте Лист 2 в Исходные данные и наберите столбец исходных данных.

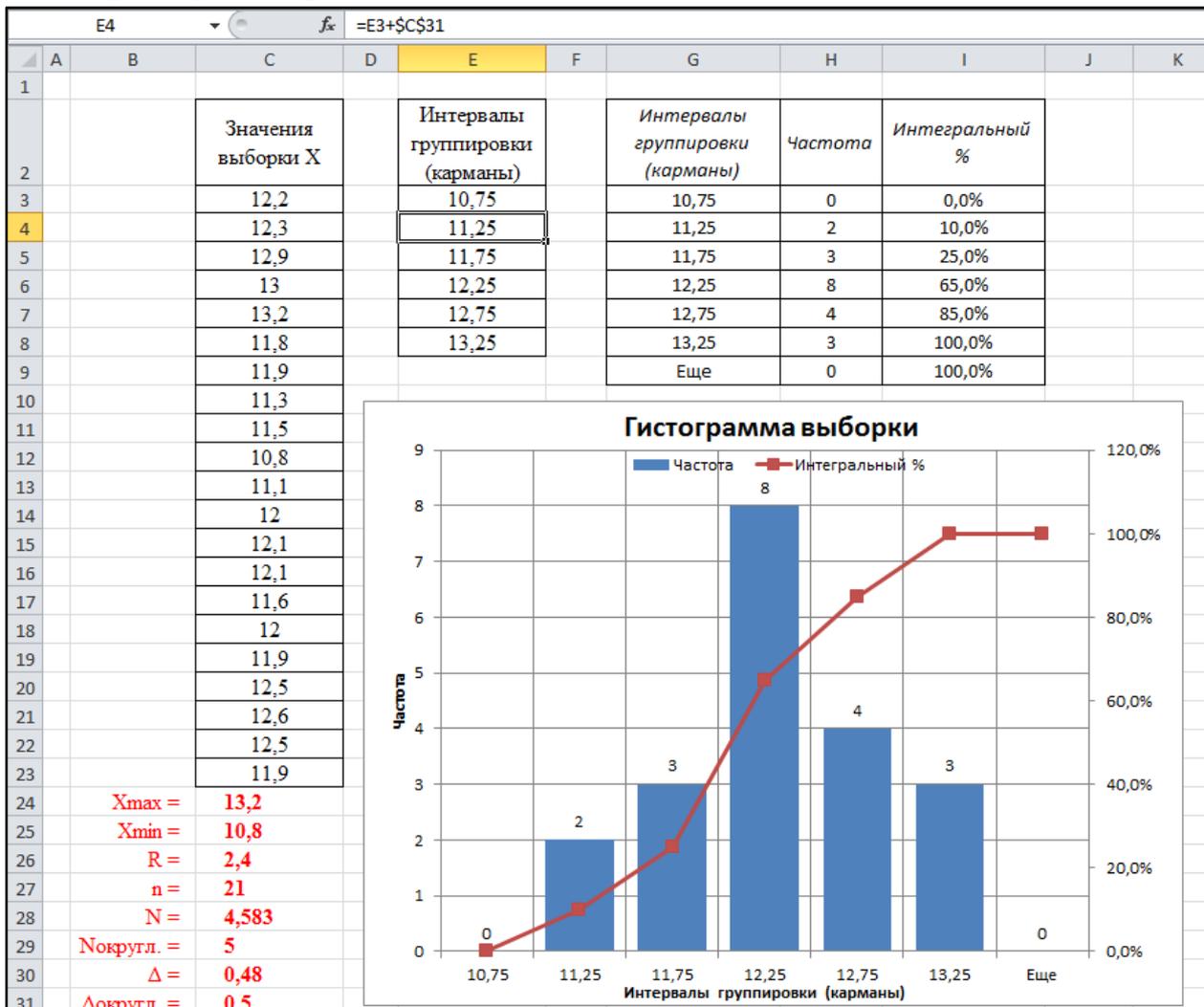
3. Вычислите величины  $U_{max}$ ,  $U_{min}$ ,  $R$ ,  $n$ ,  $N$ ,  $N_{округл.}$ ,  $\Delta$  и  $\Delta_{округл.}$ , используя встроенные функции Excel **МАКС**, **МИН**, **СЧЕТ**, **КОРЕНЬ** и **ОКРУГЛ**.

4. Сформируйте столбец интервалов группировки (карманов). Если диапазон карманов не введен, то набор отрезков, равномерно распределенных между максимальным и минимальным значениями исходных данных, будут созданы Excel автоматически.

5. Наберите команду Данные → Анализ данных → Гистограмма и в появившемся диалоговом окне выполните нужные установки.



6. Отформатируйте полученную таблицу и построенную гистограмму выборки.



7. Сделайте выводы и сохраните работу в вашем каталоге.

## Исходные данные для самостоятельного решения

**Задание.** Имеется выборка объема  $n = 26$  (табл. 2).

**Требуется:** построить гистограмму выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

Таблица 2.

№ варианта	Выборка													
1	3,13	9,56	7,03	9,18	7,85	5,59	6,85	11,3	7,90	6,00	6,68	5,66	8,64	
	8,87	4,58	11,3	5,02	4,33	9,31	10,3	5,89	6,98	5,83	8,75	7,73	9,16	
2	3,48	7,87	4,04	8,81	4,08	3,46	4,37	6,66	1,46	5,59	3,78	8,73	5,57	
	8,62	3,25	3,38	4,20	2,49	6,11	4,54	6,53	5,20	3,84	5,35	9,72	4,63	
3	4,21	5,68	3,45	6,79	3,39	2,99	3,88	3,77	1,43	5,96	4,94	6,65	5,92	
	4,20	4,25	5,64	5,58	5,87	5,05	3,55	7,95	4,45	5,85	6,68	1,24	7,09	
4	11,7	9,83	5,49	7,43	9,92	3,41	6,83	8,22	8,30	8,14	9,29	9,47	7,83	
	7,61	3,46	7,72	12,1	6,06	10,6	6,76	8,81	9,86	8,13	9,04	4,75	9,33	
5	4,89	9,25	7,94	9,10	6,37	6,77	3,47	8,84	6,48	4,92	6,98	10,1	6,32	
	6,36	5,16	7,92	12,0	7,46	7,01	13,0	7,34	6,71	5,48	9,95	11,9	8,89	
6	6,13	8,56	9,77	9,17	8,89	6,19	7,70	6,96	6,72	6,08	4,41	5,52	9,59	
	9,42	6,22	4,86	6,33	6,48	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98	
7	6,82	9,27	7,91	5,77	8,02	3,07	2,22	5,76	11,6	6,62	7,07	12,5	1,65	
	10,5	3,67	7,62	4,94	5,39	3,64	4,62	8,88	6,75	5,77	6,38	10,3	5,74	
8	8,48	9,56	6,06	5,85	6,78	5,60	10,8	7,70	6,44	8,64	6,95	5,66	4,84	
	4,98	4,62	5,57	6,47	5,97	8,42	3,66	9,24	4,13	6,58	7,41	5,67	7,89	
9	10,2	9,23	8,77	10,4	9,44	9,69	6,30	9,42	6,62	9,69	8,59	8,68	7,97	
	8,64	6,45	5,29	5,00	8,42	8,84	8,26	6,46	6,96	6,51	6,72	6,00	5,36	
10	7,83	9,62	9,77	9,17	8,89	6,19	7,71	6,96	6,72	6,08	4,41	5,52	9,59	
	8,66	6,26	4,86	6,33	6,28	8,60	7,38	7,84	7,24	6,85	6,50	8,28	4,98	

## Лабораторная работа №4

### РАСЧЕТ ЧИСЛОВЫХ ХАРАКТЕРИСТИК ВЫБОРКИ С ПОМОЩЬЮ ВСТРОЕННЫХ ФУНКЦИЙ ЭТ MSEXCEL

**Цель работы:** овладеть навыками графического представления выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

#### Краткая теория

Первый шаг к осмыслению скрытых в выборке закономерностей – это ее графическое представление, то есть построение гистограммы, полигона частот и эмпирической функции распределения. Однако выборки, имеющие похожие графические изображения, могут различаться своими числовыми характеристиками. Числовые характеристики вариационных рядов вычисляют по данным, полученным в результате наблюдений (статистическим данным), поэтому их называют также статистическими характеристиками или оценками.

Выборочные характеристики являются оценками соответствующих характеристик генеральной совокупности. Эти оценки должны удовлетворять определенным требованиям. В соответствии с важнейшими требованиями, оценки должны быть:

- *несмещенными*, то есть стремиться к истинному значению характеристики генеральной совокупности при неограниченном увеличении количества испытаний;
- *состоятельными*, то есть с ростом размера выборки оценка должна стремиться к значению соответствующего параметра генеральной совокупности с вероятностью, приближающейся к 1;
- *эффективными*, то есть для выборок равного объема используемая оценка должна иметь минимальную дисперсию.

Выборка может характеризоваться следующими числовыми характеристиками.

#### 1. *Характеристики положения.*

Самой известной и наиболее употребляемой характеристикой любого вариационного ряда является его средняя арифметическая, называемая также *выборочным средним*. Средняя арифметическая характеризует значения признака, вокруг которого концентрируются наблюдения, т.е. центральную тенденцию распределения. При

статистическом анализе выборки, кроме средней арифметической, широко применяют структурные, или порядковые, средние, к которым относятся *медиана* и *мода*.

