

**«Национальный открытый институт г.Санкт-Петербург»**

Боброва Л.В., Романова Ю.С.

# **Теория вероятностей и математическая статистика**

**Методические указания  
к выполнению практических работ**

Рекомендовано Методической комиссией по качеству  
Национального открытого института  
для студентов, обучающихся по специальности  
09.02.05 – Прикладная информатика (по отраслям)

Санкт-Петербург  
2016

**УДК 681.142.2**  
**ББК 22.19**  
**Б72**

Приведены методические указания для выполнения трех практических работ с использованием табличного процессора Excel.

Рассматриваются задачи статистического анализа числовых характеристик и функций распределения выборок, определения доверительных интервалов и проверки статистических гипотез. Даются навыки генерирования выборок случайных величин с заданными законами распределения.

Приводятся также задачи корреляционного анализа.

**УДК 004**  
**ББК 32.97**

© Боброва Л.В. 2016  
© Национальный открытый институт 2016  
© ИКЦ 2016

## ОБЩИЕ УКАЗАНИЯ

Табличный процессор Excel для решения сложных статистических задач предоставляет такие возможности, как аппарат формул и функций из категории **Статистические**, анализ данных в виде **Пакета анализа**, графические средства. Приобретение навыков использования указанных средств, освоение основных статистических методов является основной целью выполнения практических работ.

### Работа 1. АНАЛИЗ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ ВЫБОРКИ

#### 1. Цель работы

Изучить возможности табличного процессора Excel при вычислении оценок параметров случайных выборок.

#### 2. Основные теоретические положения

Экономические показатели следует рассматривать как случайные величины. Для принятия решения на основе анализа экономических показателей нужно уметь вычислять основные параметры случайных выборок. Наиболее важными параметрами являются функция и плотность распределения. Для расчета функций распределения определяют три вида частот:

- 1) Абсолютные частоты – это число попаданий значений случайной величины в заданный интервал.
- 2) Относительные частоты – это отношение абсолютной частоты к объему выборки.
- 3) Накопленные частоты образуются при последовательном суммировании относительных частот.

#### 3. Порядок выполнения работы

*Задание 1.* Проводилось исследование числа продаваемых ежедневно магазином телевизоров. Получена выборка, характеризующая спрос на этот товар.

43, 36, 42, 41, 37, 40, 42, 39, 39, 40, 44, 41, 41, 39,  
 43, 40, 38, 38, 42, 40, 41, 37, 37, 42, 40, 38, 41, 39,  
 39, 37, 40, 39, 42, 42, 37, 39, 38, 39, 38, 40, 41, 41,  
 42, 36, 40, 37, 39, 43, 39, 38, 40, 43, 41, 38, 44

С использованием функции ЧАСТОТА провести вычисление абсолютных, относительных и накопленных частот для выборки.

**Задание 2.** Построить диаграммы относительных и накопленных частот с использованием **Мастера диаграмм**.

**Задание 3.** Вычислить частоты и построить диаграммы с использованием **Пакета анализа**.

**Задание 4.** Провести оценки статистических характеристик выборки с использованием специальных функций.

**Задание 5.** Провести оценку статистических характеристик выборок с использованием **Пакета анализа**.

**Задание 6.** Провести оценку параметров выборки согласно индивидуальному заданию.

### 3.1. Выполнение задания 1. Вычисление частот.

3.1.1. Заполнить исходными данными ячейки A2:E12 электронной таблицы (табл. 1).

*Таблица 1*

	A	B	C	D	E
1	Наблюдения				
2	43	36	42	41	37
3	40	42	39	39	40
4	44	41	41	39	43
5	40	38	38	42	40
6	41	37	37	42	40
7	38	41	39	39	37
8	40	39	42	42	37
9	39	38	39	38	40
10	41	41	42	36	40
11	37	39	43	39	38
12	40	43	41	38	44

3.1.2. Ввести значения правых границ интервалов в ячейки G2:G10 (табл. 2).

3.1.3. Для вычисления массива значений абсолютных частот:

а) выделить блок ячеек H2:H10;

б) выполнить команды: *Формулы – Вставить функцию – Статистические – Частота - Ок*;

в) во втором окне Мастера функций в поле **Массив\_данных** ввести адрес диапазона исходных данных **A2:E12**;

г) в поле **Массив интервалов** ввести адрес диапазона интервалов **G2:G10** (*кнопку Ок не нажимать!*)

д) чтобы вычисления были приведены для всего выделенного диапазона, одновременно нажать клавиши Ctrl+Shift+Enter. Получаем табл. 2.

3.1.4. Для проверки вычисления суммируем абсолютные частоты данных (должны получить объем выборки  $n=55$ ).

Для этого ввести в ячейку Н12 формулу =СУММ(Н2:Н10) – см. табл. 2. Удобно воспользоваться пиктограммой автосуммирования на вкладке **Формулы**

Таблица 2

	A	B	C	D	E	F	G	Н	I	J
1	Наблюдения						Интервал (прав. границы)	Абсолют ная частота	Относи- тельная частота	Накоп- ленная частота
2	43	36	42	41	37		36	2	0,036	0,036
3	40	42	39	39	40		37	6	0,109	0,145
4	44	41	41	39	43		38	7	0,127	0,273
5	40	38	38	42	40		39	10	0,182	0,455
6	41	37	37	42	40		40	9	0,164	0,618
7	38	41	39	39	37		41	8	0,145	0,764
8	40	39	42	42	37		42	7	0,127	0,891
9	39	38	39	38	40		43	4	0,073	0,964
10	41	41	42	36	40		44	2	0,036	1,000
11	37	39	43	39	38					
12	40	43	41	38	44		Объем выборки	55		

3.1.5. Для вычисления массива относительных частот:

а) ввести в ячейку I2 формулу =Н2/\$Н\$12.

