ВОРОНЕЖСКИЙ ИНСТИТУТ МВД РОССИИ

**А. В. Меньших**

**М. А. Панкова**

# ОСНОВЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

*Практикум*

Воронеж
2025

ББК 22.171
УДК 519.2

Рецензенты:

А. В.Мельников – заведующий кафедрой правовой информатики, информационного права и естественно-научных дисциплин Центрального филиала ФГБОУ ВО «Российский государственный университет правосудия» (г. Воронеж), доктор технических наук, доцент.;

В. В. Корчагин– доцент кафедры математики и естественно-научных дисциплин ФКОУ ВО «Воронежский институт ФСИН России», кандидат технических наук, доцент.

**Меньших А. В.**

Основы статистического анализа данных : практикум [Электронный ресурс] / А. В. Меньших, М. А. Панкова. – Электр. дан. и прогр. – Воронеж : Воронежский институт МВД России, 2025. – 1 электр. опт. диск (CD-ROM) : 12 см. – Систем. требования: процессор Intel с частотой не менее 1,3 ГГц ; ОЗУ 512 Мб ; операц. система семейства Windows ; CD-ROM дисковод.

Практикум содержит систематическое изложение материала с примерами, задачами, лабораторными работами и заданиями для самостоятельной работы.

Издание предназначено для слушателей факультета переподготовки и повышения квалификации, а также для курсантов и слушателей радиотехнического факультета.

ISBN 978-5-00229-177-9

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ ..... 4
1. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ 5
Лабораторная работа № 1
Выборочный метод ..... 5
Лабораторная работа № 2
Законы распределения дискретных случайных величин ..... 21
Лабораторная работа № 3
Законы распределения непрерывных случайных величин ..... 36
2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ДАННЫХ 51
Лабораторная работа № 4
Элементы корреляционного анализа ..... 51
Лабораторная работа № 5
Парная регрессия ..... 56
Лабораторная работа № 6
Множественная регрессия ..... 65
3. ПРОГНОЗИРОВАНИЕ 79
Лабораторная работа № 7
Прогнозирование ..... 79
4. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ 108
Лабораторная работа № 8
Проверка статистических гипотез ..... 108
РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА ..... 116
Приложение 1 ..... 117
Приложение 2 ..... 118

# ВВЕДЕНИЕ

Практикум «Основы статистического анализа данных» включает в себя четыре раздела:

1. Основы математической статистики.

Рассмотрены основные понятия, связанные с выборочным методом, точечными оценками параметров распределения, дискретными и непрерывными случайными величинами, с законами их распределения и числовыми характеристиками, а также рассмотрены примеры и приведены индивидуальные задания.

2. Корреляционный анализ данных.

Содержится изложение основных понятий, связанных корреляционным и регрессионным анализом данных, а также рассмотрены примеры, приведены индивидуальные задания.

3. Прогнозирование.

Рассмотрены основные понятия, связанные с прогнозированием статистических данных, структурой временного ряда, а также рассмотрены примеры, приведены индивидуальные задания.

4. Проверка статистических гипотез.

Кроме того, важной составной частью практикума являются задания для самостоятельного выполнения по каждому типу задач. Их количество позволяет использовать их при проведении групповых занятий, выдавать в качестве индивидуальных заданий в типовом расчёте.

# 1. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

## Лабораторная работа № 1
## Выборочный метод

**Цель лабораторной работы:** изучить основные понятия, связанные с выборочным методом, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

*Определение.* **Выборочной совокупностью (выборкой)** называется совокупность случайно отобранных объектов.

*Определение.* **Генеральной совокупностью** называют совокупность объектов, из которых производится выборка.

*Определение.* **Объемом совокупности** (**выборочной** или **генеральной**) называют число объектов этой совокупности.

*Определение.* **Статистическим рядом** для выборки называют таблицу, которая в первой строке содержит значения $z_{(1)},...,z_{(m)}$, а во второй – числа их повторений.

*Определение.* Число $n_i$, $i=\overline{1,m}$, показывающее, сколько раз элемент $z_{(i)}$ встречался в выборке, называется **частотой**, а отношение $\frac{n_i}{n}$ – **относительной частотой** этого значения.

*Определение.* **Полигоном частот** называют ломаную линию, отрезки которой соединяют точки $(x_1,n_1)$, $(x_2,n_2)$, …, $(x_m,n_m)$.

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

Полигон относительных частот строится аналогично, за исключением того, что на оси ординат откладываются относительные частоты $\frac{n_i}{n}$.

*Определение.* Последовательность чисел $x_{(1)},x_{(2)},...,x_{(i)},...,x_{(n)}$, удовлетворяющих условию $x_{(1)}\le x_{(2)}\le...\le x_{(i)}\le...\le x_{(n)}$, называется **вариационным рядом выборки** или просто **вариационным рядом**; число $x_{(i)}$, $i=\overline{1,n}$, называется i-ым **членов вариационного ряда.**

*Определение.* **Мода** представляет собой значение изучаемого признака, повторяющееся с наибольшей частотой.

*Определение.* **Медианой** называется значение признака, приходящееся на середину ранжированной (упорядоченной) совокупности.

*Определение.* **Размахом** называется разность между наибольшим и наименьшим значениями ряда данных.

## Точечные оценки параметров распределения

### Выборочная средняя

Пусть для изучения генеральной совокупности относительно количественного признака X извлечена выборка объема n.

Выборочной средней называют среднее арифметическое значение признака выборочной совокупности.

Если все значения признака выборки различны, то

$$\bar{x}_в = \frac{x_1 + x_2 + ... + x_n}{n};$$

если же все значения имеют частоты $n_1$, $n_2$,…,$n_k$, то

$$\bar{x}_в = \frac{x_1 n_1 + ... + x_k n_k}{n}.$$

### Выборочная дисперсия

Для того чтобы наблюдать рассеяние количественного признака значений выборки вокруг своего среднего значения, вводят сводную характеристику – выборочную дисперсию.

Выборочной дисперсией называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения.

Если все значения признака выборки различны, то

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

если же все значения имеют частоты $n_1$, $n_2$, …, $n_k$, то

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

Для характеристики рассеивания значений признака выборки вокруг своего среднего значения пользуются сводной характеристикой – **средним квадратическим отклонением**.

Выборочным средним квадратическим отклонением называют квадратный корень из выборочной дисперсии:

$$\hat{\sigma}_в = \sqrt{\hat{D}_в}.$$

### Исправленная дисперсия

$$S^2 = \frac{n}{n+1} \hat{D}_в = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_в)^2}{n-1}$$

Для оценки среднего квадратического генеральной совокупности используют исправленное **среднее квадратическое отклонение**

$$S = \sqrt{S^2}\,.$$

*Пример 1.1*. В результате тестирования группа курсантов набрала баллы: 5, 3, 0, 1, 4, 2, 5, 4, 1, 5.

а) Построить вариационный ряд.

б) Посчитать выборочные характеристики.

**Решение с помощью табличного процессора.**

а) Простейший способ упорядочения массива данных предоставляется опцией Excel «*Данные*»→ «*Сортировка*».

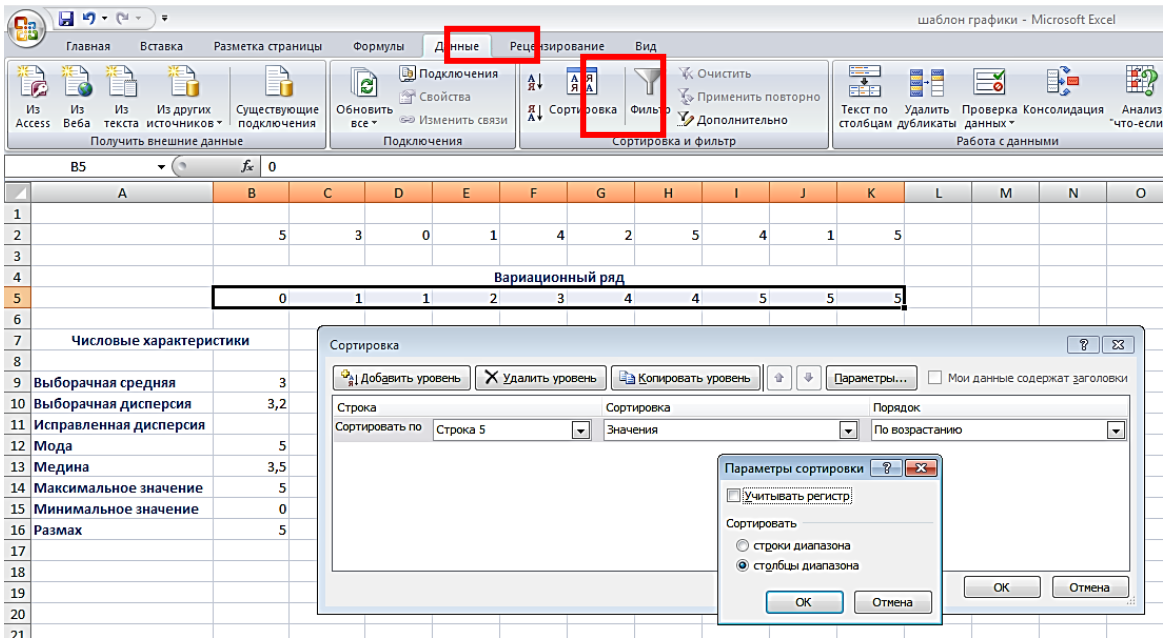



Рис. 1.1

б) Для расчета числовых характеристик используем опцию «*Вставка*»→«*Функция*»→«*Статистические*».

| | |
|---|---|
| Выборочная средняя | *=СРЗНАЧ( )* |
| Выборочная дисперсия | *=ДИСПР( )* |
| Исправленная дисперсия | *=ДИСПА( )* |
| Среднеквадратическое отклонение | *=СТАНДОТКЛОН( )* |
| Мода | *=МОДА( )* |
| Медиана | *=МЕДИАНА( )* |
| Минимальное значение | *=МИН( )* |
| Максимальное значение | *=МАКС( )* |
| Размах ряда | *=МАКС - МИН* |

Рассмотрим на примере нахождения функции «*МОДА*».