Выборочное среднее рассчитывается по формуле

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Если же анализируемые данные представлены в виде вариационного ряда, то для вычисления выборочного среднего применяется одно из следующих соотношений:

- для дискретного вариационного ряда

$$\bar{x} = \frac{\sum_{i=1}^N x^{(i)} \cdot n_i}{n} = \sum_{i=1}^N x^{(i)} \cdot \omega_i;$$

- для интервального вариационного ряда

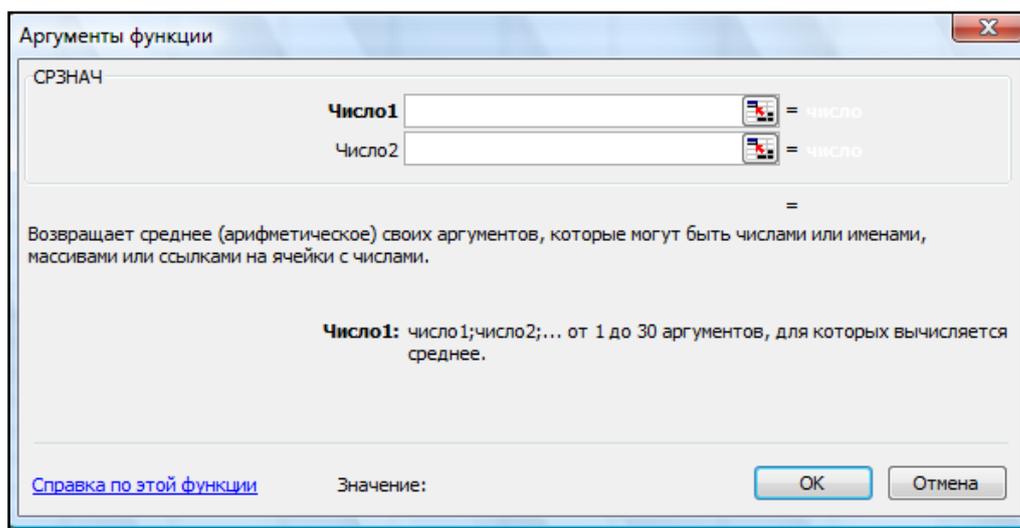
$$\bar{x} = \frac{\sum_{i=1}^N x_i^* \cdot n_i}{n} = \sum_{i=1}^m \omega_i \cdot x_i^*,$$

где  $\omega_i$  – частность (относительная частота), соответствующая  $i$ -й variante или  $i$ -му частичному интервалу;  $x_i^*$  – середина  $i$ -го частичного интервала группировки.

В программе MS Excel среднее значение находится с помощью функцией *СРЗНАЧ*.

Синтаксис функции:

**СРЗНАЧ(число1; число2; ...).**



Число1, число2, .. – это от 1 до 30 аргументов, для которых вычисляется среднее.

Аргументы должны быть либо числами, либо именами, массивами или ссылками, содержащими числа.

Достоинство *медианы* как меры центральной тенденции заключается в том, что на нее не влияет изменение крайних членов вариационного ряда, если любой из них, меньший медианы, остается меньше ее, а любой, больший медианы, продолжает быть большее ее. Медиана предпочтительнее средней арифметической для ряда, у которого крайние варианты по сравнению с остальными оказались чрезмерно большими или малыми. Особенность *моды* как меры центральной тенденции заключается в том, что она также не изменяется при изменении крайних членов ряда, т.е. обладает определенной устойчивостью к вариации признака.

Выборочная медиана разбивает выборку пополам: слева и справа от нее оказывается одинаковое число элементов выборки. Если число элементов выборки четно,  $n = 2k$ , то выборочную медиану определяют по формуле

$$Me = (x_k + x_{k+1})/2,$$

где  $x_k$  и  $x_{k+1}$  –  $k$ -е и  $(k + 1)$  -е выборочные значения из вариационного ряда.

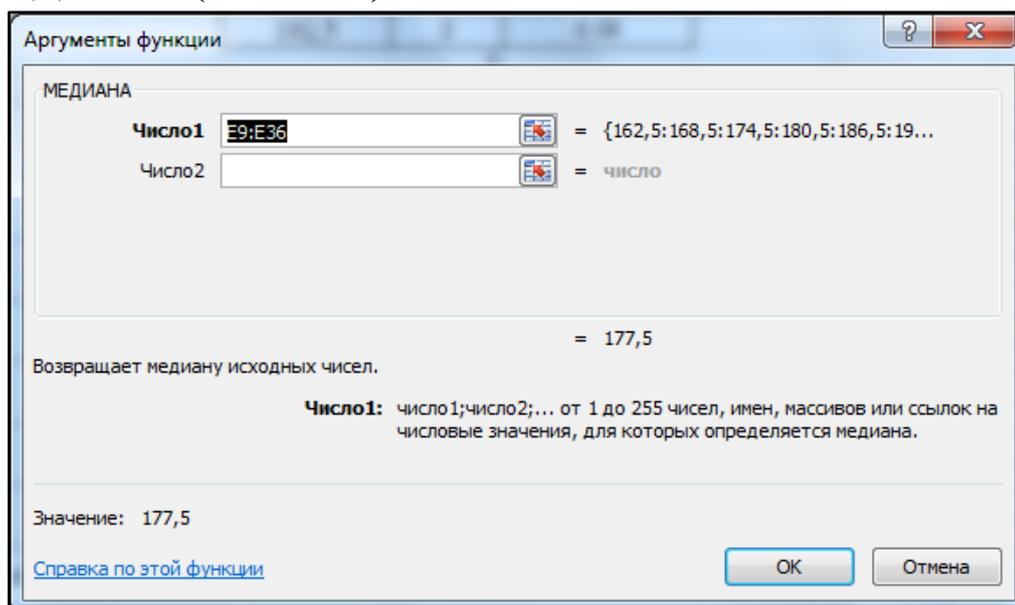
При нечетном  $n = 2k + 1$  объеме выборки медиану находят по формуле

$$Me = x_{k+1},$$

т.е. за значение медианы принимают величину  $x_{k+1}$ .

Синтаксис функции:

## МЕДИАНА (диапазон).

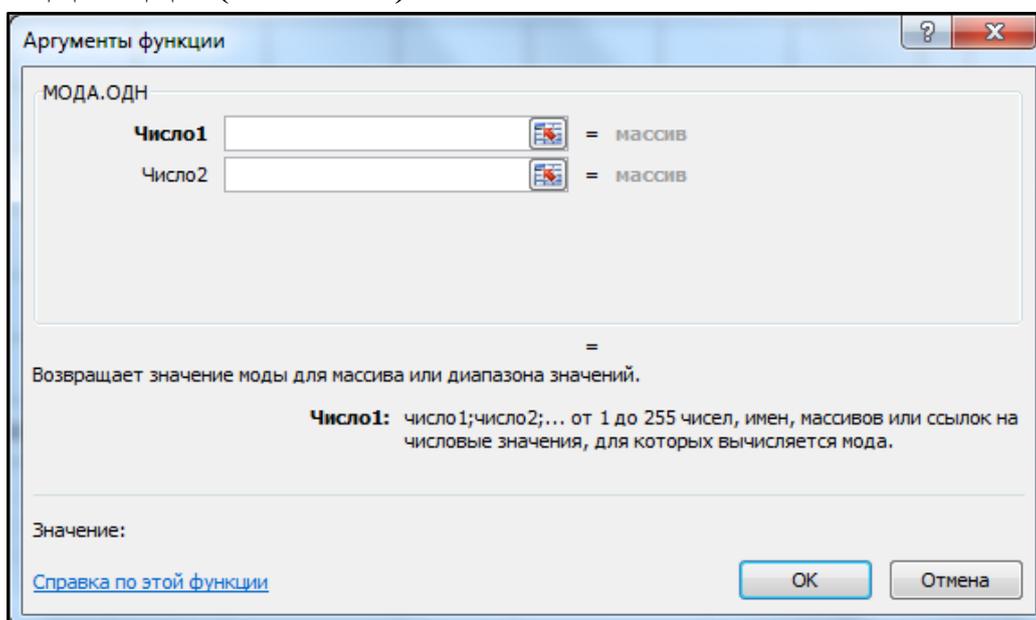


Так, например, если в диапазоне записаны значения 1, 2, 3, 4, 5, то функция МЕДИАНА вернет значение, равное 3, а если диапазон 1, 2, 3, 4,, то найденное значение 2,5.

Мода  $M_o$  используется для нахождения наиболее часто встречающегося в выборке значения.

Синтаксис функции:

## МОДА.ОДН (диапазон).



При поиске игнорируются пустые ячейки, текстовые и логические значения.

Если использовать функцию для нахождения  $M_o$  выборки 1, 2, 3, 4, 4, то функция даст 4. Если значения в выборке не повторяются, то

функция выдаст сообщение об ошибке #Н/Д.

## 2. Характеристики рассеяния.

Для получения полного представления о вариационном ряде (определив центральную тенденцию распределения с помощью характеристик положения) далее оценивают рассеяние (вариацию, изменчивость) исследуемого признака вокруг этих величин. Простейшим и, весьма приближенным показателем вариации (изменчивости), является вариационный *размах*. Размах вариации наиболее полезен, если нужен быстрый и общий взгляд на изменчивость при сравнении большого количества выборок. *Размах* выборки вычисляется по формуле

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

Но наибольший интерес представляют меры вариации (рассеяния) наблюдений вокруг средних величин, в частности, вокруг средней арифметической. К таким оценкам относятся *выборочная дисперсия* и *среднее квадратичное отклонение*.