б) скопировать ее в ячейки I3:I10.

3.1.6. Для вычисления массива накопленных частот:

а) ввести в ячейку J2 формулу =I2;

б) ввести в ячейку J3 формулу =J2+I3;

в) скопировать из J3 формулу в ячейки J4:J10.

В режиме показа формул Таблица показана в табл. 3.

3.2. Выполнение задания 2. Построение гистограммы относительных частот.

3.2.1. Выделить диапазон относительных частот.

3.2.2. Запустить режим Мастера диаграмм (**Вставка-Гистограмма**).

3.2.3. Для верного задания подписей по оси абсцисс щелкнуть **правой** клавишей мыши по области графика – **Выбрать данные – Подписи горизонтальной оси - Изменить** - ввести нужный диапазон **I2:J10**- **Ок – Ок**.

Таблица 3

	A	B	C	D	E	F	G	H	I	J
1	Наблюдения					Интервал (прав. границы)	Абсолютная частота	Относи- тельная частота	Накоп- ленная частота	
2	43	36	42	41	37	36	=ЧАСТОТА(A2:E12;G2:G10)	=H2/H\$12	=I2	
3	40	42	39	39	40	37	=ЧАСТОТА(A2:E12;G2:G10)	=H3/H\$12	=J2+I3	
4	44	41	41	39	43	38	=ЧАСТОТА(A2:E12;G2:G10)	=H4/H\$12	=J3+I4	
5	40	38	38	42	40	39	=ЧАСТОТА(A2:E12;G2:G10)	=H5/H\$12	=J4+I5	
6	41	37	37	42	40	40	=ЧАСТОТА(A2:E12;G2:G10)	=H6/H\$12	=J5+I6	
7	38	41	39	39	37	41	=ЧАСТОТА(A2:E12;G2:G10)	=H7/H\$12	=J6+I7	
8	40	39	42	42	37	42	=ЧАСТОТА(A2:E12;G2:G10)	=H8/H\$12	=J7+I8	
9	39	38	39	38	40	43	=ЧАСТОТА(A2:E12;G2:G10)	=H9/H\$12	=J8+I9	
10	41	41	42	36	40	44	=ЧАСТОТА(A2:E12;G2:G10)	=H10/H\$12	=J9+I10	
11	37	39	43	39	38					
12	40	43	41	38	44	Объем выборки	=СУММ(H2:H10)			
13										

3.2.4. Проверьте правильность подписей по осям.

3.3. Аналогичным образом, используя График, постройте график накопленных частот.

3.4. Сохраните файл с именем **Лаб\_работа\_1**

3.5.Выполнение задания 3.Построение диаграмм с использованием пакета анализа.

3.3.1. Создать копию ЭТ, табл. 2.

3.3.2. Выполните команды вкладка **Данные** – группа **Анализ** – кнопка **Анализ данных** – **Гистограмма**.

**Внимание!** Если на вкладке **Данные** отсутствует кнопка **Анализ данных**: щелкнуть по кнопке **Office** – **Параметры Excel** – **Надстройки** – **Надстройки Excel** – **Перейти** – поставить флажок **Пакет анализа** – **Ок**.

3.3.3. В появившемся окне **Гистограмма** заполните рабочие поля

- **входной интервал** – ввести **A2:E12**;
- **выходной интервал** – ввести ссылку на первую ячейку выходного диапазона **K1**;
- поставить флажки **Интегральный процент** и **Вывод графика** - Ок.

В результате появится табл. 4.

*Таблица 4*

К	L	M
Карман	Частота	Интегральный %
36	2	3,33%
38,66667	15	28,33%
41,33333	29	76,67%
Еще	14	100,00%

Сравните полученную диаграмму с диаграммой п.3.2. Диаграмма п.3.3. менее точна, т.к. диапазон карманов (интервалов) не был введен и определяется автоматически.

3.4. Выполнение задания 4. Вычисление статистических характеристик с использованием специальных функций:

3.4.1. Открыть ЭТ Лаб\_работа\_1 (табл. 2).

3.4.2. В ячейки A13:F13 ввести текст «Среднее значение», а в G13 формулу для вычисления среднего значения **=СРЗНАЧ(A2:E12)**.

3.4.3. Аналогичным образом ввести (см. табл. 5):

**Ячейка**            **Содержимое ячейки**

A14: F14 ⇒ Дисперсия выборки

G14: I14 ⇒ **=ДИСП(A2:E12)**

A15: F15 ⇒ Стандартное отклонение

G15: I15 ⇒ **=СТАНДОТКЛОН(A2:E12)**

A16: F16 ⇒ Мода

G16: I16 ⇒ **=МОДА(A2:E12)**

A17: F17 ⇒ Медиана

G17: I17 ⇒ **=МЕДИАНА(A2:E12)**.