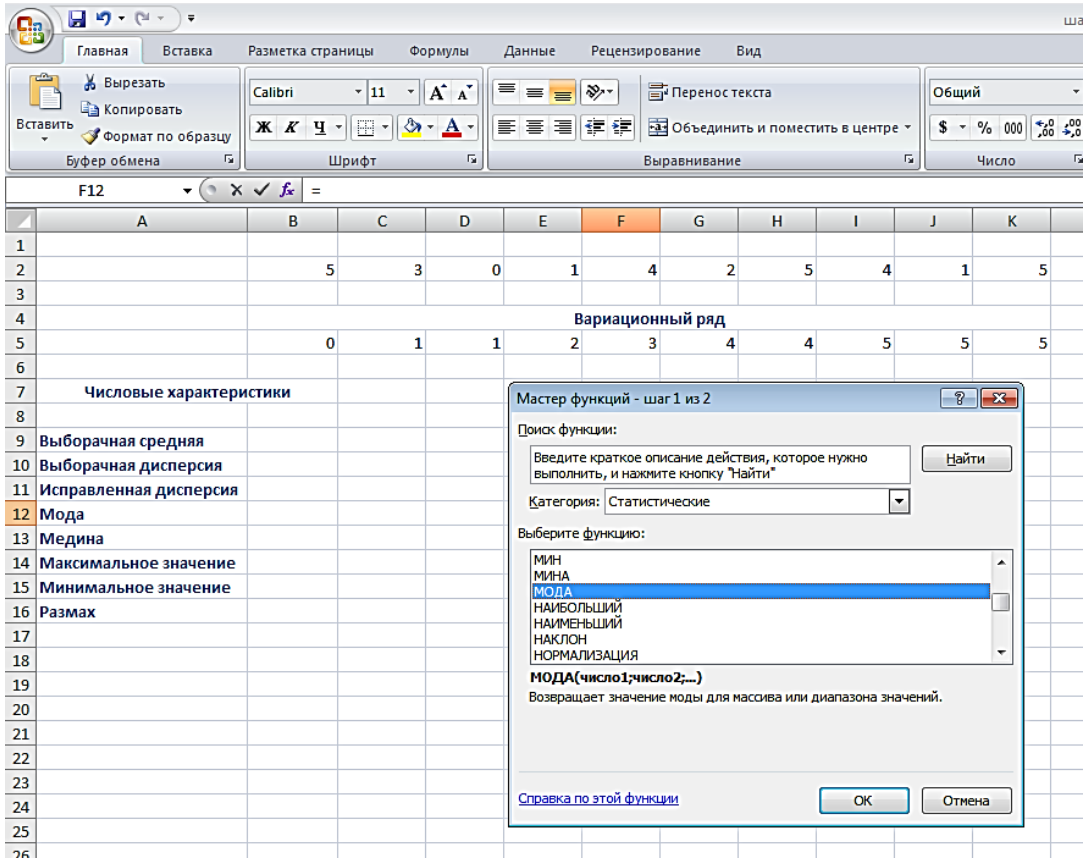



Рис. 1.2

В поле *«Число 1»* ставим курсор и мышкой выделяем нашу таблицу. После чего нажимаем *«ОК»*.

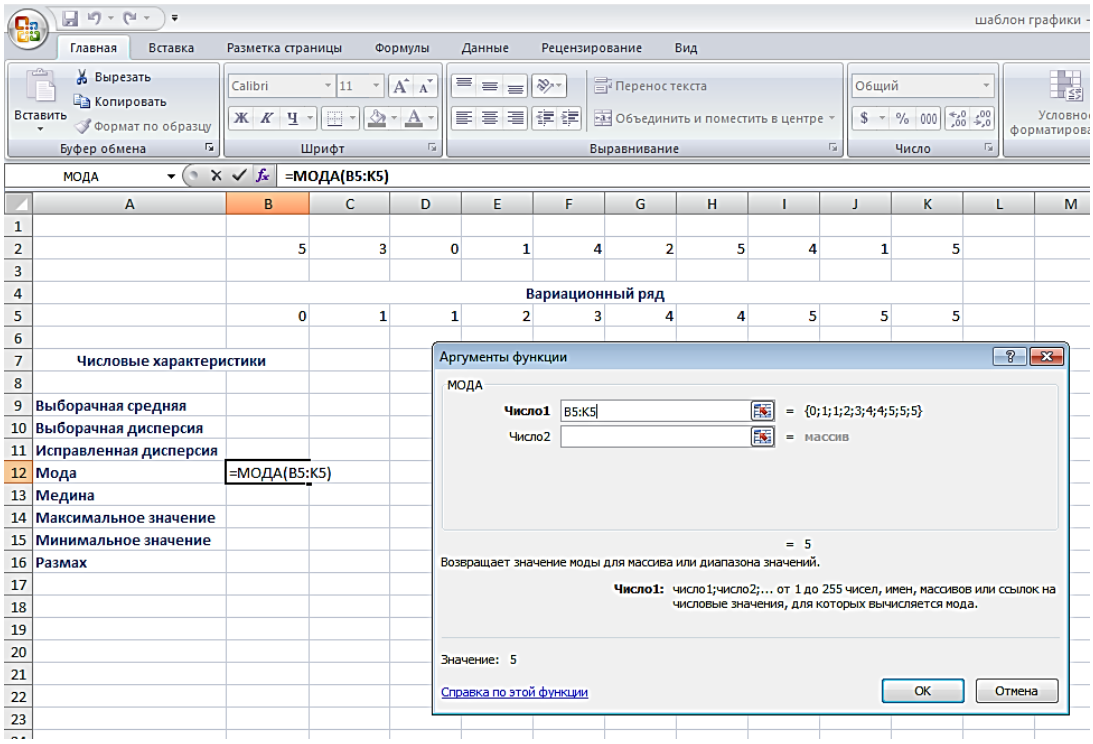



Рис. 1.3

*Пример 1.2.* Проведена проверка в 100 компаниях.
Даны значения количества работающих в компании (чел.):

23 25 24 25 30 24 30 26 28 26
32 33 31 31 25 33 25 29 30 28
23 30 29 24 33 30 30 28 26 25
26 29 27 29 26 28 27 26 29 28
29 30 27 30 28 32 28 26 30 26
31 27 30 27 33 28 26 30 31 29
27 30 30 29 27 26 28 31 29 28
33 27 30 33 26 31 34 28 32 22
29 30 27 29 34 29 32 29 29 30
29 29 36 29 29 34 23 28 24 28

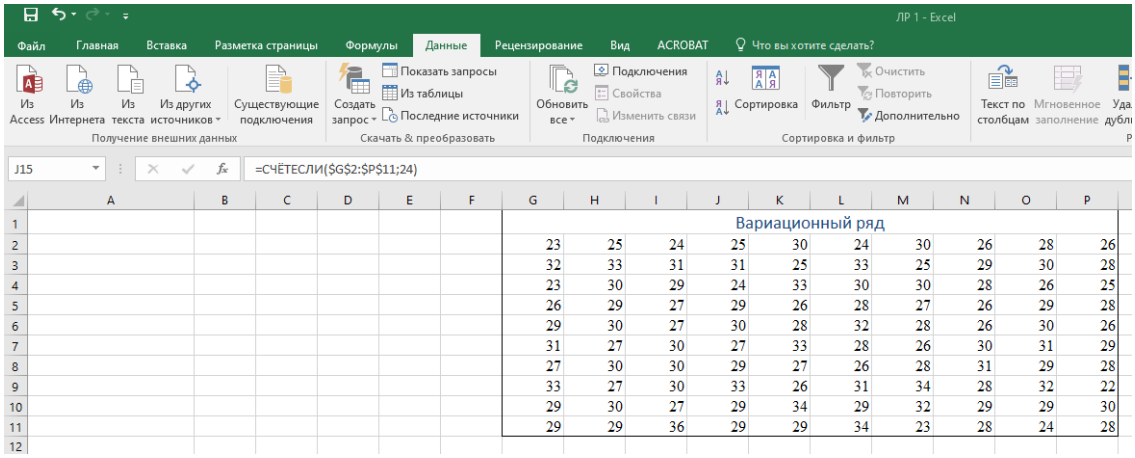
Рассчитать числовые характеристики:

- выборочное среднее;
- выборочная дисперсия;
- исправленная дисперсия;
- среднеквадратическое отклонение;
- моду;
- медиану;
- минимальное значение;
- максимальное значение;
- размах ряда.

Записать данные в виде статистического ряда.
Построить полигон частот и относительных частот.
Раскрыть смысловую сторону каждой характеристики.

**Решение с помощью табличного процессора.**

1. Занести данные в табличный процессор, каждое число в отдельную ячейку.




| | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 23 | 25 | 24 | 25 | 30 | 24 | 30 | 26 | 28 | 26 |
| 3 | 32 | 33 | 31 | 31 | 25 | 33 | 25 | 29 | 30 | 28 |
| 4 | 23 | 30 | 29 | 24 | 33 | 30 | 30 | 28 | 26 | 25 |
| 5 | 26 | 29 | 27 | 29 | 26 | 28 | 27 | 26 | 29 | 28 |
| 6 | 29 | 30 | 27 | 30 | 28 | 32 | 28 | 26 | 30 | 26 |
| 7 | 31 | 27 | 30 | 27 | 33 | 28 | 26 | 30 | 31 | 29 |
| 8 | 27 | 30 | 30 | 29 | 27 | 26 | 28 | 31 | 29 | 28 |
| 9 | 33 | 27 | 30 | 33 | 26 | 31 | 34 | 28 | 32 | 22 |
| 10 | 29 | 30 | 27 | 29 | 34 | 29 | 32 | 29 | 29 | 30 |
| 11 | 29 | 29 | 36 | 29 | 29 | 34 | 23 | 28 | 24 | 28 |

Рис. 1.4

2. Для расчета числовых характеристик используем опцию «*Вставка*»→«*Функция*»→«*Статистические*».

| | | |
|---|---|---|
| Выборочная средняя | *=СРЗНАЧ( )* | *=СУММПРОИЗВ(xi;wi)* |
| Выборочная дисперсия | *=ДИСПР( )* | |
| Исправленная дисперсия | *=ДИСПА( )* | |
| Среднеквадратическое отклонение | *=СТАНДОТКЛОН( )* | *=КОРЕНЬ(ДИСПА)* |
| Мода | *=МОДА( )* | |
| Медиана | *=МЕДИАНА( )* | |
| Минимальное значение | *=МИН( )* | |
| Максимальное значение | *=МАКС( )* | |
| Размах ряда | *=МАКС - МИН* | |

Рассмотрим на примере нахождения функции «*МОДА*».

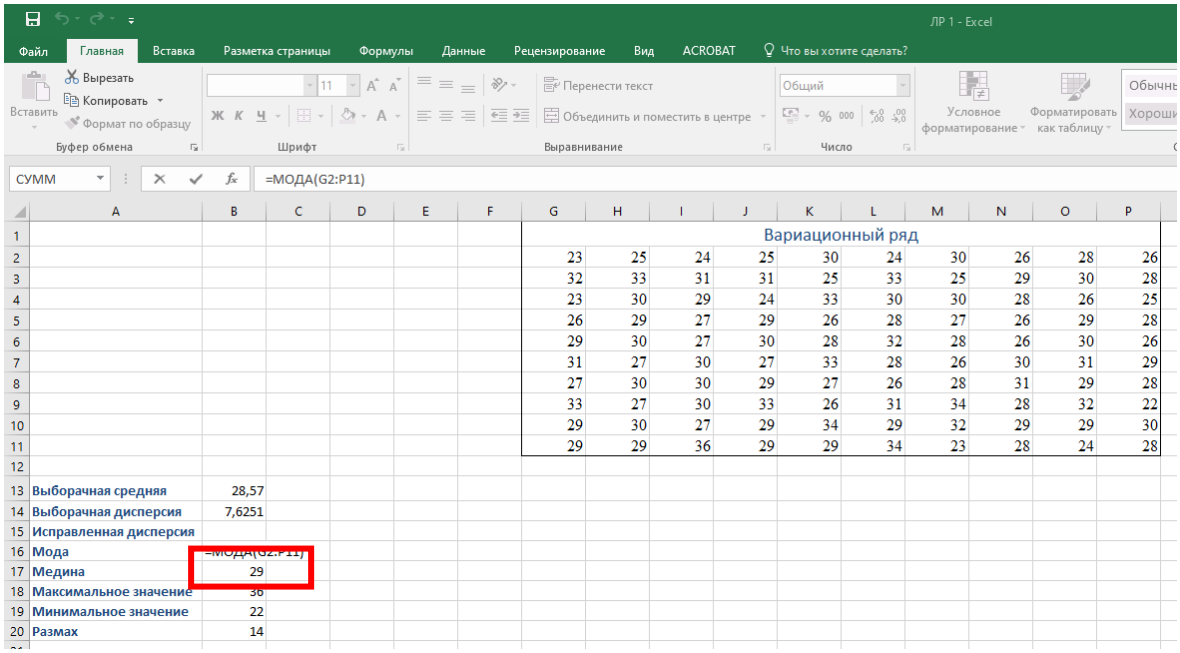



Рис. 1.5

Получили *Mo* = 29 (чел) – фирм, у которых в штате 29 человек, больше всего.