Дисперсия выборки – это параметр, характеризующий степень разброса элементов выборки относительно среднего значения  $\bar{x}$ . Чем больше дисперсия, тем дальше отклоняются значения элементов выборки от среднего значения.

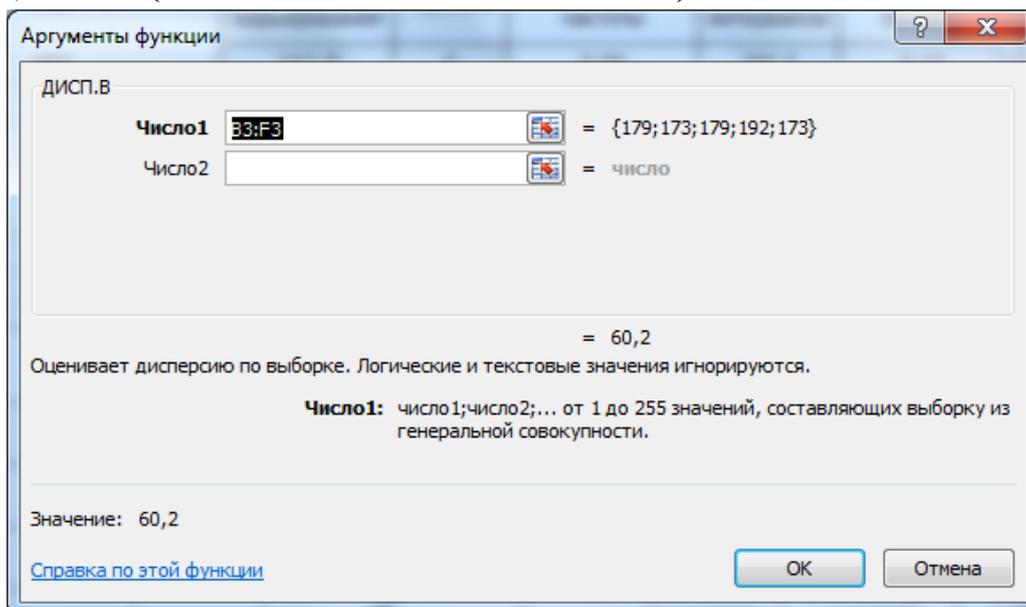
Выборочная дисперсия находится по формуле

$$D = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}.$$

Для вычисления выборочной дисперсии с помощью ЭТ MSExcel используется функция *ДИСП.В*.

Синтаксис функции:

### **ДИСП.В(число1; число2; ...; число255).**

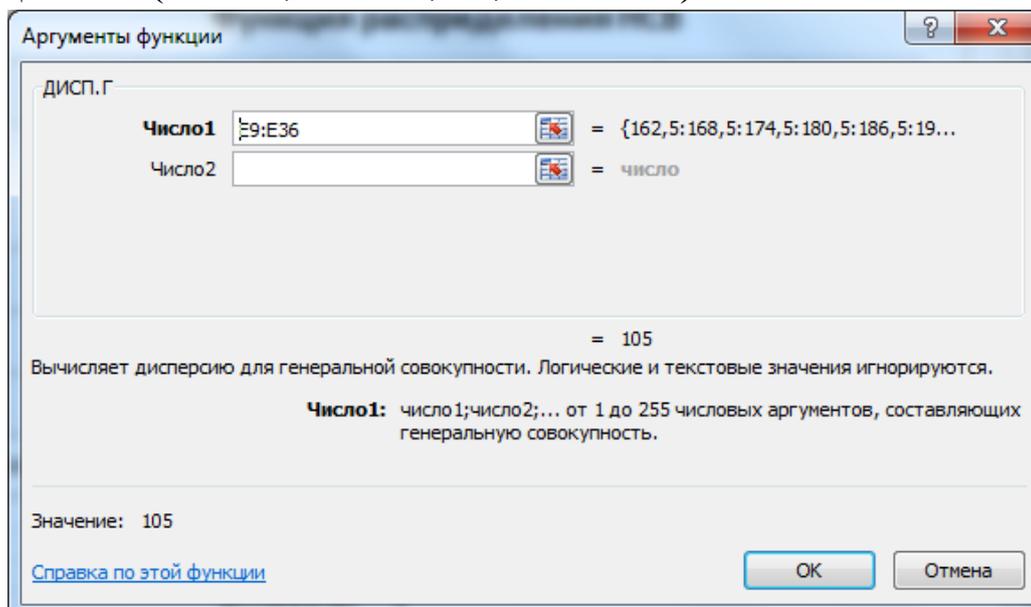


где *число1*; *число2*; ...; *число255*– числа или адреса ячеек, содержащих числовые данные. Ячейки, содержащие текстовые, логические данные или пустые, при вычислении выборочной дисперсии игнорируются.

Для вычисления дисперсии генеральной совокупности в ЭТ Excel используется функция *ДИСП.Г*.

Синтаксис функции:

### **ДИСП.Г(число1; число2; ...; число255).**



Если данные представлены в виде вариационного ряда, то целесообразно для вычисления  $D$  вместо приведенной выше формулы использовать соотношения:

- для дискретного вариационного ряда

$$D = \frac{\sum_{i=1}^N (x^{(i)} - \bar{x})^2 \cdot n_i}{n} = \sum_{i=1}^N (x^{(i)} - \bar{x})^2 \cdot \omega_i ;$$

- для интервального вариационного ряда

$$D = \frac{\sum_{i=1}^N (x_i^* - \bar{x})^2 \cdot n_i}{n} = \sum_{i=1}^N (x_i^* - \bar{x})^2 \cdot \omega_i .$$

Выборочная дисперсия обладает одним существенным недостатком: если среднее арифметическое выражается в тех же единицах, что и значения случайной величины, то, согласно определению, дисперсия выражается уже в квадратных единицах. Этого недостатка можно избежать, если использовать в качестве меры вариации признака *среднее квадратичное отклонение*  $S$

$$S = \sqrt{D} .$$

При малых объемах выборки дисперсия является смещенной оценкой, поэтому при объемах  $n \leq 30$  используют *исправленную дисперсию* и *исправленное среднее квадратичное отклонение*.

Среднее квадратичное отклонение  $S$ , полученное при выборке  $n < 30$ , носит название смещенного и его среднее значение занижено по сравнению со средним квадратичным отклонением для генеральной совокупности.

При числе испытаний  $n < 30$

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} .$$

При числе испытаний  $30 \leq n < 50$

$$S_1 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} ,$$

$$S = M_k \cdot S_1 ,$$

где  $M_k$  – коэффициент, зависящий от числа испытаний.

Значения  $M_k$  приведены в табл. 1 для  $K = n - 1$

Таблица 1

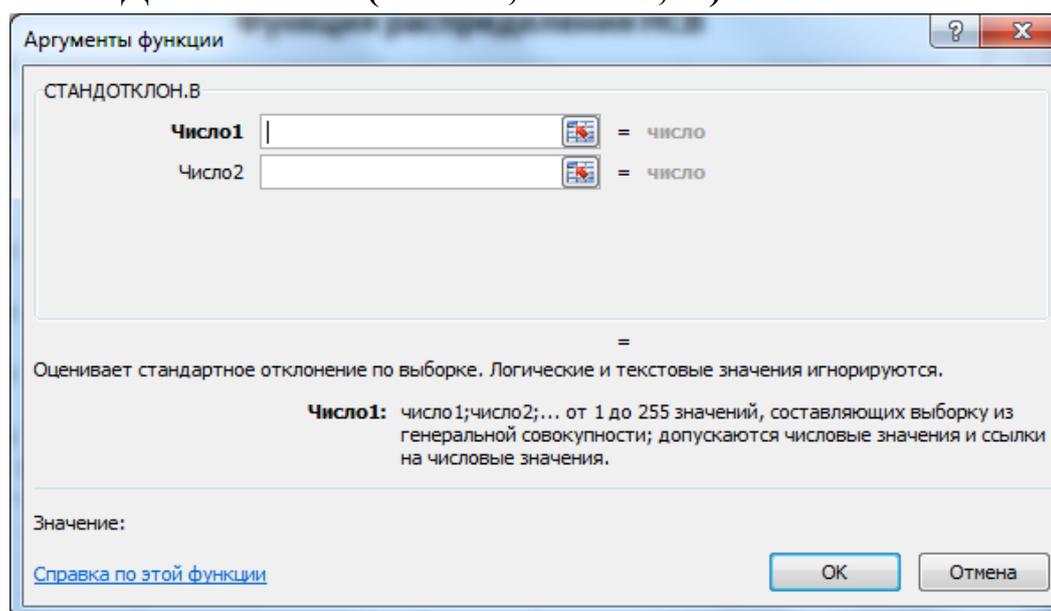
К	2	3	4	9	19	30	50	60
$M_k$	1,128	1,085	1,064	1,028	1,013	1,008	1,005	1,004

При  $n > 60$  значение коэффициента  $M_k \sim 1$ .

Для вычислений среднего квадратичного отклонения выборки применяется функция *СТАНДОТКЛОН.В*.