Таблица 5

	A	B	C	D	E	F	G	H	I	J
1	Наблюдения						Интервал (правые границы)	Абсолютная частота	Относительная частота	Накопленная частота
2	43	36	42	41	37		36	2	0,036	0,036
3	40	42	39	39	40		37	6	0,109	0,145
4	44	41	41	39	43		38	7	0,127	0,273
5	40	38	38	42	40		39	10	0,182	0,455
6	41	37	37	42	40		40	9	0,164	0,618
7	38	41	39	39	37		41	8	0,145	0,764
8	40	39	42	42	37		42	7	0,127	0,891
9	39	38	39	38	40		43	4	0,073	0,964
10	41	41	42	36	40		44	2	0,036	1,000
11	37	39	43	39	38					
12	40	43	41	38	44		Объем выборки	55		
13	Среднее значение						=СРЗНАЧ(A2:E12)			
14	Дисперсия						=ДИСП(A2:E12)			
15	Стандартное отклонение						=СТАНДОТКЛОН(A2:E12)			
16	Мода						=МОДА(A2:E12)			
17	Медиана						=МЕДИАНА(B3:B12)			

Результаты вычислений показаны в табл.6.

### 3.5. Выполнение задания 5. Вычисление статистических характеристик с использованием Пакета анализа.

3.5.1. Выполните команды **Данные - Анализ данных - Описательная статистика - Ок.**

3.5.2. В появившемся окне в поле **Входной интервал** введите **A2:E12**.

3.5.3. В поле **Выходной интервал** укажите начало размещения выходных данных – **A19**.

3.5.4. В разделе **Группировка** установить переключатель **По столбцам**.

3.5.5. Установите флажок **Итоговая статистика** и **Ок.**

Таблица 6

	A	B	C	D	E	F	G	H	I	J
1	Наблюдения						Интервал (прав. границы)	Абсолютная частота	Относи- тельная частота	Накоп- ленная частота
2	43	36	42	41	37		36	2	0,036	0,036
3	40	42	39	39	40		37	6	0,109	0,145
4	44	41	41	39	43		38	7	0,127	0,273
5	40	38	38	42	40		39	10	0,182	0,455
6	41	37	37	42	40		40	9	0,164	0,618
7	38	41	39	39	37		41	8	0,145	0,764
8	40	39	42	42	37		42	7	0,127	0,891
9	39	38	39	38	40		43	4	0,073	0,964
10	41	41	42	36	40		44	2	0,036	1,000
11	37	39	43	39	38					
12	40	43	41	38	44		Объем выборки	55		
13	Среднее значение							39,85		
14	Дисперсия							4,20		
15	Стандартное отклонение							2,05		
16	Мода							39		
17	Медиана							40		

### 3.6. Выполнение задания 6

3.6.1. Выполнить пункты 3.1-3.5 для одной из выборок, приведенных ниже. Вариант задания выбирается по последней цифре Вашего шифра (при этом создайте новый файл, скопировав файл Лаб\_работа1).

3.6.2. Сохранить файл с именем Работа1\_сам.

## **4. Отчет по работе**

- 4.1. Файл с таблицами 5 и 6.
- 4.2. Файл с таблицей индивидуального задания.

## **Индивидуальные задания к работе 1**

### **Задание 1**

Построить эмпирическое распределение числа заявок, обслуживаемых ежедневно на автостанции (гистограмму относительных частот и график накопленных частот), для заданной выборки (вариант выбирается по последней цифре шифра).

## Задание 2

Вычислить выборочное среднее, медиану, моду, дисперсию и стандартное отклонение для выборки из задания 1.

### Вариант 1

53, 44, 50, 49, 45, 48, 50, 47, 47, 48, 52, 49, 49, 47, 51, 48, 46, 46, 50, 48, 49, 45, 45, 50, 48, 46, 49, 47, 47, 45, 48, 47, 50, 50, 45, 47, 46, 47, 46, 48, 49, 49, 50, 44, 48, 45, 47, 51, 47, 46, 48, 51, 49, 46, 52.

### Вариант 2

42, 33, 39, 38, 34, 37, 39, 36, 36, 34, 41, 38, 38, 36, 40, 37, 35, 35, 39, 37, 38, 34, 34, 39, 37, 35, 39, 36, 36, 34, 37, 36, 39, 39, 34, 36, 35, 36, 35, 37, 38, 38, 39, 33, 37, 34, 36, 40, 36, 35, 37, 40, 38, 35, 41

### Вариант 3

31, 22, 28, 27, 23, 26, 28, 25, 25, 26, 30, 27, 27, 25, 29, 26, 24, 24, 28, 26, 27, 23, 23, 28, 26, 24, 27, 25, 25, 23, 26, 25, 28, 28, 23, 25, 24, 25, 24, 26, 27, 27, 28, 22, 26, 23, 25, 29, 25, 24, 26, 29, 27, 24, 30.

### Вариант 4

59, 50, 56, 55, 54, 54, 56, 53, 53, 54, 58, 55, 55, 53, 57, 54, 52, 52, 56, 54, 55, 51, 51, 56, 54, 52, 55, 53, 53, 51, 54, 53, 56, 56, 51, 53, 54, 53, 54, 54, 55, 55, 56, 50, 54, 51, 53, 57, 53, 52, 54, 57, 55, 52, 58.

### Вариант 5

#### Задание 1

33, 24, 30, 29, 25, 28, 30, 27, 27, 28, 32, 29, 29, 27, 31, 28, 26, 26, 30, 28, 29, 25, 25, 30, 28, 26, 29, 27, 27, 25, 28, 27, 30, 30, 25, 27, 26, 27, 26, 28, 29, 29, 30, 24, 28, 25, 27, 31, 27, 26, 28, 31, 29, 26, 32.