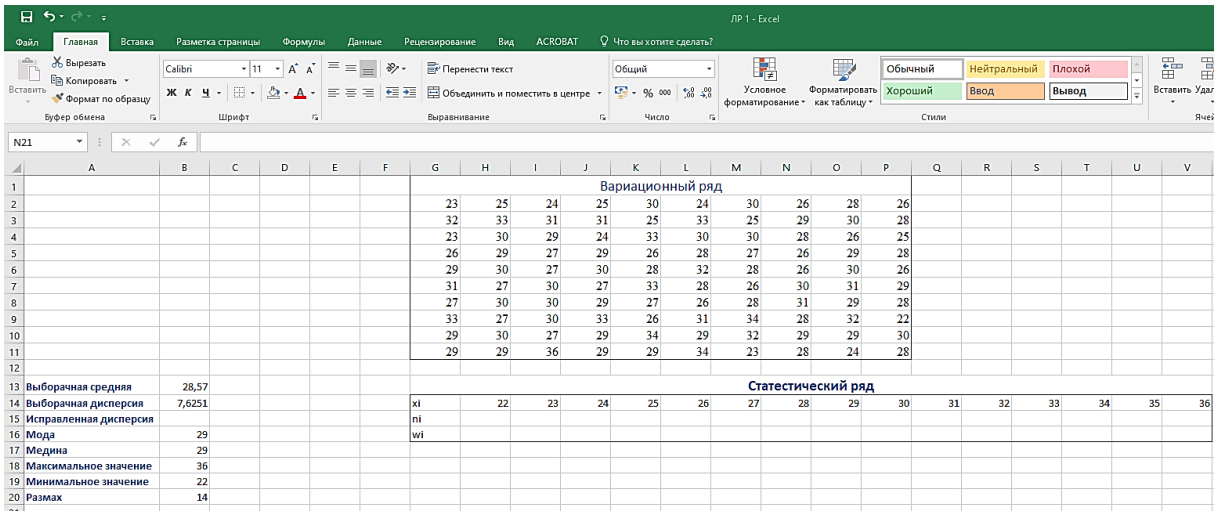
Используя тот же путь, вычисляем медиану.

Получили *Me* = 29 (чел) – среднее значение сотрудников в фирме.

Размах ряда чисел – разница между наименьшим и наибольшим возможным значением случайной величины. Для вычисления размаха ряда нужно найти наибольшее и наименьшее значения нашей выборки и вычислить их разность.

Таким образом, разница между фирмой с наибольшим штатом сотрудников и фирмой с наименьшим штатом сотрудников составила *МАКС – МИН*=36 – 22 = 14 (чел).

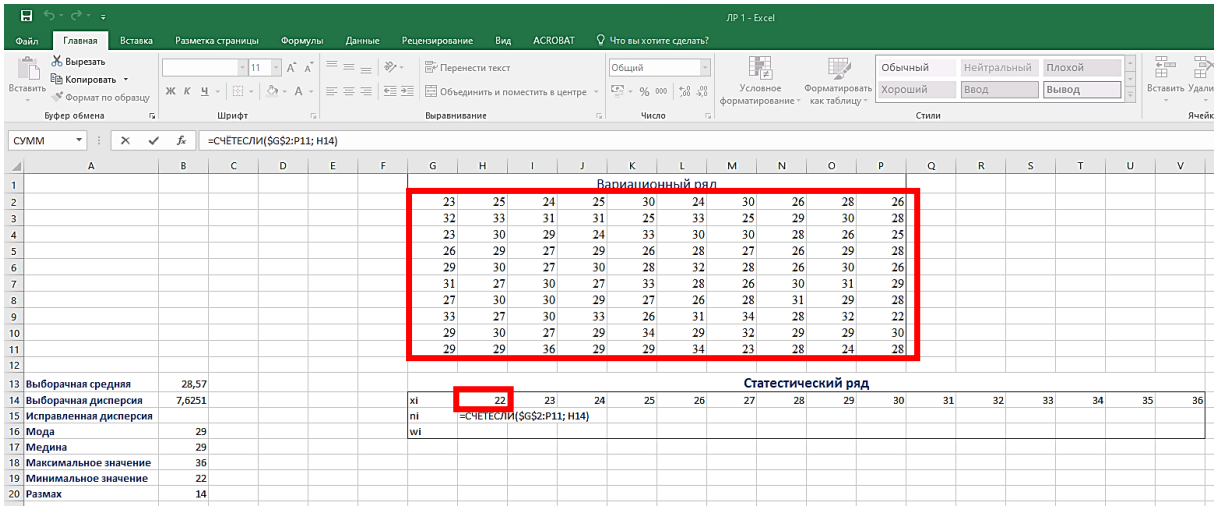
Для построения полигона частот и относительных частот необходимо задать закон распределения, т.е. составить таблицу значений случайной величины и соответствующих им частот. Мы уже знаем, что наименьшее число сотрудников в фирме = 22, а наибольшее = 36. Составим таблицу, в которой значения $x_i$ случайной величины меняются от минимального значения 22 до максимального значения 36 шагом 1.



| | A | B | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | Вариационный ряд | | | | | | | | | |
| 2 | | | 23 | 25 | 24 | 25 | 30 | 24 | 30 | 26 | 28 | 26 |
| 3 | | | 32 | 33 | 31 | 31 | 25 | 33 | 25 | 29 | 30 | 28 |
| 4 | | | 23 | 30 | 29 | 24 | 33 | 30 | 30 | 28 | 26 | 25 |
| 5 | | | 26 | 29 | 27 | 29 | 26 | 28 | 27 | 26 | 29 | 28 |
| 6 | | | 29 | 30 | 27 | 30 | 28 | 32 | 28 | 26 | 30 | 26 |
| 7 | | | 31 | 27 | 30 | 27 | 33 | 28 | 26 | 30 | 31 | 29 |
| 8 | | | 27 | 30 | 30 | 29 | 27 | 26 | 28 | 31 | 29 | 28 |
| 9 | | | 33 | 27 | 30 | 33 | 26 | 31 | 34 | 28 | 32 | 22 |
| 10 | | | 29 | 30 | 27 | 29 | 34 | 29 | 32 | 29 | 29 | 30 |
| 11 | | | 29 | 29 | 36 | 29 | 29 | 34 | 23 | 28 | 24 | 28 |
| 12 | | | | | | | | | | | | |
| 13 | Выборочная средняя | 28,57 | Статистический ряд | | | | | | | | | |
| 14 | Выборочная дисперсия | 7,6251 | xi | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 15 | Исправленная дисперсия | | ni | | | | | | | | | |
| 16 | Мода | 29 | wi | | | | | | | | | |
| 17 | Медина | 29 | | | | | | | | | | |
| 18 | Максимальное значение | 36 | | | | | | | | | | |
| 19 | Минимальное значение | 22 | | | | | | | | | | |
| 20 | Размах | 14 | | | | | | | | | | |

Рис. 1.6

Чтобы сосчитать частоту каждого значения, воспользуемся *Вставка – Функция – Статистические – СЧЕТЕСЛИ.* В окне Диапазон ставим курсор и выделяем нашу выборку, а в окне Критерий ссылаемся на ячейку, содержащую значение случайной величины 22.



| | A | B | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | Вариационный ряд | | | | | | | | | |
| 2 | | | 23 | 25 | 24 | 25 | 30 | 24 | 30 | 26 | 28 | 26 |
| 3 | | | 32 | 33 | 31 | 31 | 25 | 33 | 25 | 29 | 30 | 28 |
| 4 | | | 23 | 30 | 29 | 24 | 33 | 30 | 30 | 28 | 26 | 25 |
| 5 | | | 26 | 29 | 27 | 29 | 26 | 28 | 27 | 26 | 29 | 28 |
| 6 | | | 29 | 30 | 27 | 30 | 28 | 32 | 28 | 26 | 30 | 26 |
| 7 | | | 31 | 27 | 30 | 27 | 33 | 28 | 26 | 30 | 31 | 29 |
| 8 | | | 27 | 30 | 30 | 29 | 27 | 26 | 28 | 31 | 29 | 28 |
| 9 | | | 33 | 27 | 30 | 33 | 26 | 31 | 34 | 28 | 32 | 22 |
| 10 | | | 29 | 30 | 27 | 29 | 34 | 29 | 32 | 29 | 29 | 30 |
| 11 | | | 29 | 29 | 36 | 29 | 29 | 34 | 23 | 28 | 24 | 28 |
| 12 | | | | | | | | | | | | |
| 13 | Выборочная средняя | 28,57 | Статистический ряд | | | | | | | | | |
| 14 | Выборочная дисперсия | 7,6251 | xi | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 15 | Исправленная дисперсия | | ni | =СЧЁТЕСЛИ($G$2:P11; H14) | | | | | | | | |
| 16 | Мода | 29 | wi | | | | | | | | | |
| 17 | Медина | 29 | | | | | | | | | | |
| 18 | Максимальное значение | 36 | | | | | | | | | | |
| 19 | Минимальное значение | 22 | | | | | | | | | | |
| 20 | Размах | 14 | | | | | | | | | | |

Рис. 1.7

Нажимаем клавишу ОК, получаем значение 1, т.е. число 22 в нашей выборке встречается 1 раз и его частота =1. Аналогичным образом заполняем всю таблицу.

Для проверки вычисляем объем выборки, сумму частот (Вставка – Функция – Математические – СУММА). Должно получиться 100 (количество всех фирм).

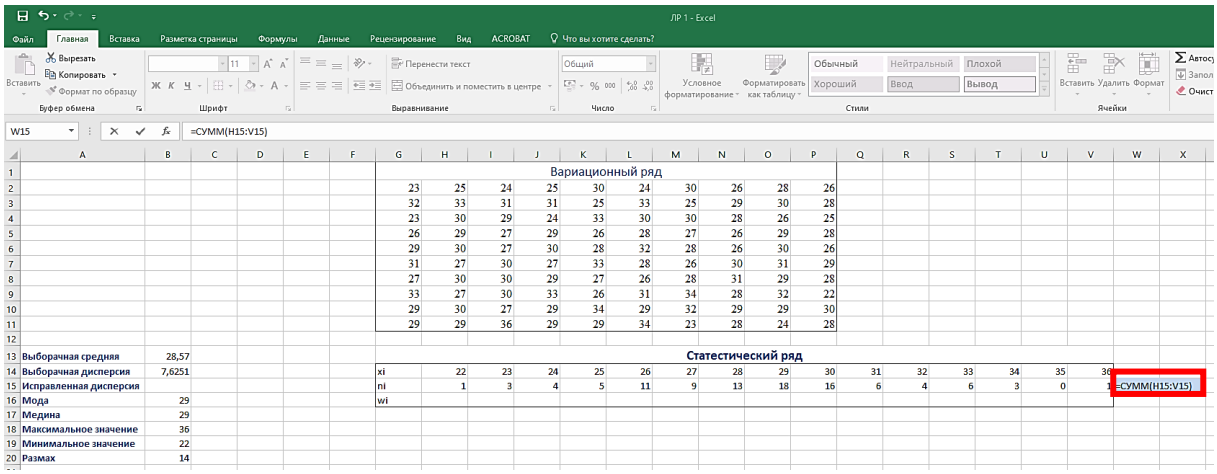



Рис. 1.8

Относительная частота находится по формуле ni/100, где 100 – объем выборки.