Синтаксис функции:

**СТАНДОТКЛОН.В(число1; число2; ...).**

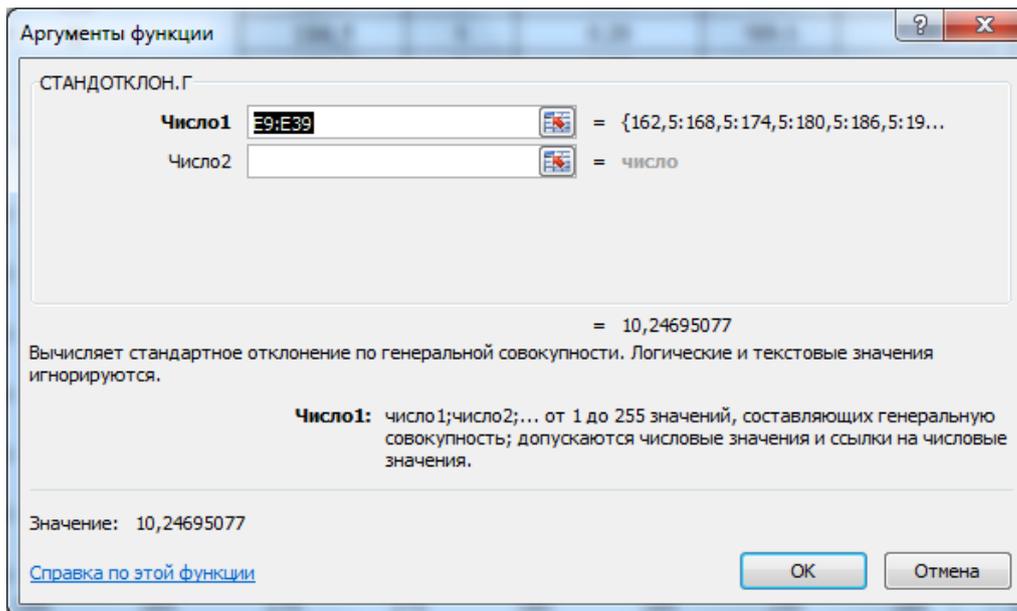


Число1, число2,... –от 1 до 255 числовых аргументов, соответствующих выборке из генеральной совокупности. Вместо аргументов, разделенных точкой с запятой, можно также использовать массив или ссылку на массив.

Функция *СТАНДОТКЛОН.В* оценивает среднее квадратичное отклонение (стандартное отклонение) по выборке. Стандартное отклонение – это мера того, насколько широко разбросаны точки данных относительно их среднего.

*СТАНДОТКЛОН.В* предполагает, что аргументы являются только выборкой из генеральной совокупности.

Если данные представляют всю генеральную совокупность, то стандартное отклонение следует вычислять с помощью функции *СТАНДОТКЛОН.Г*.

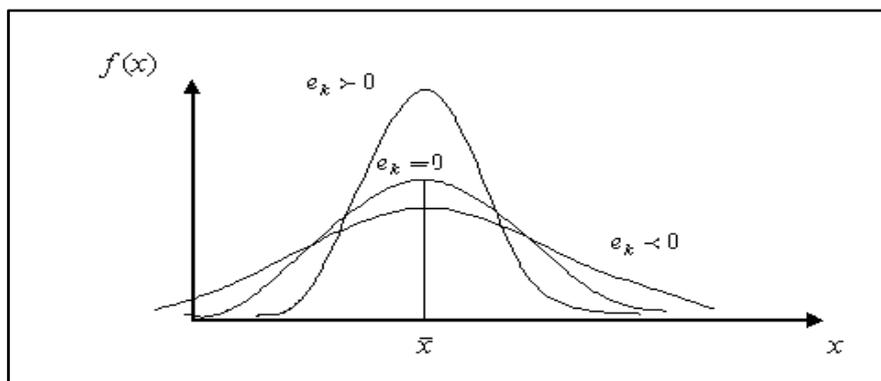


Стандартное отклонение вычисляется с использованием «несмещенного» или « $n - 1$ » метода.

### 3. Характеристики формы.

К характеристикам формы относят коэффициент асимметрии и эксцесс. *Выборочный эксцесс* характеризует островершинность эмпирического распределения относительно стандартного нормального. Эксцесс стандартного нормального распределения равен трем. Если эксцесс положителен ( $e_k > 0$ ), то полигон вариационного ряда имеет более крутую вершину. Это говорит о скоплении значений признака в центральной зоне ряда распределения, т.е. о преимущественном появлении в данных значений, близких к средней величине.

Если эксцесс отрицателен ( $e_k < 0$ ), то полигон имеет более пологую вершину по сравнению с нормальной кривой. Это означает, что значения признака не концентрируются в центральной части ряда, а достаточно равномерно рассеяны по всему диапазону от минимального до максимального значения. Чем больше абсолютная величина эксцесса, тем существеннее распределение отличается от нормального, смотри рисунок.



Выборочный эксцесс может быть найден по формуле

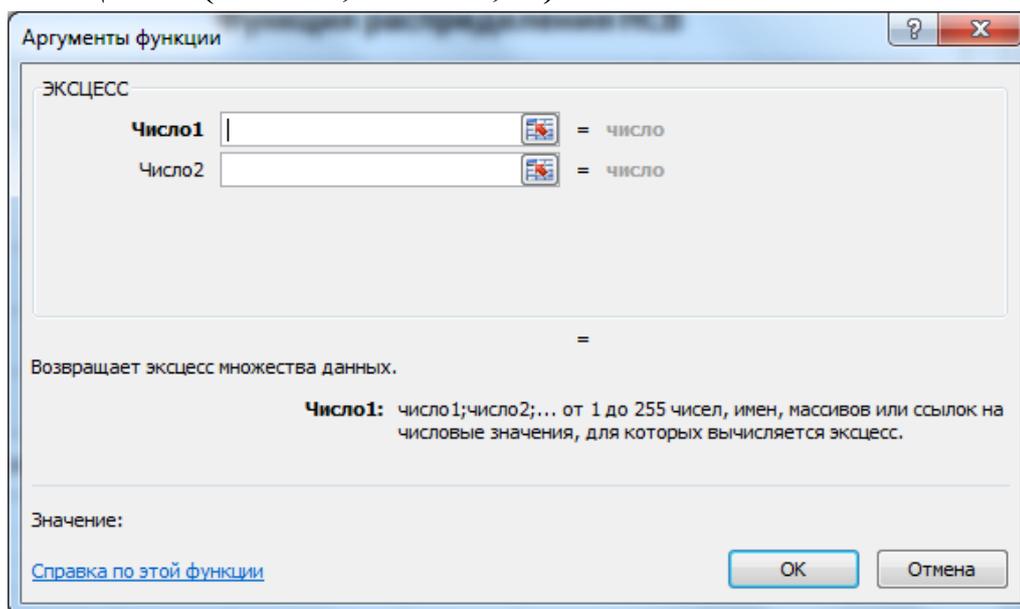
$$e_k = \frac{\mu_4}{S^4} - 3,$$

где  $\mu_4 = \sum_{i=1}^n (x_i - \bar{x})^4$ .

Для вычислений выборочного эксцесса выборки применяется функция MS Excel ЭКСЦЕСС.

Синтаксис функции:

**ЭКСЦЕСС (число1; число2; ...).**

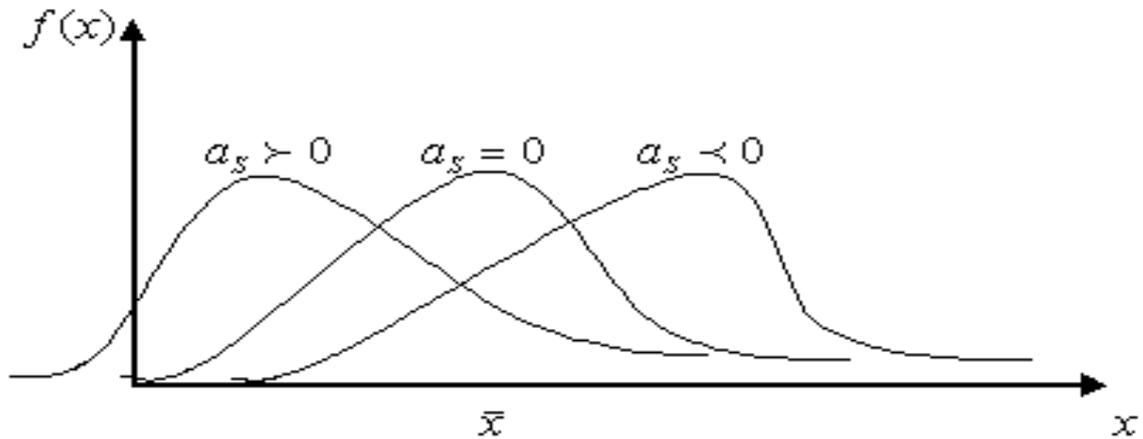


*Коэффициент асимметрии* характеризует симметрию распределения выборочных данных около центра выборки  $\bar{x}$ , для стандартного нормального распределения коэффициент асимметрии равен 0 ( $a_3=0$ ).

Если распределение асимметрично, одна из ветвей построенного полигона частот имеет более пологий спуск, чем другая.

Если правая ветвь графика более пологая то это означает

преимущественное появление в распределении более высоких значений признака, при этом коэффициент асимметрии  $a_s > 0$ . В противном случае  $a_s < 0$ , при этом в распределении чаще встречаются более низкие значения (смотри рисунок).



Чем больше значение коэффициента асимметрии, тем более асимметрично распределение (до 0,25 асимметрия незначительная; от 0,25 до 0,5 умеренная; свыше 0,5 – существенная).