### Вариант 6

23, 14, 20, 19, 15, 18, 20, 17, 17, 18, 22, 19, 19, 17, 21, 18, 16, 16, 20, 18, 19, 15, 15, 20, 18, 16, 19, 17, 17, 15, 18, 17, 20, 20, 15, 17, 16, 17, 16, 18, 19, 19, 20, 14, 18, 15, 17, 21, 17, 16, 18, 21, 19, 16, 22.

### Вариант 7

68, 59, 65, 64, 60, 63, 65, 62, 62, 63, 65, 64, 64, 62, 66, 63, 61, 61, 65, 63, 64, 60, 60, 65, 63, 61, 64, 62, 62, 60, 63, 62, 65, 65, 60, 62, 61, 62, 61, 63, 64, 64, 65, 59, 63, 60, 62, 66, 62, 61, 63, 66, 65, 61, 67.

### **Вариант 8**

44, 35, 41, 40, 36, 39, 41, 38, 38, 39, 43, 40, 40, 39, 42, 39, 37, 37, 41, 39, 40, 36, 36, 41, 39, 37, 40, 39, 39, 36, 39, 38, 41, 41, 36, 36, 37, 38, 37, 37, 40, 40, 41, 35, 39, 36, 39, 42, 39, 37, 39, 42, 40, 37, 43.

### **Вариант 9**

73, 74, 70, 69, 65, 68, 70, 67, 67, 68, 72, 69, 69, 67, 71, 68, 66, 66, 70, 68, 69, 65, 65, 70, 68, 66, 69, 67, 67, 65, 68, 67, 70, 70, 65, 67, 66, 67, 66, 68, 66, 70, 64, 68, 68, 65, 67, 71, 67, 66, 68, 71, 69, 66, 72.

### **Вариант 0**

57, 48, 54, 53, 49, 52, 54, 51, 51, 52, 56, 53, 53, 51, 55, 52, 50, 50, 54, 52, 53, 49, 49, 54, 52, 50, 53, 51, 51, 49, 52, 51, 54, 54, 49, 51, 50, 51, 50, 52, 53, 53, 54, 48, 52, 49, 51, 55, 51, 50, 52, 55, 53, 50, 56.

## **Работа 2. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ. ОЦЕНКА ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ**

### **1. Цель работы**

Научиться проверять предположения о возможных значениях параметров и законов распределения выборок.

### **2. Основные теоретические положения**

#### **2.1. Проверка статистических гипотез**

Пусть имеется выборка значений случайной величины. При решении большинства задач реальные законы распределения эмпирических данных неизвестны. Обычно выдвигается статистическая гипотеза о предполагаемом виде распределения, которая проверяется с помощью критериев согласия.

**Статистическая гипотеза** – это предположение о виде закона распределения выборки или отдельных параметрах этого распределения.

Проверка статистических гипотез – процесс формирования решения о возможности принять или отвергнуть гипотезу. Метод проверки гипотезы называется **статистическим критерием**.

Обычно гипотеза формулируется так, чтобы на основе эксперимента ее можно было принять или отвергнуть с заранее заданной вероятностью ошибки  $\alpha$ . Эта вероятность  $\alpha$  называется **уровнем значимости**. В статистике, как правило, берут уровни значимости, равные 0,05; 0,01 или 0,001.

Наибольшее распространение получил критерий согласия  $\chi^2$  (**хи-квадрат**) или **критерий Пирсона**.

Критерий Пирсона основан на сравнении эмпирических частот с теоретическими (ожидаемыми) частотами, рассчитанными по формуле распределения. Лучшие результаты критерий согласия  $\chi^2$  дает при работе с нормальным распределением. Это самое распространенное на практике распределение.

## **2.2. Построение доверительных интервалов для среднего значения**

Кроме параметров случайной величины оценивают их доверительные интервалы. Например, оценка доверительного интервала среднего значения (т.е. математического ожидания случайной величины) производится по формуле

$$[m_n; m_g] = \left[ m - t \frac{s}{\sqrt{n}}; m + t \frac{s}{\sqrt{n}} \right]$$

здесь  $m$  – среднее значение;

$m_n$  – нижняя граница доверительного интервала;

$m_g$  – его верхняя граница;

$s$  – стандартное отклонение;

$n$  – объем выборки;

$t$  – параметр, который определяется распределением

Стьюдента.

В книгах по статистике приводится таблица распределения Стьюдента. Параметр  $t$  выбирается в зависимости от заданной доверительной вероятности  $p$  и объема выборки  $n$ .

### 3. Порядок выполнения работы

*Задание 1.* Проверить гипотезу о том, что приведенная выборка имеет нормальный закон распределения.

*Задание 2.* Найти границы 95% доверительного интервала для математического ожидания выборки из задания 1.

*Задание 3.* Проверить гипотезу о нормальном распределении выборки для индивидуального варианта. Найти границы 95% доверительного интервала для математического ожидания этой выборки.