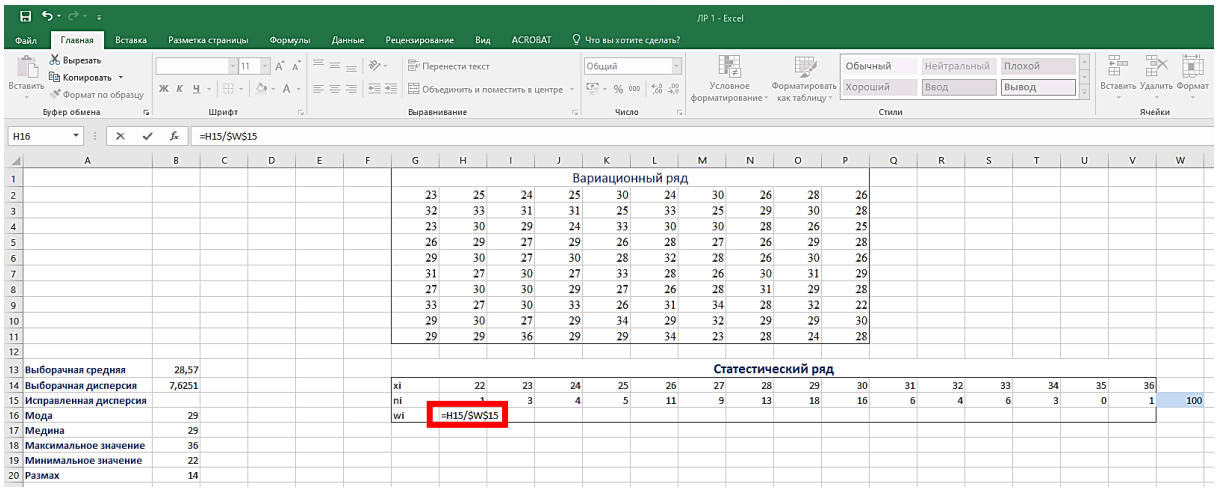



Рис. 1.9

Проверка: сумма всех wi равняется 1.

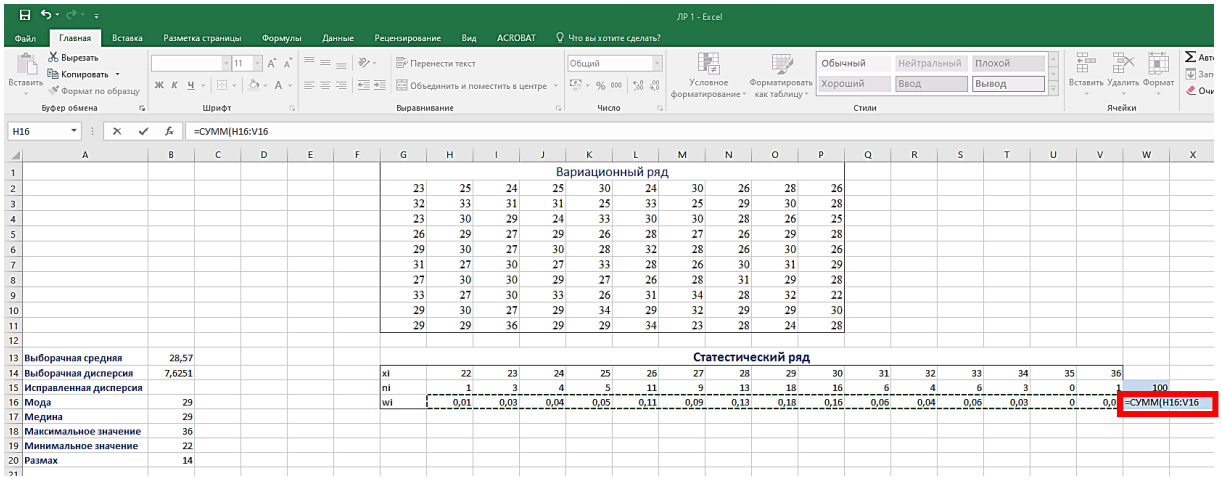



Рис. 1.10

Чтобы построить полигон частот, выделяем таблицу – Вставка – Диаграмма – Стандартные – Точечная (точечная диаграмма, на которой значения соединены отрезками).

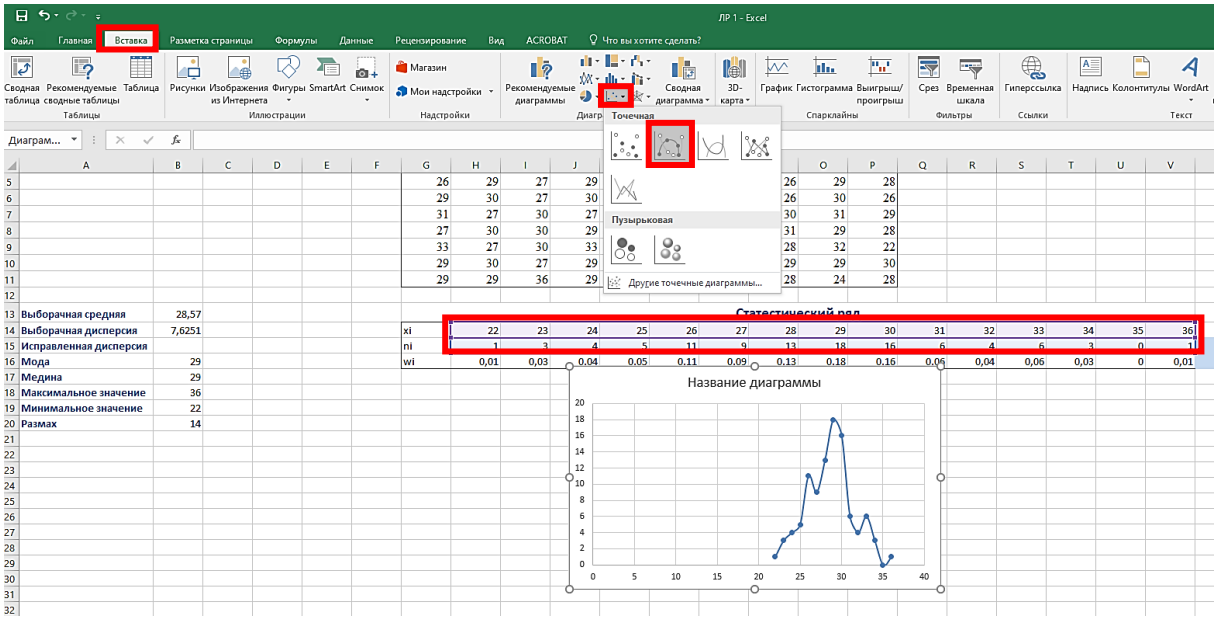



Рис. 1.11

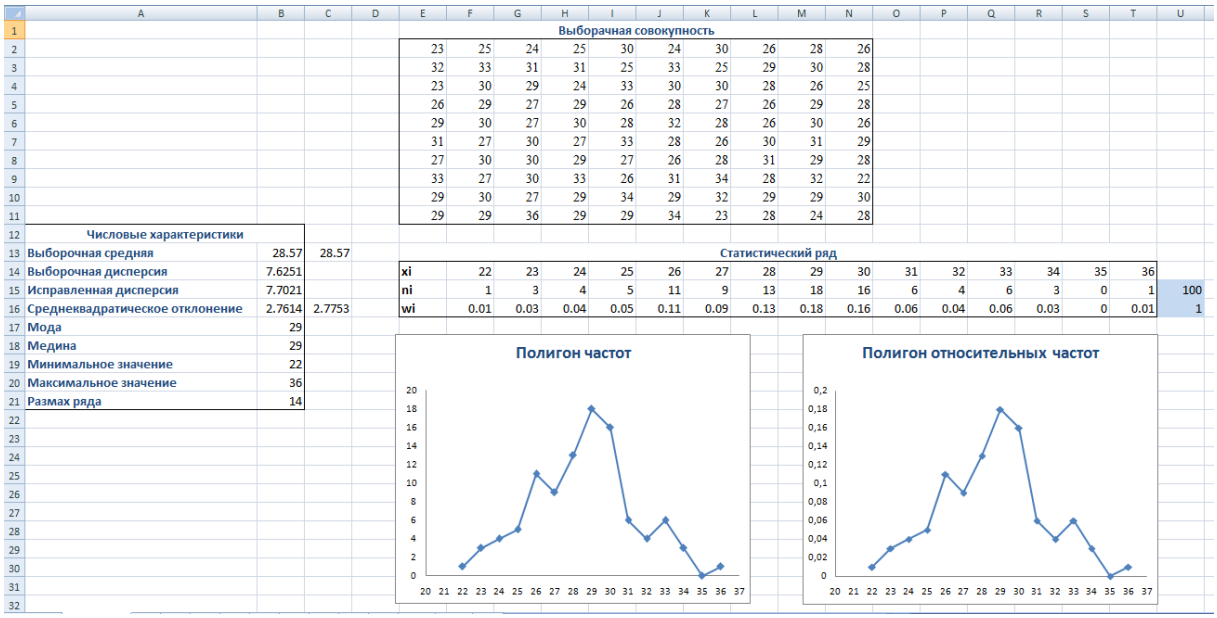
Получаем:


Рис. 1.12

Аналогично строим полигон относительных частот.