Коэффициент асимметрии вычисляется по формуле

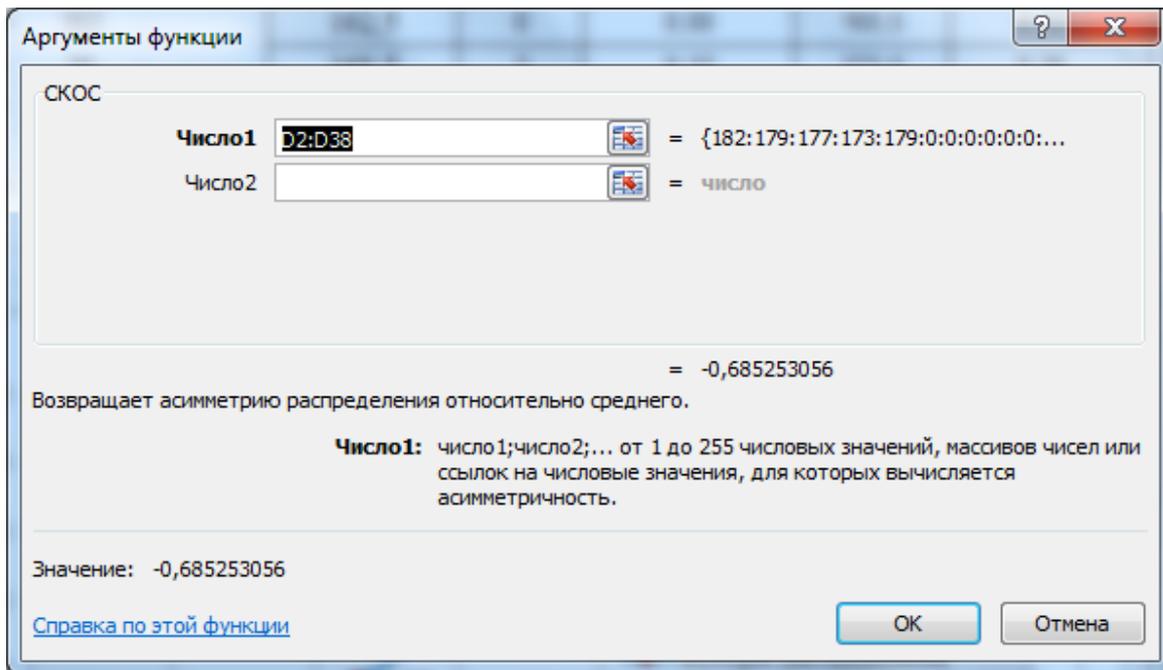
$$a_s = \frac{\mu_3}{S^3},$$

где  $\mu_3 = \sum_{i=1}^n (x_i - \bar{x})^3$ .

Для вычисления коэффициента асимметрии выборки применяется функция *СКОС*.

Синтаксис функции:

**СКОС(число1; число2; ...).**



### Пример выполнения

**Постановка задачи.** Приведены размеры месячных зарплат (в тыс. руб.) 27швей-мотористок, работающих по сдельно-премиальной системе оплаты труда (табл. 1).

Таблица 1.

18,8	20,2	19,3	19,9	23,2	22,5	17,4	21,8	19,2
19,4	18,7	16,3	18,4	19,3	18,8	19,4	18,7	16,3
20,5	20,6	19,4	18,7	16,3	18,4	19,3	18,8	17,8

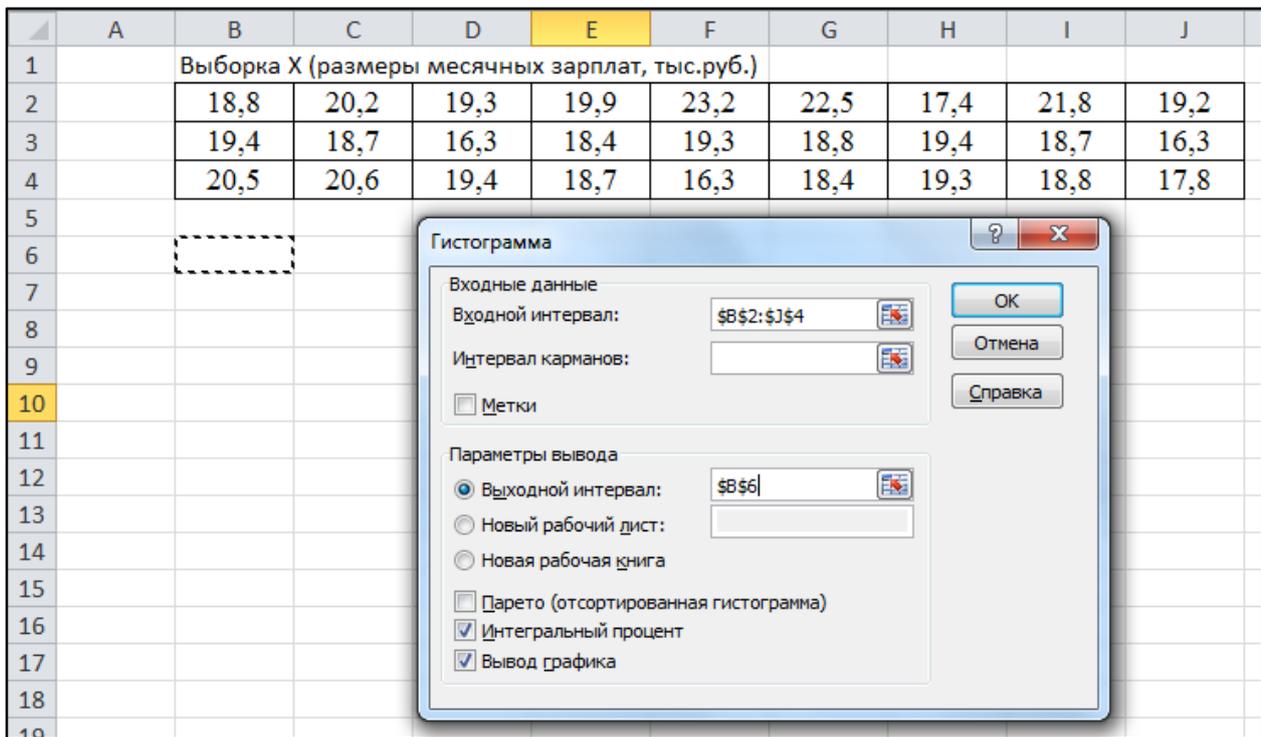
**Требуется:** найти числовые характеристики выборки с помощью встроенных функций ЭТ MS Excel.

**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MS Excel необходимо выполнить следующие действия:

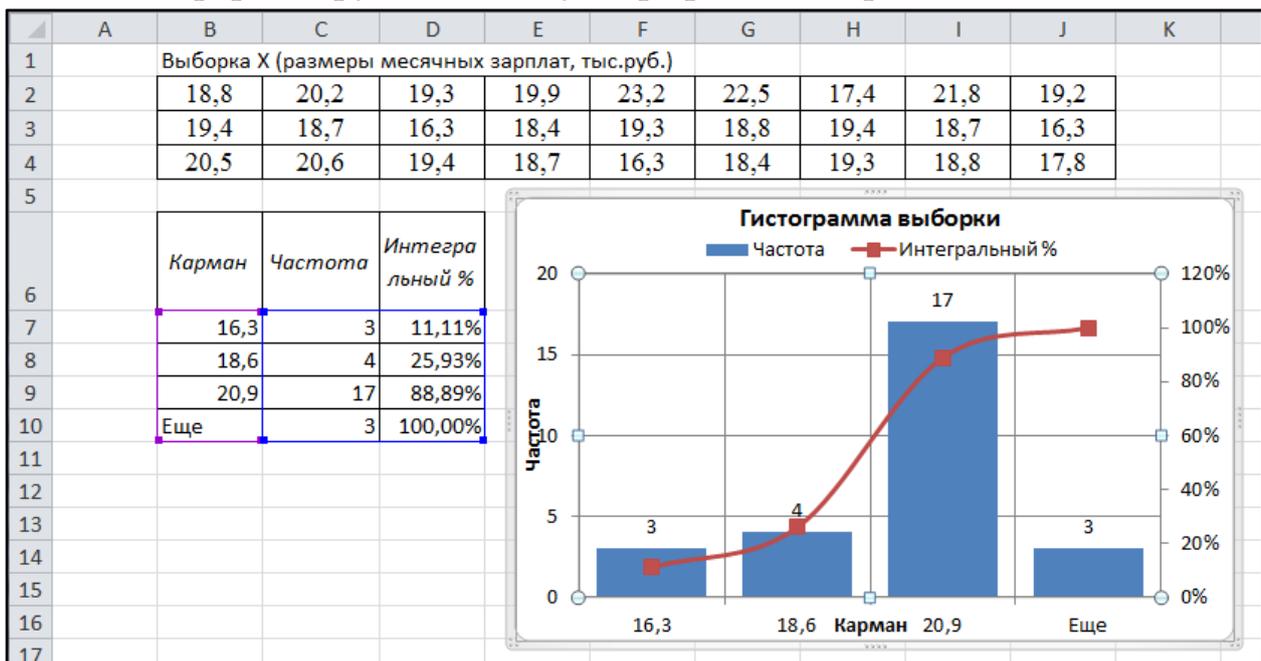
1. Идентифицируйте свою работу, переименовав Лист1 в Титульный лист и записав номер лабораторной работы, ее название, кто выполнил и проверил.