#### 3.1. Выполнение задания 1

1. Вводим исходную выборку в диапазон A2:F12 и определяем интервал ее изменения (рис. 1):

	A	B	C	D	E	F	G
1	<b>ВЫБОРКА</b>						
2	<b>24</b>	<b>24</b>	<b>22</b>	<b>29</b>	<b>24</b>	<b>25</b>	
3	<b>21</b>	<b>23</b>	<b>27</b>	<b>23</b>	<b>28</b>	<b>26</b>	
4	<b>23</b>	<b>27</b>	<b>24</b>	<b>28</b>	<b>26</b>	<b>26</b>	
5	<b>24</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>25</b>	<b>26</b>	
6	<b>28</b>	<b>26</b>	<b>25</b>	<b>27</b>	<b>22</b>	<b>25</b>	
7	<b>22</b>	<b>24</b>	<b>25</b>	<b>25</b>	<b>25</b>	<b>27</b>	
8	<b>26</b>	<b>27</b>	<b>26</b>	<b>27</b>	<b>27</b>	<b>28</b>	
9	<b>26</b>	<b>29</b>	<b>23</b>	<b>24</b>	<b>24</b>	<b>25</b>	
10	<b>27</b>	<b>25</b>	<b>27</b>	<b>26</b>	<b>26</b>	<b>24</b>	
11	<b>24</b>	<b>26</b>	<b>24</b>	<b>27</b>	<b>27</b>	<b>25</b>	
12	<b>24</b>	<b>25</b>	<b>25</b>	<b>23</b>	<b>23</b>	<b>25</b>	
13							
14				Функция			Значение
15	<b>Минимум</b>			<b>=МИН(A2:F12)</b>			<b>21</b>
16							
17	<b>Максимум</b>			<b>=МАКС(A2:F12)</b>			<b>29</b>

Рис. 1

2. Разбиваем весь диапазон значений случайной величины на несколько интервалов. Задаем правые границы интервалов в столбце G (рис. 2)

	A	B	C	D	E	F	G	H
1	<b>ВЫБОРКА</b>						Правые границы интервалов	Абсолютные частоты
2	24	24	22	29	24	25	21	1
3	21	23	27	23	28	26	22	3
4	23	27	24	28	26	26	23	7
5	24	23	24	25	25	26	24	13
6	28	26	25	27	22	25	25	14
7	22	24	25	25	25	27	26	11
8	26	27	26	27	27	28	27	11
9	26	29	23	24	24	25	28	4
10	27	25	27	26	26	24	29	2
11	24	26	24	27	27	25		
12	24	25	25	23	23	25		
13							Сумма	66

	F	G	H
1		Правые границы интервалов	Абсолютные частоты
2	25	21	=ЧАСТОТА(A2:F12;G2:G10)
3	26	22	=ЧАСТОТА(A2:F12;G2:G10)
4	26	23	=ЧАСТОТА(A2:F12;G2:G10)
5	26	24	=ЧАСТОТА(A2:F12;G2:G10)
6	25	25	=ЧАСТОТА(A2:F12;G2:G10)
7	27	26	=ЧАСТОТА(A2:F12;G2:G10)
8	28	27	=ЧАСТОТА(A2:F12;G2:G10)
9	25	28	=ЧАСТОТА(A2:F12;G2:G10)
10	24	29	=ЧАСТОТА(A2:F12;G2:G10)
11	25		
12	25		
13		Сумма	=СУММ(H2:H10)

Рис.2

3. Вычисляем абсолютные частоты (число попаданий значений случайной величины в каждый интервал) – рис. 2:

- Выделить диапазон H2:H10;
- Вставка - Функция – Частота – обвести диапазоны A2:F12 и G2:G10.
- Для контроля находим сумму абсолютных частот (в ячейке H13).

4. Вычисляем значения плотности вероятности (относительные частоты):

- В ячейку I2 вводим формулу = H2/\$H\$13;
- Копируем в ячейки I3:I10 (рис.3).
- Ctrl + Shift + Enter.;

	A	B	C	D	E	F	G	H	I		
1	<b>ВЫБОРКА</b>						Правые границы интервалов	Абсолютные частоты	Относительные частоты		I
2	24	24	22	29	24	25	21	1	0,015		Относительные частоты
3	21	23	27	23	28	26	22	3	0,045		
4	23	27	24	28	26	26	23	7	0,106	=H2/SHS13	
5	24	23	24	25	25	26	24	13	0,197	=H3/SHS13	
6	28	26	25	27	22	25	25	14	0,212	=H4/SHS13	
7	22	24	25	25	25	27	26	11	0,167	=H5/SHS13	
8	26	27	26	27	27	28	27	11	0,167	=H6/SHS13	
9	26	29	23	24	24	25	28	4	0,061	=H7/SHS13	
10	27	25	27	26	26	24	29	2	0,030	=H8/SHS13	
11	24	26	24	27	27	25				=H9/SHS13	
12	24	25	25	23	23	25				=H10/SHS13	
13							Сумма	66			

Рис. 3

5. Строим гистограмму плотности вероятности:
- Выделить значения плотности вероятности (I2:I10);
  - Вставка-Гистограмма –выбрать первую слева сверху);
- Для задания значений по оси X щелкнуть правой клавишей по диаграмме
- – Подписи горизонтальной оси – Изменить – обвести G2:G10 (рис. 4).