Рис. 1.13

Образец оформления задачи:



Выборочная совокупность

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 23 | 25 | 24 | 25 | 30 | 24 | 30 | 26 | 28 | 26 |
| 32 | 33 | 31 | 31 | 25 | 33 | 25 | 29 | 30 | 28 |
| 23 | 30 | 29 | 24 | 33 | 30 | 30 | 28 | 26 | 25 |
| 26 | 29 | 27 | 29 | 26 | 28 | 27 | 26 | 29 | 28 |
| 29 | 30 | 27 | 30 | 28 | 32 | 28 | 26 | 30 | 26 |
| 31 | 27 | 30 | 27 | 33 | 28 | 26 | 30 | 31 | 29 |
| 27 | 30 | 30 | 29 | 27 | 26 | 28 | 31 | 29 | 28 |
| 33 | 27 | 30 | 33 | 26 | 31 | 34 | 28 | 32 | 22 |
| 29 | 30 | 27 | 29 | 34 | 29 | 32 | 29 | 29 | 30 |
| 29 | 29 | 36 | 29 | 29 | 34 | 23 | 28 | 24 | 28 |

Числовые характеристики

| | | |
|---|---|---|
| Выборочная средняя | 28.57 | 28.57 |
| Выборочная дисперсия | 7.6251 | |
| Исправленная дисперсия | 7.7021 | |
| Среднеквадратическое отклонение | 2.7614 | 2.7753 |
| Мода | 29 | |
| Медина | 29 | |
| Минимальное значение | 22 | |
| Максимальное значение | 36 | |
| Размах ряда | 14 | |

Статистический ряд

| xi | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ni | 1 | 3 | 4 | 5 | 11 | 9 | 13 | 18 | 16 | 6 | 4 | 6 | 3 | 0 | 1 | 100 |
| wi | 0.01 | 0.03 | 0.04 | 0.05 | 0.11 | 0.09 | 0.13 | 0.18 | 0.16 | 0.06 | 0.04 | 0.06 | 0.03 | 0 | 0.01 | 1 |



Рис. 1.14

## Задания для самостоятельного выполнения

**Задание 1.** Построить вариационный ряд.

**Задание 2.** Посчитать выборочные характеристики.

| | | | |
|---|---|---|---|
| **1)** | 1, 1, 6, 3, 8, 3, 6, 2, 8, 5, 9, 1, 9 | **16)** | 3, 7, 2, 2, 3, 8, 9, 8, 5, 3, 3, 2, 4 |
| **2)** | 9, 4, 6, 2, 3, 2, 2, 5, 7, 5, 9, 1, 5 | **17)** | 8, 8, 8, 2, 4, 3, 1, 8, 8, 3, 3, 2, 7 |
| **3)** | 4, 2, 7, 7, 7, 4, 5, 2, 6, 1, 5, 1, 1 | **18)** | 4, 5, 2, 4, 7, 2, 2, 2, 7, 4, 9, 2, 1 |
| **4)** | 3, 7, 2, 6, 3, 8, 9, 8, 5, 3, 3, 2, 4 | **19)** | 3, 2, 2, 6, 2, 8, 2, 8, 5, 2, 3, 2, 2 |
| **5)** | 8, 8, 8, 2, 4, 3, 1, 8, 8, 3, 3, 2, 7 | **20)** | 7, 8, 7, 2, 4, 3, 1, 7, 8, 3, 7, 2, 7 |
| **6)** | 4, 5, 2, 4, 7, 2, 1, 2, 7, 4, 9, 2, 1 | **21)** | 2, 2, 2, 4, 7, 2, 1, 2, 7, 2, 9, 2, 1 |
| **7)** | 4, 2, 7, 7, 7, 4, 5, 2, 6, 1, 5, 1, 6 | **22)** | 4, 2, 9, 9, 9, 4, 5, 2, 6, 9, 5, 9, 6 |
| **8)** | 9, 4, 6, 2, 3, 2, 2, 1, 7, 5, 6, 1, 5 | **23)** | 5, 5, 6, 5, 3, 5, 2, 1, 5, 5, 6, 1, 5 |
| **9)** | 1, 3, 6, 1, 9, 3, 6, 2, 8, 5, 6, 1, 6 | **24)** | 1, 1, 6, 1, 9, 3, 1, 2, 1, 5, 1, 1, 6 |
| **10)** | 5, 5, 2, 4, 7, 2, 5, 2, 7, 5, 9, 5, 1 | **25)** | 5, 2, 2, 4, 5, 5, 5, 2, 7, 5, 9, 5, 5 |
| **11)** | 8, 7, 8, 2, 3, 8, 9, 8, 5, 3, 8, 2, 4 | **26)** | 0, 1, 6, 1, 0, 3, 6, 2, 8, 5, 9, 1, 9 |
| **12)** | 8, 9, 8, 2, 4, 3, 9, 8, 8, 3, 9, 2, 7 | **27)** | 9, 1, 6, 0, 3, 2, 2, 5, 7, 5, 9, 1, 5 |
| **13)** | 4, 1, 2, 2, 7, 2, 2, 2, 7, 4, 9, 2, 2 | **28)** | 1, 2, 7, 7, 0, 4, 5, 1, 6, 1, 0, 1, 1 |
| **14)** | 9, 2, 1, 9, 2, 8, 2, 8, 5, 2, 9, 2, 9 | **29)** | 3, 0, 2, 1, 3, 8, 9, 8, 1, 3, 3, 2, 0 |
| **15)** | 0, 8, 7, 2, 0, 3, 1, 0, 8, 1, 7, 2, 0 | **30)** | 8, 8, 0, 2, 4, 3, 0, 8, 8, 3, 3, 2, 7 |

**Задание 3.** Рассчитать числовые характеристики:

- выборочное среднее;
- выборочная дисперсия;
- исправленная дисперсия;
- среднеквадратическое отклонение;
- моду;
- медиану;
- минимальное значение;
- максимальное значение;
- размах ряда.

Записать данные в виде статистического ряда.
Построить полигон частот и относительных частот.
Раскрыть смысловую сторону каждой характеристики.

**Вариант 1.**

Проведен медицинский осмотр. Ниже приведены года рождения обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1990 | 1997 | 1994 | 1996 | 1989 | 1989 | 1997 | 1997 | 1993 | 1994 |
| 1999 | 1996 | 1998 | 1995 | 1995 | 1992 | 1994 | 1999 | 1997 | 1998 |
| 1995 | 1994 | 1989 | 1993 | 1990 | 1989 | 1990 | 1999 | 1992 | 1998 |
| 1989 | 1989 | 1993 | 1994 | 1992 | 1992 | 1990 | 1994 | 1992 | 1998 |
| 1990 | 1993 | 1995 | 1999 | 1997 | 1991 | 1991 | 1998 | 1995 | 1998 |
| 1999 | 1992 | 1989 | 1992 | 1989 | 1994 | 1989 | 1993 | 1995 | 1999 |
| 1992 | 1999 | 1999 | 1992 | 1992 | 1994 | 1989 | 1994 | 1991 | 1989 |
| 1990 | 1990 | 1992 | 1994 | 1999 | 1993 | 1992 | 1999 | 1989 | 1993 |
| 1991 | 1989 | 1994 | 1995 | 1993 | 1997 | 1990 | 1994 | 1992 | 1998 |
| 1989 | 1993 | 1999 | 1989 | 1993 | 1992 | 1996 | 1999 | 1993 | 1996 |

**Вариант 2.**

Проведен медицинский осмотр учащихся. Приведены результаты измерения роста обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 175 | 170 | 174 | 172 | 170 | 173 | 170 | 178 | 182 | 183 |
| 179 | 170 | 174 | 185 | 185 | 168 | 174 | 173 | 176 | 177 |
| 173 | 174 | 171 | 181 | 172 | 173 | 177 | 170 | 171 | 172 |
| 173 | 170 | 170 | 181 | 172 | 181 | 174 | 185 | 179 | 181 |
| 178 | 180 | 185 | 185 | 179 | 179 | 180 | 183 | 181 | 175 |
| 174 | 184 | 172 | 171 | 183 | 183 | 179 | 168 | 178 | 177 |
| 178 | 168 | 184 | 170 | 185 | 172 | 182 | 175 | 168 | 180 |
| 178 | 171 | 172 | 182 | 179 | 169 | 173 | 171 | 179 | 174 |
| 168 | 185 | 181 | 182 | 169 | 172 | 168 | 177 | 180 | 184 |
| 178 | 181 | 183 | 179 | 178 | 185 | 173 | 181 | 175 | 179 |

**Вариант 3.**

Выпускники школы сдавали ЕГЭ. Результаты ЕГЭ приведены ниже:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 62 | 59 | 71 | 58 | 71 | 69 | 69 | 65 | 65 | 63 |
| 67 | 61 | 59 | 67 | 70 | 64 | 69 | 64 | 58 | 70 |
| 59 | 67 | 59 | 66 | 61 | 62 | 71 | 71 | 68 | 67 |
| 60 | 68 | 64 | 68 | 60 | 66 | 66 | 62 | 59 | 58 |
| 71 | 62 | 64 | 71 | 65 | 60 | 62 | 58 | 69 | 70 |
| 58 | 66 | 69 | 66 | 60 | 68 | 63 | 67 | 64 | 59 |
| 58 | 61 | 61 | 71 | 59 | 63 | 67 | 66 | 62 | 60 |
| 68 | 58 | 66 | 61 | 65 | 71 | 60 | 59 | 61 | 64 |
| 60 | 68 | 62 | 58 | 61 | 70 | 63 | 61 | 62 | 67 |
| 69 | 67 | 68 | 62 | 71 | 62 | 58 | 67 | 65 | 66 |

**Вариант 4.**

Было задано 100 задач по математике. Ниже приведены результаты проверки (количество правильно решенных задач):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 60 | 65 | 70 | 66 | 67 | 63 | 62 | 68 | 68 | 66 |
| 60 | 68 | 70 | 61 | 68 | 62 | 64 | 70 | 61 | 66 |
| 71 | 71 | 67 | 69 | 65 | 68 | 61 | 61 | 65 | 65 |
| 71 | 70 | 65 | 65 | 62 | 69 | 58 | 66 | 66 | 69 |
| 68 | 69 | 60 | 69 | 68 | 68 | 60 | 58 | 58 | 68 |
| 67 | 71 | 62 | 58 | 58 | 68 | 60 | 71 | 61 | 69 |
| 58 | 70 | 58 | 59 | 68 | 62 | 71 | 65 | 64 | 62 |
| 58 | 71 | 66 | 59 | 63 | 61 | 58 | 69 | 61 | 62 |
| 59 | 60 | 68 | 67 | 67 | 63 | 67 | 66 | 66 | 61 |
| 68 | 66 | 71 | 67 | 71 | 58 | 68 | 68 | 69 | 69 |

**Вариант 5.**

Процент прибывших по сигналу «ТРЕВОГА» в подразделениях составил:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 87 | 93 | 93 | 91 | 81 | 98 | 86 | 90 | 87 | 94 |
| 85 | 89 | 87 | 84 | 90 | 84 | 96 | 97 | 99 | 100 |
| 83 | 85 | 88 | 90 | 97 | 91 | 90 | 99 | 84 | 100 |
| 95 | 99 | 98 | 96 | 87 | 84 | 83 | 97 | 97 | 93 |
| 95 | 86 | 91 | 83 | 82 | 89 | 84 | 91 | 96 | 83 |
| 81 | 97 | 90 | 81 | 91 | 89 | 87 | 91 | 89 | 91 |
| 84 | 84 | 96 | 96 | 87 | 99 | 98 | 85 | 81 | 100 |
| 98 | 100 | 97 | 93 | 100 | 85 | 97 | 83 | 94 | 88 |
| 85 | 92 | 86 | 92 | 92 | 94 | 94 | 86 | 95 | 89 |
| 89 | 89 | 88 | 97 | 91 | 94 | 86 | 82 | 84 | 84 |

**Вариант 6.**

Проведен медицинский осмотр. Ниже приведен возраст обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 29 | 23 | 27 | 23 | 30 | 22 | 24 | 24 | 29 | 30 |
| 20 | 32 | 28 | 32 | 24 | 18 | 26 | 30 | 20 | 32 |
| 28 | 32 | 21 | 23 | 24 | 18 | 29 | 23 | 26 | 23 |
| 22 | 26 | 18 | 31 | 19 | 22 | 22 | 19 | 18 | 32 |
| 22 | 29 | 30 | 32 | 23 | 25 | 24 | 24 | 29 | 20 |
| 22 | 30 | 27 | 22 | 22 | 30 | 32 | 19 | 18 | 32 |
| 31 | 27 | 24 | 29 | 30 | 25 | 21 | 19 | 26 | 28 |
| 22 | 32 | 31 | 27 | 32 | 27 | 30 | 20 | 26 | 20 |
| 27 | 25 | 19 | 25 | 18 | 25 | 29 | 20 | 26 | 31 |
| 32 | 29 | 31 | 28 | 29 | 24 | 18 | 19 | 21 | 19 |