2. Переименуйте Лист 2 в Исходные данные и наберите таблицу исходных данных.

3. Постройте гистограмму выборки с помощью Надстройки Пакет анализа, не указывая интервалы группировки выборки.



#### 4. Отформатируйте таблицу и график гистограммы.



5. Найдите числовые характеристики выборки с помощью встроенных функций ЭТ MS Excel, рассмотренных выше.

18	1. Характеристики положения:				
19		Выборочное среднее	<b>19,163</b>		
20		Медиана	<b>19,200</b>		
21		Мода	<b>18,800</b>		
22	2. Характеристики рассеяния:				
23		Наибольшее	<b>23,2</b>		
24		Наименьшее	<b>16,3</b>		
25		Размах выборки	<b>6,9</b>		
26		Дисперсия выборки		<b>2,763</b>	
27		Дисперсия ген.совокупности		<b>2,661</b>	
28		Среднее квадратич. отклонение		<b>1,662</b>	
29		СКО генеральной совокупности		<b>1,631</b>	
30	3. Характеристики формы:				
31		Выборочный эксцесс	<b>0,8245</b>		
32		Кoeffиц. асимметрии	<b>0,4384</b>		
33					

6. Сделайте выводы и сохраните работу в вашем каталоге.

### Исходные данные для самостоятельного решения

**Задание.** Имеется выборка объема  $n = 24$  (табл. 2).

**Требуется:** найти числовые характеристики выборки с помощью встроенных функций ЭТ MS Excel.

Таблица 2.

№ варианта	Выборка							
	1	18,8	20,2	19,3	19,7	23,2	22,5	17,4
19,4		18,7	16,3	18,4	19,3	18,8	19,4	18,7
20,5		20,7	19,4	18,7	16,3	18,4	19,3	18,8
2	19,9	23,2	22,5	17,4	21,8	18,8	20,2	19,3
	18,4	19,3	18,7	19,4	18,7	19,4	18,7	16,3
	18,7	16,3	18,4	19,3	18,8	20,5	20,6	19,4
3	19,3	18,7	16,3	18,4	19,3	18,8	19,4	18,7
	20,7	20,6	19,4	18,7	16,3	18,7	19,3	18,8
	19,9	23,2	22,7	17,4	21,8	18,8	20,2	19,3
4	16,3	18,4	19,3	18,3	19,4	18,7	21,8	18,8
	19,7	17,7	16,3	18,4	19,3	18,8	18,7	19,4
	22,5	17,4	21,8	17,7	20,2	19,3	18,8	20,5
5	16,5	18,7	19,3	18,8	19,4	18,7	19,3	18,8
	18,8	16,3	18,4	19,3	18,8	20,5	20,6	19,4
	19,7	18,7	16,3	18,4	19,3	18,8	19,4	18,7

6	16,3	20,6	19,4	18,7	16,3	18,7	19,3	18,8
	21,8	23,2	22,7	17,4	21,8	18,8	20,2	19,3
	19,4	18,4	19,3	18,1	19,4	19,7	21,8	18,8
7	19,3	18,7	16,7	17,4	19,3	18,8	18,7	19,4
	20,2	17,4	21,8	18,8	20,2	19,3	17,8	20,5
	19,4	18,7	19,3	18,3	19,4	18,7	19,3	18,8
8	19,7	18,7	16,3	18,7	19,3	18,8	19,3	18,8
	16,3	17,4	21,8	18,8	20,2	19,3	18,4	19,3
	21,8	18,1	19,4	19,7	21,8	18,8	18,7	16,3
9	19,4	17,4	19,3	18,8	18,7	19,4	17,4	21,8
	19,3	18,8	20,2	19,3	17,8	20,5	18,1	19,4
	20,2	18,3	19,4	18,7	19,3	18,8	17,4	19,3
10	19,3	20,6	19,4	18,7	16,3	18,7	18,8	20,2
	20,2	23,2	22,7	17,4	21,8	18,8	18,3	19,4
	19,4	18,4	19,3	18,1	19,4	19,7	19,7	21,8

## Лабораторная работа №5

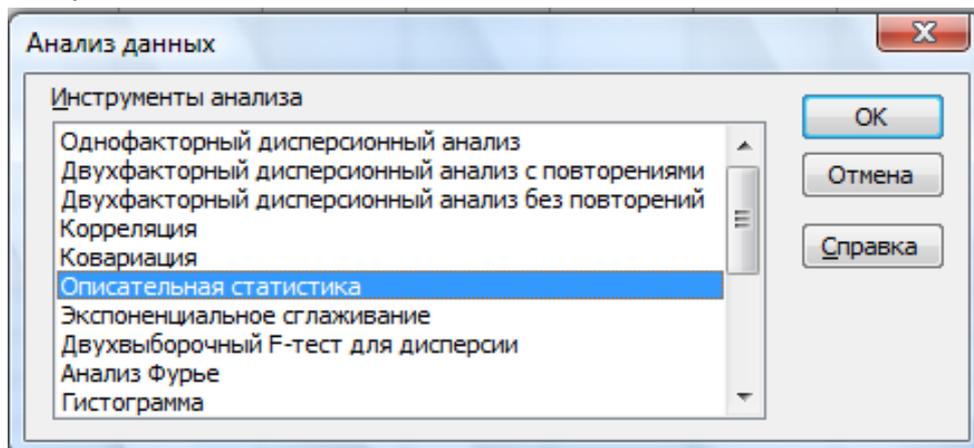
### НАХОЖДЕНИЕ ЧИСЛОВЫХ ХАРАКТЕРИСТИК ВЫБОРКИ СТАНДАРТНЫМИ СРЕДСТВАМИ ЭТ MS EXCEL

**Цель работы:** овладеть навыками расчета числовых характеристик выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

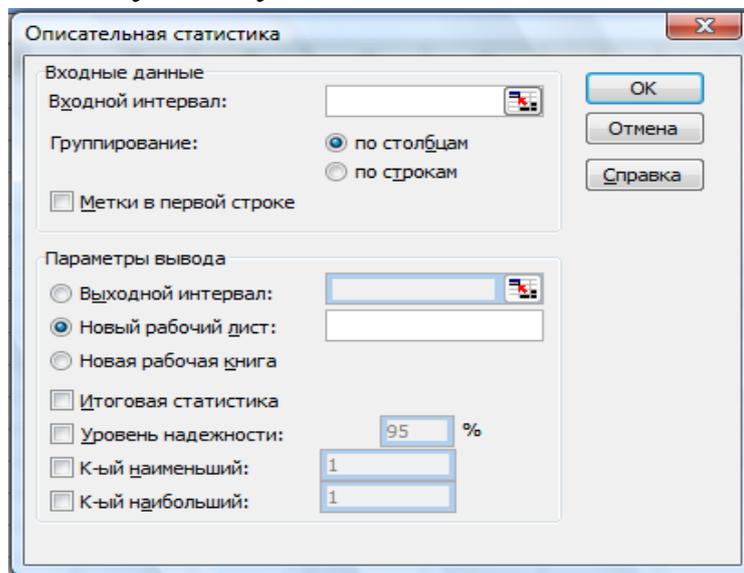
#### Краткая теория

В ЭТ MS Excel имеется набор мощных инструментов для работы с выборками и углубленного статистического анализа данных, называемый Пакет анализа, который может быть использован для решения задач статистической обработки выборочных данных.

Надстройка Пакет анализа вызывается командой главного меню Данные → Анализ данных. В появившемся окне Анализ данных выбираем пункт Описательная статистика.



Далее откроется окно Описательная статистика, в котором необходимо сделать нужные установки.



*Входной диапазон.* Ссылка на диапазон, содержащий анализируемые данные. Ссылка должна состоять не менее чем из двух смежных диапазонов данных, данные в которых расположены по строкам или столбцам.

*Группирование.* Установите переключатель в положение «По столбцам» или «По строкам» в зависимости от расположения данных во входном диапазоне.

*Метки в первой строке/Метки в первом столбце.* Если первая строка исходного диапазона содержит названия столбцов, установите переключатель в положение Метки в первой строке. Если названия строк находятся в первом столбце входного диапазона, установите переключатель в положение Метки в первом столбце. Если входной диапазон не содержит меток, то необходимые заголовки в выходном диапазоне будут созданы автоматически.

*Уровень надежности.* Установите флажок, если в выходную таблицу необходимо вывести границу доверительного интервала для среднего. В поле введите требуемое значение в процентах. Например, значение 95% вычисляет уровень надежности среднего с уровнем значимости 0,05.

*К-ый наибольший.* Установите флажок, если в выходную таблицу необходимо включить строку для k-го наибольшего значения для каждого диапазона данных. В соответствующем окне введите число k. Если k равно 1, эта строка будет содержать максимальное значение выборки.

*К-ый наименьший.* Установите флажок, если в выходную таблицу необходимо включить строку для k-го наименьшего значения для каждого диапазона данных. В соответствующем окне введите число k. Если k равно 1, эта строка будет содержать минимальное значение выборки.

*Выходной диапазон.* Введите ссылку на левую верхнюю ячейку выходного диапазона. Этот инструмент анализа выводит два столбца сведений для каждого набора данных. Левый столбец содержит метки статистических данных; правый столбец содержит статистические данные. Состоящий из двух столбцов диапазон статистических данных будет выведен для каждого столбца или для каждой строки входного диапазона в зависимости от положения переключателя Группирование.