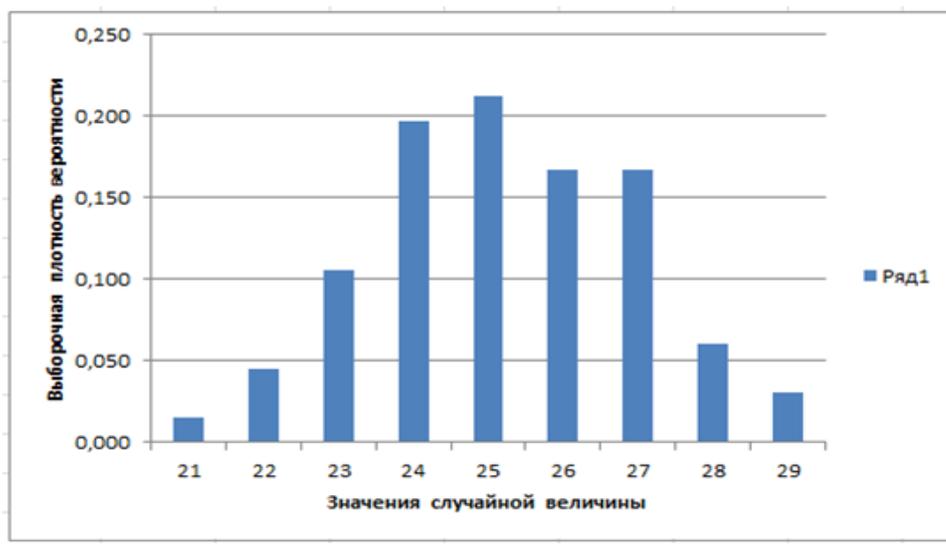


Рис. 4

5. Находим среднее значение и стандартное отклонение выборки:

функции  $=СРЗНАЧ(А2:F12)$  и  $=СТАНДОТКЛОН.В(А2:F12)$

6. Для проверки гипотезы о нормальном распределении выборки моделируем теоретическую плотность вероятности нормального закона распределения с вычисленными значениями среднего и стандартного отклонения:

- Ввести в ячейку J2 формулу  $=НОРМРАСП(G2;25,2;1,77;0)$  и скопировать в J3:J10 (рис. 5).

I	J
Относительные частоты	Теоретические относительные частоты
0,015	0,013
0,045	0,044
0,106	0,104
0,197	0,179
0,212	0,224
0,167	0,204
0,167	0,134
0,061	0,064
0,030	0,022

I	J
Относительные частоты	Теоретические относительные частоты
$=H2/SHS13$	$=НОРМ.РАСП(G2;25,2;1,77;0)$
$=H3/SHS13$	$=НОРМ.РАСП(G3;25,2;1,77;0)$
$=H4/SHS13$	$=НОРМ.РАСП(G4;25,2;1,77;0)$
$=H5/SHS13$	$=НОРМ.РАСП(G5;25,2;1,77;0)$
$=H6/SHS13$	$=НОРМ.РАСП(G6;25,2;1,77;0)$
$=H7/SHS13$	$=НОРМ.РАСП(G7;25,2;1,77;0)$
$=H8/SHS13$	$=НОРМ.РАСП(G8;25,2;1,77;0)$
$=H9/SHS13$	$=НОРМ.РАСП(G9;25,2;1,77;0)$
$=H10/SHS13$	$=НОРМ.РАСП(G10;25,2;1,77;0)$

Рис. 5

6. Строим график теоретической плотности вероятности:

- Выделить диапазон J2:J10;
- вставка – График – График с маркерами (рис. 6).

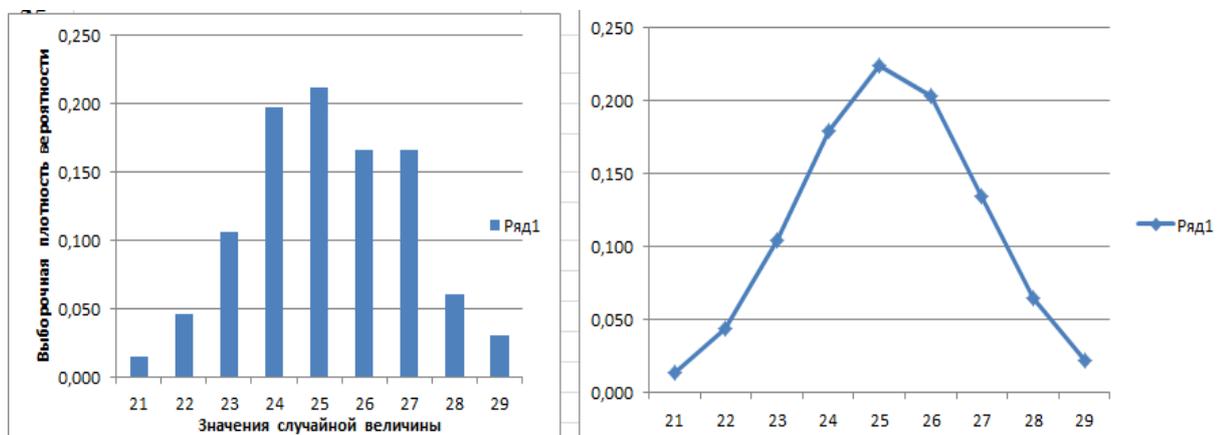


Рис. 6

7. Для проверки гипотезы о нормальном распределении выборки:
- Активизируем ячейку K2;
  - Вставка – Функция – Статистические- ХИ2ТЕСТ;
  - ввести Фактический интервал I2:I10;
  - Ожидаемый интервал J2:J10.

**Сделать вывод о законе распределения.**

3.2. Выполнение задания 2. Найти границы 95% доверительного интервала для математического ожидания выборки из задания 1.