**Вариант 7.**

В 100 фирмах проведён социологический опрос. Процент курящих в каждой из фирм составил:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 35 | 40 | 48 | 46 | 47 | 36 | 44 | 43 | 34 | 42 |
| 49 | 42 | 44 | 41 | 45 | 34 | 48 | 50 | 45 | 49 |
| 43 | 52 | 48 | 44 | 45 | 45 | 34 | 49 | 51 | 48 |
| 39 | 34 | 47 | 47 | 45 | 52 | 40 | 52 | 52 | 38 |
| 48 | 46 | 51 | 48 | 42 | 40 | 46 | 36 | 35 | 49 |
| 35 | 44 | 34 | 40 | 46 | 43 | 36 | 46 | 36 | 44 |
| 35 | 39 | 46 | 41 | 50 | 41 | 49 | 44 | 42 | 43 |
| 38 | 36 | 46 | 42 | 43 | 50 | 44 | 42 | 34 | 38 |
| 44 | 44 | 41 | 48 | 46 | 39 | 48 | 50 | 40 | 35 |
| 45 | 36 | 47 | 37 | 38 | 34 | 38 | 41 | 35 | 34 |

**Вариант 8.**

Проведен медицинский осмотр. Ниже приведены года рождения обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1987 | 1982 | 1985 | 1985 | 1994 | 1990 | 1990 | 1982 | 1992 | 1996 |
| 1985 | 1985 | 1986 | 1997 | 1993 | 1999 | 1989 | 1993 | 1991 | 1997 |
| 1993 | 1997 | 1982 | 1998 | 1982 | 1988 | 1982 | 1997 | 1981 | 1995 |
| 1988 | 1995 | 1991 | 1982 | 1986 | 1989 | 1986 | 1985 | 1987 | 1997 |
| 1996 | 1996 | 1994 | 1995 | 1991 | 1985 | 1999 | 1994 | 1984 | 1989 |
| 1996 | 1981 | 1995 | 1993 | 1987 | 1997 | 1986 | 1994 | 1999 | 1988 |
| 1992 | 1993 | 1997 | 1988 | 1986 | 1989 | 1987 | 1985 | 1999 | 1987 |
| 1981 | 1986 | 1981 | 1992 | 1984 | 1989 | 1991 | 1990 | 1988 | 1986 |
| 1981 | 1987 | 1992 | 1992 | 1985 | 1986 | 1999 | 1998 | 1981 | 1990 |
| 1981 | 1987 | 1991 | 1992 | 1987 | 1998 | 1987 | 1987 | 1996 | 1984 |

**Вариант 9.**

Проведен медицинский осмотр учащихся. Приведены результаты измерения роста обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 168 | 172 | 169 | 174 | 170 | 177 | 185 | 179 | 168 | 171 |
| 173 | 182 | 171 | 182 | 179 | 170 | 184 | 173 | 173 | 181 |
| 175 | 172 | 183 | 174 | 182 | 168 | 183 | 181 | 178 | 183 |
| 180 | 183 | 183 | 183 | 182 | 181 | 176 | 170 | 175 | 174 |
| 172 | 182 | 182 | 168 | 185 | 171 | 176 | 174 | 179 | 170 |
| 178 | 174 | 168 | 172 | 174 | 173 | 182 | 169 | 170 | 171 |
| 184 | 177 | 183 | 184 | 175 | 173 | 182 | 176 | 171 | 173 |
| 181 | 177 | 185 | 176 | 185 | 185 | 176 | 176 | 174 | 171 |
| 181 | 179 | 176 | 182 | 169 | 171 | 170 | 179 | 174 | 173 |
| 169 | 170 | 182 | 172 | 174 | 172 | 183 | 183 | 173 | 177 |

**Вариант 10.**

Выпускники школы сдавали ЕГЭ. Результаты ЕГЭ приведены ниже:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 58 | 66 | 69 | 61 | 69 | 71 | 71 | 67 | 67 | 71 |
| 60 | 67 | 66 | 60 | 65 | 70 | 64 | 71 | 62 | 62 |
| 69 | 59 | 67 | 63 | 67 | 71 | 61 | 61 | 67 | 68 |
| 69 | 64 | 68 | 61 | 71 | 64 | 67 | 71 | 59 | 67 |
| 60 | 58 | 67 | 70 | 59 | 70 | 65 | 64 | 71 | 60 |
| 64 | 61 | 66 | 68 | 71 | 70 | 60 | 58 | 59 | 59 |
| 61 | 65 | 58 | 61 | 62 | 68 | 71 | 60 | 59 | 70 |
| 60 | 67 | 69 | 68 | 63 | 71 | 67 | 60 | 59 | 62 |
| 58 | 67 | 69 | 70 | 68 | 65 | 70 | 64 | 69 | 71 |
| 59 | 68 | 70 | 59 | 67 | 65 | 69 | 58 | 71 | 63 |

**Вариант 11.**

Было задано 100 задач по математике. Ниже приведены результаты проверки (количество правильно решенных задач):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 59 | 63 | 65 | 63 | 59 | 67 | 65 | 70 | 60 | 64 |
| 69 | 68 | 68 | 68 | 68 | 60 | 65 | 58 | 70 | 68 |
| 61 | 60 | 67 | 70 | 64 | 69 | 62 | 58 | 59 | 64 |
| 59 | 69 | 69 | 71 | 60 | 71 | 62 | 64 | 59 | 65 |
| 61 | 63 | 62 | 63 | 71 | 70 | 63 | 65 | 58 | 60 |
| 60 | 66 | 59 | 61 | 62 | 65 | 70 | 60 | 66 | 62 |
| 58 | 67 | 66 | 67 | 64 | 58 | 65 | 61 | 61 | 69 |
| 71 | 68 | 63 | 71 | 67 | 60 | 59 | 71 | 70 | 69 |
| 66 | 60 | 62 | 58 | 69 | 68 | 61 | 70 | 59 | 66 |
| 71 | 65 | 71 | 58 | 63 | 70 | 61 | 67 | 62 | 66 |

**Вариант 12.**

Процент прибывших по сигналу «ТРЕВОГА» в подразделениях составил:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 95 | 98 | 92 | 85 | 85 | 81 | 100 | 89 | 88 | 89 |
| 100 | 98 | 98 | 94 | 92 | 82 | 91 | 98 | 85 | 98 |
| 84 | 93 | 98 | 88 | 94 | 84 | 98 | 96 | 98 | 90 |
| 94 | 84 | 84 | 82 | 99 | 97 | 86 | 92 | 86 | 86 |
| 95 | 91 | 98 | 97 | 82 | 84 | 90 | 84 | 91 | 96 |
| 99 | 82 | 98 | 94 | 85 | 85 | 90 | 96 | 84 | 89 |
| 83 | 98 | 85 | 97 | 86 | 92 | 86 | 84 | 92 | 94 |
| 94 | 81 | 81 | 87 | 100 | 95 | 82 | 85 | 92 | 92 |
| 97 | 96 | 100 | 98 | 83 | 90 | 91 | 81 | 94 | 91 |
| 96 | 82 | 98 | 87 | 91 | 90 | 90 | 100 | 91 | 88 |

**Вариант 13.**

Проведен медицинский осмотр. Ниже приведен возраст обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 30 | 21 | 28 | 30 | 30 | 20 | 26 | 25 | 22 | 20 |
| 19 | 19 | 31 | 27 | 21 | 21 | 24 | 22 | 20 | 20 |
| 27 | 30 | 32 | 24 | 24 | 27 | 20 | 29 | 32 | 23 |
| 19 | 25 | 28 | 23 | 32 | 22 | 21 | 29 | 32 | 21 |
| 30 | 32 | 20 | 30 | 24 | 23 | 29 | 24 | 18 | 29 |
| 24 | 26 | 25 | 30 | 30 | 28 | 29 | 20 | 28 | 20 |
| 21 | 27 | 30 | 28 | 19 | 28 | 30 | 27 | 23 | 22 |
| 29 | 23 | 27 | 30 | 26 | 23 | 25 | 24 | 31 | 24 |
| 24 | 19 | 26 | 27 | 21 | 26 | 20 | 25 | 24 | 25 |
| 31 | 19 | 28 | 24 | 29 | 28 | 25 | 27 | 29 | 29 |

**Вариант 14.**

В 100 фирмах проведён социологический опрос. Процент курящих в каждой из фирм составил:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 35 | 52 | 42 | 36 | 48 | 34 | 47 | 45 | 35 | 50 |
| 48 | 46 | 52 | 44 | 50 | 42 | 44 | 47 | 45 | 45 |
| 37 | 47 | 43 | 49 | 41 | 42 | 44 | 49 | 38 | 41 |
| 51 | 46 | 46 | 37 | 40 | 50 | 51 | 48 | 37 | 41 |
| 43 | 51 | 49 | 49 | 38 | 47 | 42 | 43 | 49 | 38 |
| 46 | 45 | 50 | 41 | 48 | 38 | 43 | 46 | 47 | 43 |
| 43 | 51 | 52 | 47 | 35 | 37 | 42 | 52 | 35 | 41 |
| 38 | 39 | 43 | 34 | 44 | 40 | 45 | 46 | 50 | 37 |
| 39 | 41 | 37 | 36 | 44 | 47 | 49 | 37 | 39 | 37 |
| 42 | 39 | 47 | 42 | 45 | 52 | 42 | 35 | 42 | 37 |

**Вариант 15.**

Проведен медицинский осмотр. Ниже приведены года рождения обследуемых:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1995 | 1990 | 1996 | 1995 | 1995 | 1990 | 1998 | 1992 | 1992 | 1996 |
| 1992 | 1999 | 1996 | 1994 | 1989 | 1998 | 1991 | 1995 | 1996 | 1992 |
| 1997 | 1993 | 1994 | 1995 | 1993 | 1991 | 1993 | 1992 | 1993 | 1990 |
| 1995 | 1998 | 1991 | 1997 | 1996 | 1996 | 1997 | 1991 | 1992 | 1997 |
| 1996 | 1993 | 1995 | 1999 | 1998 | 1989 | 1997 | 1998 | 1990 | 1998 |
| 1990 | 1996 | 1998 | 1995 | 1995 | 1998 | 1997 | 1993 | 1993 | 1997 |
| 1996 | 1997 | 1998 | 1993 | 1994 | 1998 | 1998 | 1994 | 1994 | 1991 |
| 1997 | 1990 | 1995 | 1993 | 1994 | 1990 | 1989 | 1994 | 1989 | 1992 |
| 1998 | 1998 | 1999 | 1992 | 1998 | 1989 | 1999 | 1993 | 1996 | 1999 |
| 1997 | 1997 | 1995 | 1990 | 1993 | 1989 | 1992 | 1998 | 1997 | 1999 |

## Контрольные вопросы

1. Что такое генеральная совокупность?
2. Что такое выборка (выборочная совокупность)?
3. Что такое объем совокупности?
4. Что такое статистический ряд?
5. Что такое частота?
6. Что такое относительная частота?
7. Что такое полигон частот (относительных частот)?
8. Что такое вариационный ряд?
9. Что такое выборочная (генеральная) средняя?
10. Что такое выборочная (генеральная) дисперсия?
11. Что такое среднее квадратическое отклонение?
12. Что такое исправленная дисперсия?
13. Что такое мода?
14. Что такое медиана?
15. Что такое размах ряда?