Если хотим вывести результаты расчета на новый лист, то установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки А1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

Если хотим вывести результаты расчета в новой книге, то установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку А1 на первом листе в этой книге.

*Итоговая статистика.* Установите флажок, если в выходном диапазоне необходимо получить по одному полю для каждого из следующих видов статистических данных, представленных в таблице 2.

Таблица 2.

Значение	Примечания
Среднее	Выборочное среднее $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$ . Функция СРЗНАЧ.
Стандартная ошибка	Оценка среднеквадратичного отклонения выборочного среднего. Вычисляется по формуле $\sqrt{\frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$
Медиана	Число, которое является серединой множества чисел, то есть половина чисел имеют значения большие, чем медиана, а половина чисел имеют значения меньшие, чем медиана. Функция МЕДИАНА.
Мода	Наиболее часто встречающееся значение в выборке. Если нет одинаковых значений, то возвращается значение ошибки #Н/Д. Функция МОДА.ОДН.
Стандартное отклонение	Оценка среднеквадратичного отклонения генеральной совокупности $S = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$ . Функция СТАНДОТКЛОН.В.
Дисперсия выборки	Оценка дисперсии генеральной совокупности $D = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ . Функция ДИСП.В.
Эксцесс	Выборочный эксцесс. Функция ЭКСЦЕСС.
Асимметричность	Коэффициент асимметрии. Функция СКОС.
Интервал	Размах варьирования $R = X_{\max} - X_{\min}$ .
Минимум	Минимальное значение в выборке. Функция МИН.

Максимум	Максимальное значение в выборке. Функция МАКС.
Сумма	Сумма всех значений в выборке. Функция СУММ.
Счет	Объем выборки. Функция СЧЕТ.
Наибольший	k-тое наибольшее значение выборки. Если k=1, то выводится максимальное значение. Функция НАИБОЛЬШИЙ.
Наименьший	k-тое наименьшее значение выборки. Если k=1, то выводится минимальное значение. Функция НАИМЕНЬШИЙ
Уровень надежности	Параметр показывает возможность отклонения среднего по выборке, от среднего для генеральной совокупности, при заданном уровне надежности.

*Замечание.* Следует обратить внимание на то, что расчет параметров в режиме *Описательная статистика* имеет ряд важных особенностей:

1. В качестве значений параметров: Стандартное отклонение, Дисперсия выборки, Эксцесс, Асимметричность – Excel генерирует оценки соответствующих параметров для генеральной совокупности, а не для выборки.

2. Для применения *Описательной статистики* предварительное ранжирование исходных данных не требуется: при вычислении показателей ранжирование выполняется автоматически.

3. Появление в ячейке *Мода* индикатора ошибки #Н/Д указывает на то, что в анализируемых данных нет одинаковых значений признака. В этом случае в качестве моды  $M_o$  выбирается то значение признака, которое соответствует максимальной ординате теоретической кривой распределения.

4. Индикатор ошибки #ДЕЛ/0! В ячейке *Эксцесс* и/или *Асимметричность* означает, что в результирующей таблице стандартное отклонение является нулевым или же заданный входной диапазон данных содержит менее четырех элементов данных

5. Стандартная ошибка – это разность между ожидаемыми и наблюдаемыми значениями исследуемого признака.

*Стандартная ошибка* или *ошибка среднего* находится из выражения

$$m = \frac{S}{\sqrt{n}}$$

Стандартная ошибка – это параметр, характеризующий степень возможного отклонения среднего значения, полученного на исследуемой ограниченной выборке, от истинного среднего значения, полученного на всей совокупности элементов. С помощью стандартной ошибки задается так называемый доверительный интервал. 95%-ый доверительный интервал, равный  $x \pm 2m$ , обозначает диапазон, в который с вероятностью  $p = 0,95$  (при достаточно большом числе наблюдений  $n > 30$ ) попадает среднее значение генеральной совокупности.

### Пример выполнения

**Постановка задачи.** Приведены объемы дневной выручки (в тыс. руб.) 24 продавцов колбасных изделий, работающих в разных районах города (см. табл.1).

Таблица 1.

20,2	19,3	19,9	23,1	18,8	17,4
19,9	18,3	16,4	17,3	18,3	15,8
20,5	20,6	19,4	18,7	16,3	18,4
21,6	21,2	19,3	19,1	19,3	18,8

**Требуется:** выполнить описательную статистику выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

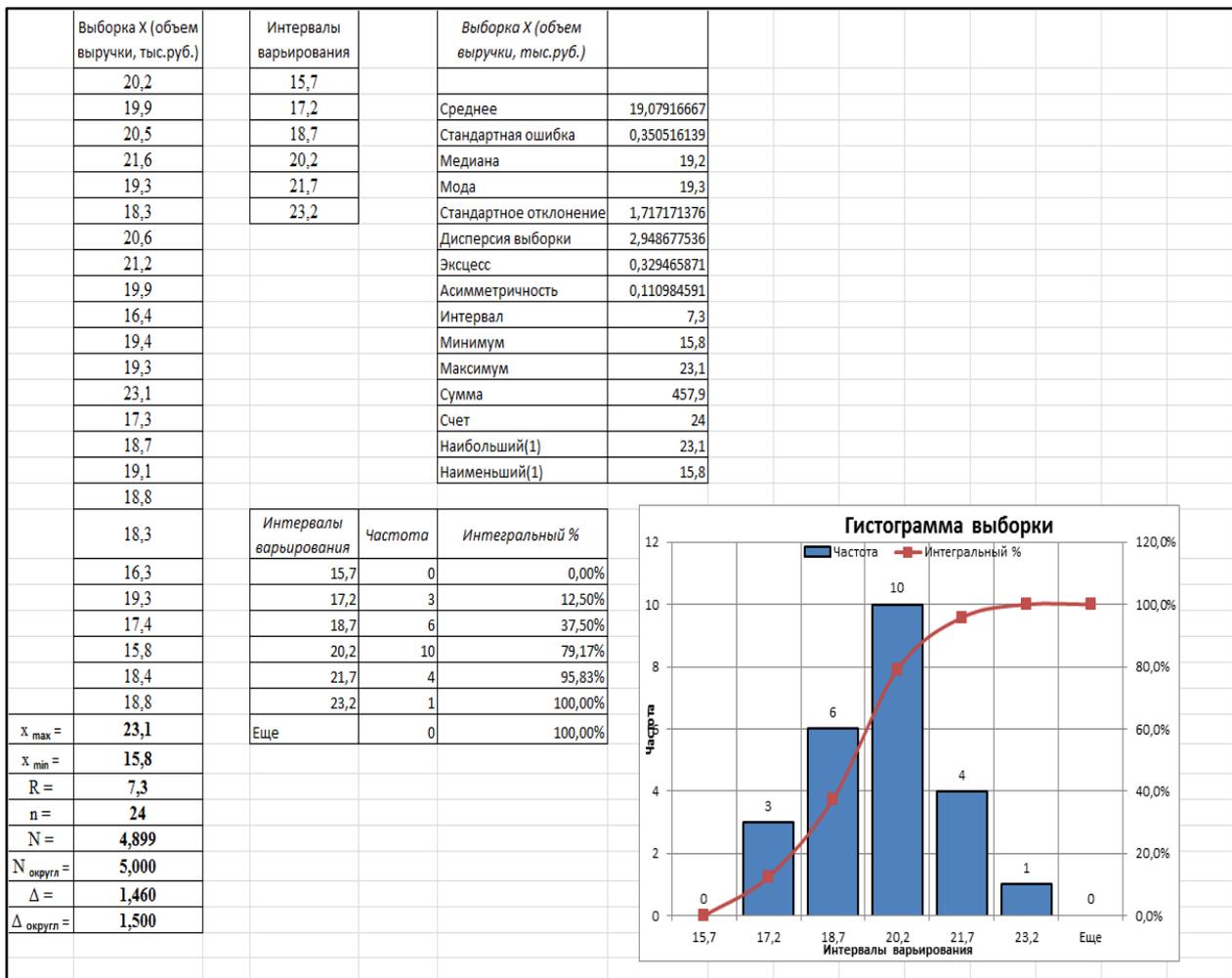
**Решение задачи в среде ЭТ MS Excel.** Для решения задачи в среде ЭТ MS Excel необходимо выполнить следующие действия:

1. Идентифицируйте свою работу, переименовав Лист1 в Титульный лист и записав номер лабораторной работы, ее название, кто выполнил и проверил.