Используется функция  
=ДОВЕРИТ(альфа;стандартное\_откл;размер)

Здесь **Альфа** =  $\alpha$ ;

**Стандартное отклонение** =  $s$ ;

**Размер** – объем выборки =  $n$ .

Проведем вычисления в Excel.

а). Установить курсор в любую ячейку электронной таблицы.

б). Вставка – Функция – Статистические – ДОВЕРИТ - **Ок.**

- в). Во втором окне **Мастера функций** ввести параметры
- в поле **Альфа** 0,05
  - **Стандартное отклонение** 1,77;
  - **Размер** 66.

Нажать **Ок.**

В выбранной ячейке появится значение **0,43** для среднего значения выборки A2:E12. Это половина доверительного интервала.

Значит, сам доверительный интервал для математического ожидания равен  $[25.2-0.43; 25.2+0.43] = [24,77; 25,63]$

3.3. Выполнение задания 3. Самостоятельная работа

Для выборки из индивидуального задания **Работы 1** проверить гипотезу о нормальном распределении и найти доверительный интервал для математического ожидания этой выборки.

#### **4. Отчет по работе**

Файл с выполненными заданиями. Выводы.

# Работа 3. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

## 1. Цель работы

Изучить основные понятия корреляционного анализа и возможности оценки корреляционной зависимости в пакете Excel.

## 2. Основные теоретические положения

### 2.1. Корреляционная зависимость

Одна из важнейших задач статистики – установление связи между наблюдаемыми явлениями. В случае **корреляционной** связи между изменением результирующего и фактор-признака нет полного соответствия. Воздействие проявляется лишь в среднем, при массовом наблюдении данных.

Корреляционная связь двух факторов (x и y) называется **парной корреляцией**. Влияние нескольких факторов на результирующий признак называется **множественной корреляцией**.

### 2.2. Коэффициент корреляции

Для определения степени тесноты связи парной линейной зависимости служит коэффициент корреляции.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{nS_x S_y}$$

Здесь  $S_x$ ,  $S_y$  – стандартные отклонения выборок x и y.

Линейный коэффициент корреляции может принимать значение в пределах  $-1 \div +1$ . Чем ближе он по абсолютной величине к 1, тем теснее связь между признаками.

### 2.3. Корреляционная матрица

При большом числе наблюдений (выборок) коэффициенты корреляции вносятся в таблицы, называемые корреляционными матрицами.

В такой матрице на пересечении строк и столбцов находятся коэффициенты корреляции между соответствующими параметрами.

### 3. Порядок выполнения работы

*Задание 1.* В табл. 7 приведены статистические данные о численности населения городов и численности больных сердечно-сосудистыми заболеваниями.

Таблица 7

Город	Москва	Рязань	Тула	Воронеж	Орел
Численность населения	8000000	300000	200000	400000	450000
Число больных	75000	10000	8000	5000	2000
Город	С.-Петербург	Екатеринбург	Красноярск	Чита	Якутск
Численность населения	4000000	1000000	800000	600000	100000
Число больных	100000	20000	6000	5000	2000

Установить, существует ли взаимосвязь между этими параметрами. Если такая связь есть, определить тесноту связи.

*Задание 2.* Провести анализ корреляционной связи для выборок согласно индивидуальному заданию.

3.1. Выполнение задания 1. Оценим, существует ли взаимосвязь между числом жителей городов и числом больных в них.

Для оценки парного коэффициента корреляции в Excel используются функция =КОРРЕЛ(Массив 1; Массив 2) или процедура КОРРЕЛЯЦИЯ в Пакете анализа.

3.1.1. Создать ЭТ табл. 8. Активизировать ячейку **В13**.

3.1.2. Выполнить команды: **Вставка – Функция – Статистические – КОРРЕЛ – Ок.**

Таблица 8

	А	В	С
1	Оценка коэффициента корреляции		
2	Город	Численность населения (тыс. чел)	Число больных (тыс. чел)
3	Москва	8000	75
4	Рязань	300	10
5	Тула	200	8
6	Воронеж	400	15
7	Орел	450	20
8	С.-Перербург	4000	100
9	Екатеринбург	1000	20
10	Красноярск	800	6
11	Чита	600	5
12	Якутск	100	2

3.1.3. Ввести в окно

**КОРРЕЛ:**

– **Массив 1** B3:B12;

– **Массив 2** C3:C12;

**Ок.**

В ячейке **B13** – значение коэффициента корреляции, равное 0,83 (табл. 9).

• Таблица  
9

	А	В	С
1	Оценка коэффициента корреляции		
2	Город	Численность населения (тыс. чел.)	Число больных (тыс. чел.)
3	Москва	8000	75
4	Рязань	300	10
5	Тула	200	8
6	Воронеж	400	15
7	Орел	450	20
8	С.-Петербург	4000	100
9	Екатеринбург	1000	20
10	Красноярск	800	6
11	Чита	600	5
12	Якутск	100	2
13		<b>0,8340689</b>	
14		Столбец 1	Столбец 2
15	Столбец 1		1

16	Столбец 2	0,8340689	1
----	-----------	-----------	---

Корреляция считается тесной (сильной), если  $|r| > 0,7$ .

Если  $0,4 \leq |r| \leq 0,7$ , можно утверждать, что корреляционная связь есть, но не тесная.

При  $|r| < 0,4$  корреляционная связь отсутствует.