## Лабораторная работа № 2

### Законы распределения дискретных случайных величин

**Цель лабораторной работы:** изучить основные понятия, связанные с законами распределения дискретных случайных величин, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

*Определение.* **Случайная величина** это величина, которая принимает определенное значение в результате опыта.

*Определение.* **Дискретной случайной величиной** называется такая величина, которая в результате опыта может принимать определенные значения с определенной вероятностью, образующие счетное множество (множество, элементы которого могут быть занумерованы).

*Определение.* **Непрерывной случайной величиной** называется такая величина, которая может принимать любые значения из некоторого конечного или бесконечного промежутка.

*Определение.* **Законом распределения дискретной случайной величины** называют соответствие между возможными значениями случайной величины и их вероятностями.

### Способы задания случайных величин

*Определение.* Таблица соответствия значений случайной величины и их вероятностей называется **рядом распределения**.

Таблица имеет следующий вид:

| **X** | $x_1$ | $x_2$ | …. | $x_n$ |
|---|---|---|---|---|
| **P** | $p_1$ | $p_2$ | ….. | $p_n$ |

где $\sum_{i=1}^{n} p_i = 1$.

Графическое представление этой таблицы называется **многоугольником распределения**. При этом сумма всех ординат многоугольника распределения представляет собой вероятность всех возможных значений случайной величины, а следовательно, равна единице.

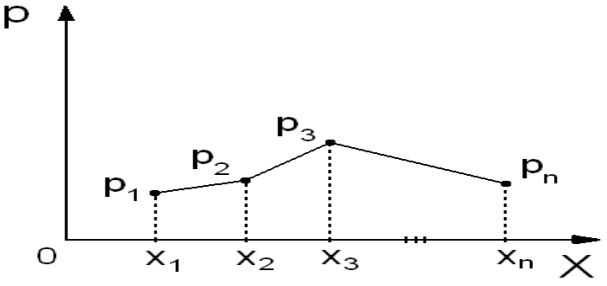


Рис. 2.1

*Определение.* **Функцией распределения** называют функцию F(x), определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее *x*.

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

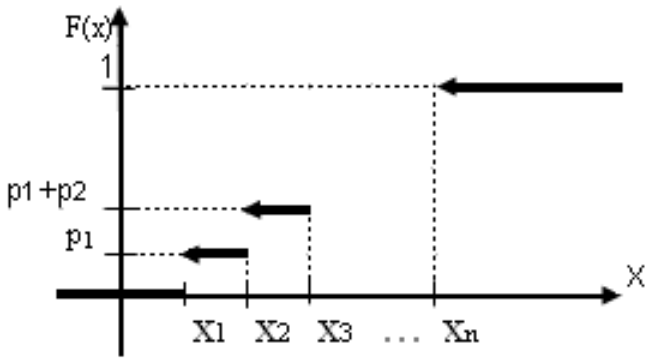



Рис. 2.2

## Свойства функции распределения

1) Значения функции распределения принадлежат отрезку [0, 1].

$$0 \le F(x) \le 1$$

2) F(x) – неубывающая функция.

$$F(x_2) \ge F(x_1) \text{ при } x_2 \ge x_1$$

3) Вероятность того, что случайная величина примет значение, заключенное в интервале (*a, b*), равна приращению функции распределения на этом интервале.

$$P(a \le X < b) = F(b) - F(a)$$

4) На минус бесконечности функция распределения равна нулю, на плюс бесконечности функция распределения равна единице.

*F(-∞)=P (X<-∞)=0* как вероятность невозможного события *X*<-∞,

*F(+∞)=P(X<+∞)=1* как вероятность достоверного события X<-∞.

5) Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

## Числовые характеристики случайных величин

### I. Математическое ожидание

*Определение.* **Математическим ожиданием M(X)** дискретной случайной величины X называется сумма произведений всех ее значений на соответствующие им вероятности:

$$M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

Свойства математического ожидания

1) $M(C)=C$, где C – постоянная величина;
2) $M(C \bullet X)=C \bullet M(X)$,
3) $M(X \pm Y)=M(X) \pm M(Y)$;
4) $M(X \bullet Y)=M(X) \bullet M(Y)$, где $X,Y$ – независимые случайные величины;
5) $M(X \pm C)=M(X) \pm C$, где $C$ – постоянная величина.

## II. Дисперсия

*Определение.* **Дисперсией D(X)** случайной величины X называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания:

$$D(X) = M\left[X - M(X)\right]^2$$

Свойства дисперсии

*1)* $D(C)=0$, где $C$ – постоянная величина;

*2)* $D(X)>0$, где $X$ – случайная величина;

*3)* $D(C \bullet X)=C^2 \bullet D(X)$, где $C$– постоянная величина;

*4)* $D(X+Y)=D(X)+D(Y)$, где $X,Y$ – независимые случайные величины;

*5)* Дисперсия разности двух независимых случайных величин равна сумме дисперсий этих величин. $D(X-Y) = D(X) + D(Y)$

Для вычисления дисперсии часто бывает удобно пользоваться формулой:

$$D(X) = M(X^2) - M^2(X)$$

## III. Среднеквадратическое отклонение

*Определение.* **Средним квадратическим отклонением σ(X)** случайной величины X называется квадратный корень из дисперсии:

$$\sigma(X) = \sqrt{D(X)}$$

## Основные распределения дискретных случайных величин

### I. Равномерное распределение на конечном множестве

| ***X*** | $x_1$ | $x_2$ | … | $x_n$ |
|---|---|---|---|---|
| ***P*** | $\frac{1}{n}$ | $\frac{1}{n}$ | … | $\frac{1}{n}$ |

$$M(x)=\frac{x_1}{n}+\frac{x_2}{n}+\ldots+\frac{x_n}{n}=\frac{1}{n}(x_1+x_2+\ldots+x_n)$$

*Пример 2.1.* Найти математическое ожидание числа очков, выпадающих при бросании игральной кости.

*Решение.* Случайная величина X числа очков принимает значения 1, 2, 3, 4, 5, 6. Вероятность того, что выпадет одно из данных значений, равна 1/6. Закон распределения представим в виде таблицы:

| **X** | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **P** | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

Найдем математическое ожидание величины X:

$$M(X)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{1+2+3+4+5+6}{6}=\frac{21}{6}=3{,}5$$

**Решение с помощью табличного процессора.**

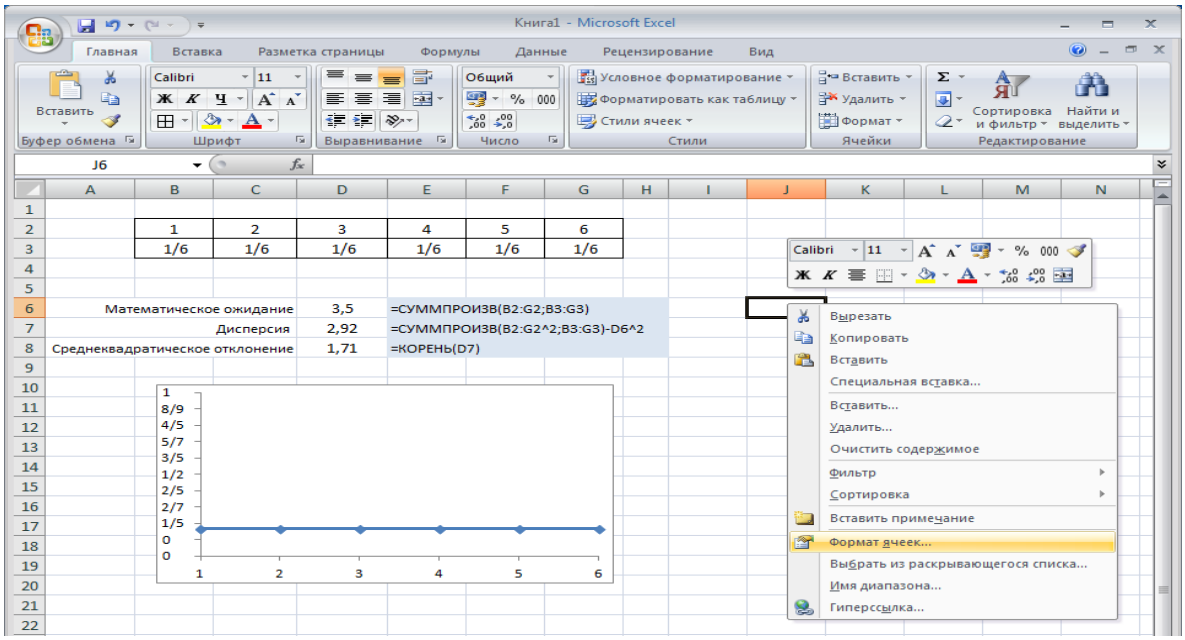



Рис. 2.3.

➢Указание: Необходимо поменять формат ячейки: «*Формат ячейки*» → «*Число*» → «*Дробный*».

## II. Биномиальное распределение *Bi(n,p)*

Биномиальным называют закон распределения дискретной случайной величины *X* – числа появлений события в ***n*** независимых испытаниях, в каждом из которых вероятность наступления события постоянна.

Вероятности $p_i$ вычисляют по формуле Бернулли $P(X=k)=P_n(k)=C_n^k p^k (1-p)^{n-k}$.

| **X** | 0 | 1 | … | k | … | n |
|---|---|---|---|---|---|---|
| **P** | $C_n^0 p^0 (1-p)^n$ | $C_n^1 p^1 (1-p)^{n-1}$ | … | $C_n^k p^k (1-p)^{n-k}$ | … | $C_n^n p^n (1-p)^0$ |

Для биномиального распределения:

- математическое ожидание *M(X) = np*,
- дисперсия *D(X) = npq*.

В пределе при n→∞ биномиальное распределение по своим значениям приближается к **нормальному с параметрами** $a=np$ и $\sigma=\sqrt{npq}$.

В пределе при n→∞ и при p→0 биномиальное распределение превращается в **распределение Пуассона** с параметром $\lambda=np$.

*Пример 2.2.* Построить ряд распределения числа попаданий мячом в корзину при трех бросках, если вероятность попадания при одном броске равна 0,6. Найти среднее число попаданий и дисперсию.