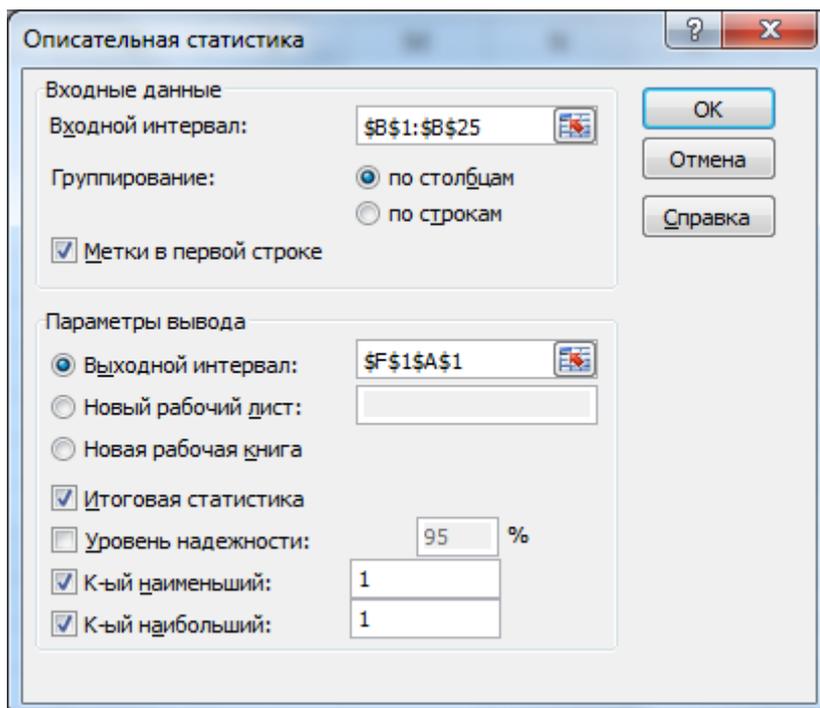
2. Переименуйте Лист 2 в Исходные данные и наберите столбец исходных данных.

3. Вычислите величины  $x_{max}$ ,  $x_{min}$ ,  $R$ ,  $n$ ,  $N$ ,  $N_{округл.}$ ,  $\Delta$  и  $\Delta_{округл.}$ , используя встроенные функции Excel **МАКС**, **МИН**, **СЧЕТ**, **КОРЕНЬ** и **ОКРУГЛ**.

4. Сформируйте столбец интервалов группировки. Наберите команду Данные → Анализ данных → Гистограмма и в появившемся диалоговом окне выполните нужные установки. Отформатируйте полученную таблицу и построенную гистограмму выборки.



5. Наберите команду Данные → Анализ данных → Описательная статистика и в появившемся диалоговом окне выполните нужные установки.



6. Щелчок по кнопке «ОК» приводит к появлению результирующей таблицы статистических характеристик выборки.

F	G
Выборка X (объем выручки, тыс.руб.)	
Среднее	19,07916667
Стандартная ошибка	0,350516139
Медиана	19,2
Мода	19,3
Стандартное отклонение	1,717171376
Дисперсия выборки	2,948677536
Эксцесс	0,329465871
Асимметричность	0,110984591
Интервал	7,3
Минимум	15,8
Максимум	23,1
Сумма	457,9
Счет	24
Наибольший(1)	23,1
Наименьший(1)	15,8

7. Повторно вычислим найденные характеристики с помощью встроенных функций MS Excel или формул. Сравним полученные результаты.

F	G	H	I	J	K	L
Выборка X (объем выручки, тыс.руб.)						
Среднее	19,07916667	19,07916667	Функция СРЗНАЧ	$t = \frac{S}{\sqrt{n}}$		
Стандартная ошибка	0,350516139	0,350516139	←			
Медиана	19,2	19,2	Функция МЕДИАНА			
Мода	19,3	19,3	Функция МОДА.ОДН			
Стандартное отклонение	1,717171376	1,717171376	Функция СТАНДОТКЛОН.В			
Дисперсия выборки	2,948677536	2,948677536	Функция ДИСП.В			
Эксцесс	0,329465871	0,329465871	Функция ЭКСЦЕСС			
Асимметричность	0,110984591	0,110984591	Функция СКОС			
Интервал	7,3	7,3	$R = x_{\max} - x_{\min}$			
Минимум	15,8	15,8	Функция МИН			
Максимум	23,1	23,1	Функция МАКС			
Сумма	457,9	457,9	Функция СУММ			
Счет	24	24	Функция СЧЕТ			
Наибольший(1)	23,1	23,1	Функция НАИБОЛЬШИЙ			
Наименьший(1)	15,8	15,8	Функция НАИМЕНЬШИЙ			

8. Сделайте выводы и сохраните работу в вашем каталоге.

### Исходные данные для самостоятельного решения

**Задание.** Имеется выборка объема  $n = 27$  (табл. 2).

**Требуется:** выполнить описательную статистику выборки с помощью Надстройки Пакет Анализа ЭТ MS Excel.

Таблица 2.

№ варианта	Выборка								
1	22,5	20,2	19,3	19,9	23,1	18,8	17,4	21,6	19,1
	21,6	19,9	18,3	16,4	17,3	18,3	15,8	21,2	19,3
	17,8	20,5	20,6	19,4	18,7	16,3	18,4	19,3	18,8
2	18,8	20,2	19,3	19,9	23,2	22,5	17,4	21,8	19,2
	19,4	18,7	16,3	18,4	19,3	18,8	19,4	18,7	16,3
	20,5	20,6	19,4	18,7	16,3	18,4	19,3	18,8	17,8
2	20,2	19,3	19,9	23,1	18,8	17,4	21,6	19,1	22,4
	18,7	20,2	19,3	19,9	23,2	22,5	17,4	21,8	19,2
	18,1	19,8	18,2	16,4	17,2	21,8	15,8	21,2	19,2
3	19,4	18,7	16,3	18,4	19,3	18,8	19,4	18,7	16,3
	18,5	20,6	19,4	20,7	16,3	18,4	19,3	18,8	17,8
	20,1	19,3	19,9	23,1	18,8	17,4	21,6	19,1	22,4
4	19,7	20,2	19,3	18,9	23,2	22,5	17,4	21,8	19,2
	18,3	19,8	18,2	16,4	17,2	21,8	15,8	21,2	19,2
	19,7	18,7	16,3	18,4	19,3	18,8	19,4	18,7	16,3
5	19,4	20,7	16,3	18,4	19,3	18,8	17,8	18,7	20,2
	19,9	23,1	18,8	17,4	21,6	19,1	22,4	18,1	19,8
	19,3	18,9	23,2	22,5	17,4	21,8	19,2	19,4	18,7
6	18,7	16,3	18,4	19,3	18,8	19,4	18,7	18,5	20,6
	20,6	19,4	20,7	16,3	18,4	19,3	18,8	18,4	19,3
	19,3	19,9	23,1	18,8	17,4	21,6	19,1	18,4	19,3
7	16,3	18,4	19,3	18,8	19,4	18,7	18,5	20,6	18,7
	19,4	20,7	16,3	18,4	19,3	18,8	18,4	19,3	20,6
	19,9	23,1	18,8	17,4	21,6	19,1	18,4	19,3	19,3
8	19,3	19,9	23,1	18,8	17,4	21,6	19,1	22,5	20,2
	18,3	16,4	17,3	18,3	15,8	21,2	19,3	21,6	19,9
	20,6	19,4	18,7	16,3	18,4	19,3	18,8	17,8	20,5
9	19,4	20,7	16,3	18,4	19,3	18,8	17,8	18,7	20,2
	19,9	23,1	18,8	17,4	21,6	19,1	22,4	18,1	19,8
	19,3	18,9	23,2	22,5	17,4	21,8	19,2	19,4	18,7
10	18,7	16,3	18,4	19,3	18,8	19,4	18,7	18,5	20,6
	20,6	19,4	20,7	16,3	18,4	19,3	18,8	18,4	19,3
	16,4	20,4	20,8	19,4	18,7	17,8	18,4	19,4	18,8

## Список использованных источников

1. Гмурман В. Е. Теория вероятностей и математическая статистика [Текст]: Учеб.пособие для вузов / В. Е. Гмурман. – М. : Высшая школа, 2005. – 480 с.
2. Горелова Г. В. Теория вероятностей и математическая статистика в примерах и задачах с применением Excel [Текст]: Учеб. пособие для вузов / Г. В. Горелова, И. А. Кацко. – Ростов н/Д.: Феникс, 2005. – 477 с.
3. Воскобойников Ю.Е. Математическая статистика (с примерами в Excel) [Текст]: Учеб. пособие / Ю.Е.Воскобойников, Е.И.Тимошенко; Новосиб. гос. архитектур.-строит. ун-т (Сибстрин). – 2-е изд., перераб. и доп. – Новосибирск: НГАСУ (Сибстрин), 2006. – 152 с.
4. Ильченко М.А., Струкова Л.С. Excel в математических и статистических расчетах [Текст]: Учеб. пособие / М.А.Ильченко, Л.С.Струкова; Мичурин. гос. аграр. ун-т. – Мичуринск: МГАУ, 2007. – 136 с.
5. Смирнов В.А. Прикладная статистика в пакете анализа MS Excel [Текст]: Учебное пособие / В.А. Смирнов. – Пенза: ПГУАС, 2008. – 88 с.
6. Макарова Н.В. Статистика в Excel [Текст]: Учеб. пособие / Н.В.Макарова, В.Я. Трофимец. – М.: Финансы и статистика, 2002. – 368 с.
7. Никитина Н. Ш. Математическая статистика для экономистов [Текст]: Учеб. пособие для вузов / Н. Ш. Никитина. – М.: ИНФРА-М, 2001. – 170 с.

Подписано в печать 23.12.2014.  
Формат 60×84 <sup>1</sup>/<sub>16</sub>. Усл. печ. л. 3,72. Уч.-изд. лист. 1,39.  
Тираж 50 экз. Заказ № 4136.

Издательство Южного федерального университета.

Отпечатано в отделе полиграфической, корпоративной и сувенирной продукции  
Издательско-полиграфического комплекса КИБИ МЕДИА ЦЕНТРА ЮФУ  
344090, г. Ростов-на-Дону, пр. Стачки, 200/1. Тел. (863) 247-80-51.