В нашем случае  $r > 0,7$  значит можно утверждать, что корреляционная связь тесная.

3.1.4. Проверим вычисления с использованием Пакета анализа:

а) Выполнить команды: **Данные – Анализ данных – Корреляция - Ок;**

б) Ввести:

- **Входной интервал** ;
- **Группировка** По столбцам;
- **Выходной интервал** ;
- **Ок.**

Программа помещает результат в ячейки A14:C16 табл. 9.

Парный коэффициент корреляции находится в ячейке B16. Как видим, результат полностью совпал с п.3.1.3.

### 3.3. Выполнение задания 2

Провести оценку корреляционной зависимости согласно индивидуальному заданию. Номер варианта выбирается по последней цифре шифра.

## **Индивидуальные задания**

**Вариант 1.** По группе акционерных коммерческих банков региона имеются следующие данные:

Номер банка	1	2	3	4	5	6	7	8	9	10
Активы банка, млн. руб.	866	328	207	185	109	104	327	113	91	849
Прибыль, млн.руб.	39,5	17,8	12,7	14,9	4,0	15,5	6,4	10,3	3,4	13,4

Определите, есть ли взаимосвязь между этими параметрами.

**Вариант 2..** По предприятиям имеются следующие данные о ёмкости электросталеплавильных печей и расходе электроэнергии на одну тонну выплавленной стали:

Номер предприятия	1	2	3	4	5	6	7	8	9	10
Ёмкость печи, т	1,0	1,5	1,0	10,0	10,0	5,0	2,0	1,5	3,5	1,0
Расход электроэнергии на 1 т стали, кВт-ч/т	924	909	1010	541	682	657	888	835	602	890

Определите, имеется ли взаимосвязь между этими параметрами?

**Вариант 3..** По группе однородных предприятий имеются следующие данные:

Номер предприятия	1	2	3	4	5	6	7	8	9	10
Уровень механизации работ, %	22	85	67	36	21	40	39	39	31	62
Объём продукции, млн. руб.	117	186	89	112	52	133	141	158	120	197

Определите, имеется ли взаимосвязь между этими параметрами?

**Вариант 4.** Поступление доходов в консолидированный бюджет Санкт-Петербурга приведено в таблице.

	Район города					
	Курорт- ный	Примо- рский	Невс- кий	Киров- ский	Выбор- гский	Крон- штадт
Число работа- ющих на пред- приятиях райо- на (тыс. чел.)	18	49	106	132	125	6
Доход (млрд. руб)	20,8	48,7	132,4	150,6	114,2	8,1

Определите, имеется ли взаимосвязь между этими параметрами?

**Вариант 5.** Определите, имеется ли взаимосвязь между рождаемостью и смертностью (количество на 1000 чел.) в С. Петербурге:

Годы	Рождаемость	Смертность
1991	9,3	12,5
1992	7,4	13,5
1993	6,5	17,4
1994	7,1	17,2
1995	7,0	15,9
1996	6,6	14,2

**Вариант 6.** Проверьте, имеется ли взаимосвязь между среднемесячными доходами на семью ( в тыс. руб.) и расходами на покупку кондитерских изделий (в руб.) ?

Доходы семьи	4,8	3,8	5,4	4,2	3,4	4,6	3,4	4,8	5,0	3,8	5,2	4,0	3,8	4,6	4,4
Расходы на кон- дитерск. изделия	75	68	78	71	64	73	66	75	75	65	77	69	67	72	70

**Вариант 7.** Имеются экспериментальные данные о времени вулканизации и сопротивлении резины разрыву.

Номер анализа	1	2	3	4	5	6	7	8	9	10
Время вулканизации, мин.	35	40	30	42	37	38	34	33	36	31
Сопротивление разрыву, кг/см <sup>2</sup>	162	174	155	172	173	166	162	160	167	153

Определите, имеется ли взаимосвязь между этими параметрами?

**Вариант 8.** В мартеновском цехе завода произведены испытания для определения зависимости производительности печи от содержания углерода в металле. Получены следующие результаты:

Номер анализа	1	2	3	4	5	6	7	8	9	10
Процент углерода в металле	0,9 5	0,9 8	0,6 5	0,9 4	0,9 9	0,7 8	0,8 2	1,1 2	0,9 2	1,1 2
Производительность печи, т/час	16, 3	16, 0	17, 4	16, 5	16, 0	16, 7	15, 8	16, 4	15, 7	16, 0

Определите, имеется ли взаимосвязь между этими параметрами?

**Вариант 9.** Определите, есть ли взаимосвязь между стоимостью квартиры, ее площадью и удаленностью от центра города:

Удаленность от центра, км	74	46	92	48	93	72	42	50	64	78
Общая площадь, кв. м.	56	70	29	69	25	60	71	68	65	49
Стоимость, тыс.руб.	44	69	27	78	30	48	79	65	57	43

**Вариант 10.** По группе предприятий имеются следующие данные:

Номер предприятия	1	2	3	4	5	6	7	8	9	10
Годовая производительность труда работника, тыс. руб.	360	298	328	330	366	316	334	300	314	320
Наличие современного оборудования, тыс.руб./чел	15,2	12,8	13,8	14,0	16,3	12,6	13,2	12,9	13,1	12,5
Текучесть кадров, %	9,1	10,1	5,0	7,0	9,0	4,0	12,1	6,5	8,0	7,0

Определите, есть ли взаимосвязь между этими параметрами.