*Решение.* Случайная величина X – число попаданий в корзину при трёх бросках. Соответствующие вероятности найдём по формуле Бернулли.

$$P_3(0)=C_3^0 p^0 (1-p)^{3-0}=C_3^0 p^0 q^3=\frac{3!}{0!3!}\cdot 0{,}6^{01}\cdot 0{,}4^3=1\cdot 1\cdot 0{,}064=0{,}064$$

$$P_3(1)=C_3^1 p^1 (1-p)^{3-1}=C_3^1 p^1 q^2=\frac{3!}{1!2!}\cdot 0{,}6^1\cdot 0{,}4^2=3\cdot 0{,}6\cdot 0{,}16=0{,}288$$

$$P_3(2)=C_3^2 p^2 (1-p)^{3-2}=C_3^2 p^2 q^1=\frac{3!}{2!1!}\cdot 0{,}6^2\cdot 0{,}4^1=3\cdot 0{,}36\cdot 0{,}4=0{,}432$$

$$P_3(3)=C_3^3 p^3 (1-p)^{3-3}=C_3^3 p^3 q^0=\frac{3!}{3!0!}\cdot 0{,}6^3\cdot 0{,}4^0=1\cdot 0{,}216\cdot 1=0{,}216$$

Искомый закон распределения:

| **X** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| **P** | 0,064 | 0,288 | 0,432 | 0,216 |

Контроль: 0,064 + 0,288 + 0,432 + 0,216 = 1

Математическое ожидание:

$M(X)=\Sigma x_i p_i = 0\cdot 0{,}064 + 1\cdot 0{,}288 + 2\cdot 0{,}432 + 3\cdot 0{,}216 = 1{,}8$

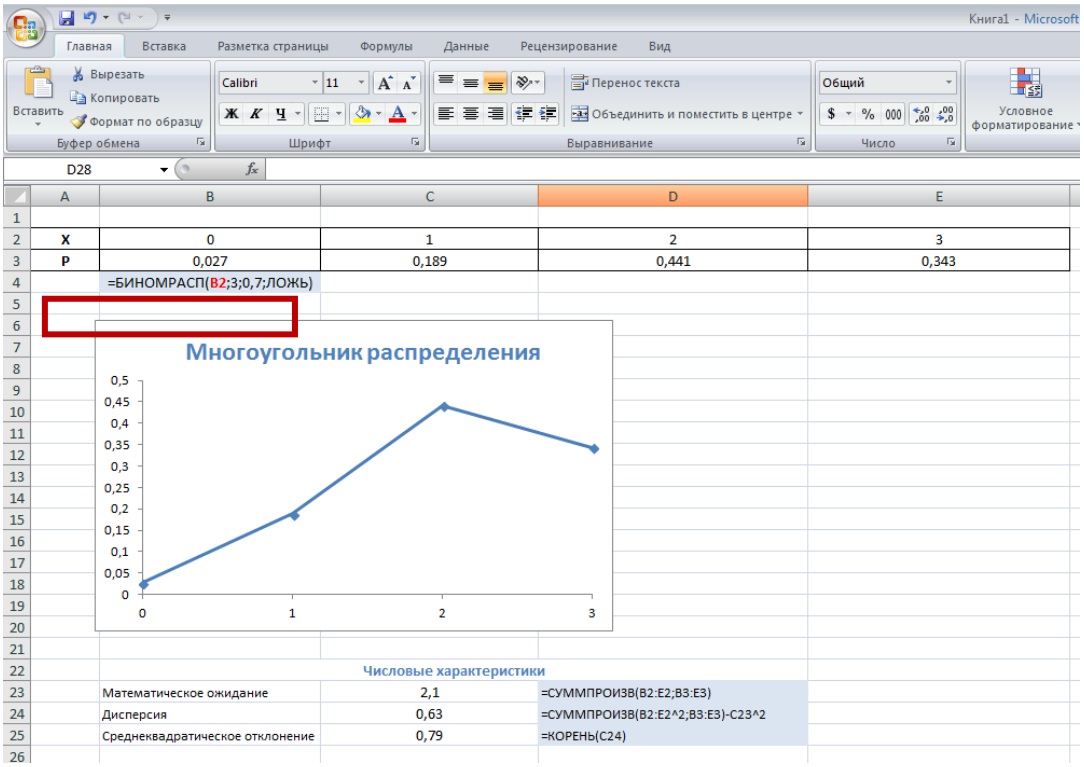
или: $M(X)=np=3\cdot 0{,}6=1{,}8$

Дисперсия:

$D(X)=\Sigma x_i^2 p_i - (M(X))^2 = 0^2\cdot 0{,}064 + 1^2\cdot 0{,}288 + 2^2\cdot 0{,}432 + 3^2\cdot 0{,}216 - 1{,}8^2 = 0{,}72$ или: $D(X)=npq=3\cdot 0{,}6\cdot 0{,}4=0{,}72$

Среднее квадратическое отклонение: $\sigma(X) = \sqrt{D(X)} \approx 0{,}85$

**Решение с помощью табличного процессора.**




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | X | 0 | 1 | 2 | 3 |
| 3 | P | 0,027 | 0,189 | 0,441 | 0,343 |
| 4 | | =БИНОМРАСП(B2;3;0,7;ЛОЖЬ) | | | |
| 22 | | Числовые характеристики | | | |
| 23 | | Математическое ожидание | 2,1 | =СУММПРОИЗВ(B2:E2;B3:E3) | |
| 24 | | Дисперсия | 0,63 | =СУММПРОИЗВ(B2:E2^2;B3:E3)-C23^2 | |
| 25 | | Среднеквадратическое отклонение | 0,79 | =КОРЕНЬ(C24) | |

Рис. 2.4

➢Указание: Вероятность того, что в ***n*** независимых испытаниях, в каждом из которых вероятность наступления события равна ***p***, событие наступит ровно ***k*** раз. Вероятности $P_i$ вычисляют по формуле Бернулли

$$БИНОМРАСП(k,n,p,ЛОЖЬ) = P_n(k) = C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

## III. Распределение пуассона

Если число испытаний ***n*** очень велико, а вероятность появления события ***A*** в каждом испытании очень мала ***(p≤0,1)***, то для вычисления ***P(X=k)*** используют формулу Пуассона:

$$P(X=k) = P_n(k) = (np)^k \frac{e^{-np}}{k!}$$

| **X** | 0 | 1 | 2 | … | k | … |
|---|---|---|---|---|---|---|
| **P** | $(np)^0 \frac{e^{-np}}{0!}$ | $(np)^1 \frac{e^{-np}}{1!}$ | $(np)^2 \frac{e^{-np}}{2!}$ | … | $(np)^k \frac{e^{-np}}{k!}$ | … |

При решении задач можно использовать таблицу значений вероятностей распределения Пуассона (см. Приложение 1)

*Пример 2.3.*_Станок-автомат штампует детали. Вероятность того, что изготовленная деталь окажется бракованной, равна 0,002. Найти вероятность того, что среди 1000 отобранных деталей окажется:

***а)*** 5 бракованных;

***б)*** хотя бы одна бракованная.

*Решение.* Число n=1000 велико, вероятность изготовления бракованной детали p=0,002 мала, и рассматриваемые события (деталь окажется бракованной) независимы, поэтому имеет место формула Пуассона:

$$P(X=k)=P_n(k)=(np)^k\frac{e^{-np}}{k!}$$

Найдем λ=np=1000•0,002=2.

***а)*** Найдем вероятность того, что будет 5 бракованных деталей (m=5):

$$P(X=5)=P_{1000}(5)=(1000\cdot 0{,}002)^5\frac{e^{-1000\cdot 0{,}002}}{5!}=2^5\frac{e^{-2}}{5!}=0{,}0361$$

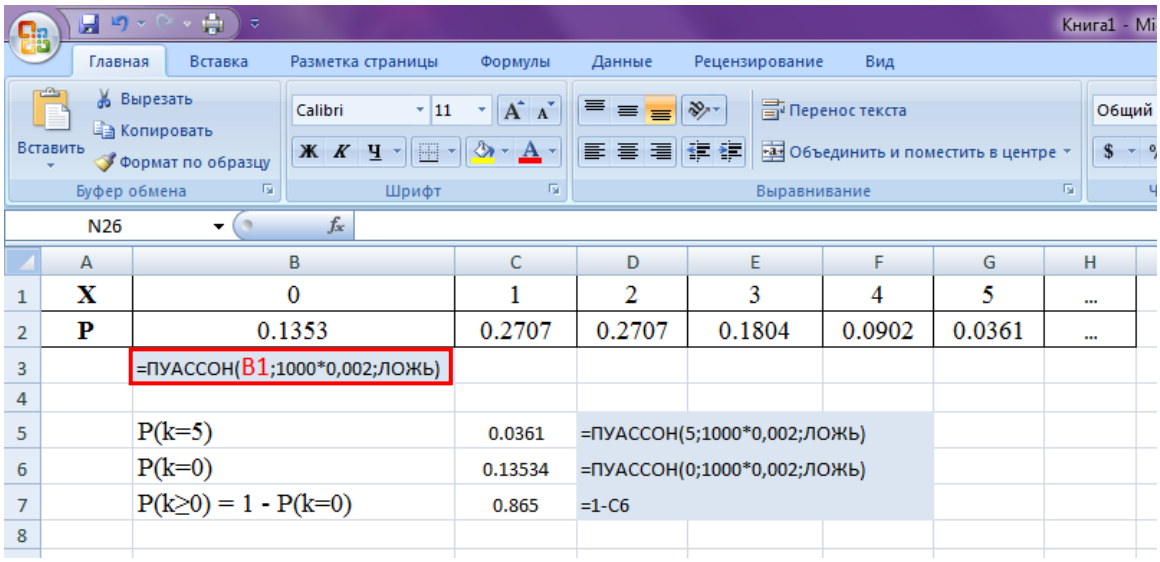
***б)*** Найдем вероятность того, что будет хотя бы одна бракованная деталь. Событие ***A*** – «хотя бы одна из отобранных деталей бракованная» является противоположным событию $\overline{A}$ — «все отобранные детали не бракованные». Следовательно, $P(A)=1-P(\overline{A})$. Отсюда искомая вероятность равна:

$$P(X=5)=1-P_{1000}(0)=1-(1000\cdot 0{,}002)^0\frac{e^{-1000\cdot 0{,}002}}{0!}=1-2^0\frac{e^{-2}}{0!}=1-0{,}13534=0{,}865$$

**Решение с помощью таблицы значений вероятностей распределения Пуассона** (см. Приложение 1).

| np / k | 1 | 2 | 3 | 4 | 5 | … |
|---|---|---|---|---|---|---|
| 0 | 0,3679 | 0,1353 | 0,0498 | 0,0183 | 0,0067 | … |
| 1 | 0,3679 | 0,2707 | 0,1494 | 0,0733 | 0,0337 | … |
| 2 | 0,1839 | 0,2707 | 0,2240 | 0,1465 | 0,0842 | … |
| 3 | 0,0613 | 0,1804 | 0,2240 | 0,1954 | 0,1404 | … |
| 4 | 0,0153 | 0,0902 | 0,1680 | 0,1954 | 0,1755 | … |
| 5 | 0,0031 | **0,0361** | 0,1008 | 0,1563 | 0,1755 | … |
| 6 | 0,0005 | 0,0120 | 0,0504 | 0,1042 | 0,1462 | … |
| 7 | 0,0001 | 0,0037 | 0,0216 | 0,0595 | 0,1044 | … |
| 8 | 0 | 0,0009 | 0,0081 | 0,0298 | 0,0653 | … |
| 9 | 0 | 0,0002 | 0,0027 | 0,0132 | 0,0363 | … |
| 10 | 0 | 0 | 0,0008 | 0,0053 | 0,0181 | … |
| … | … | … | … | … | … | … |

**Решение с помощью табличного процессора.**




| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | **X** | 0 | 1 | 2 | 3 | 4 | 5 | … |
| 2 | **P** | 0.1353 | 0.2707 | 0.2707 | 0.1804 | 0.0902 | 0.0361 | … |
| 3 | | =ПУАССОН(B1;1000*0,002;ЛОЖЬ) | | | | | | |
| 4 | | | | | | | | |
| 5 | | P(k=5) | 0.0361 | =ПУАССОН(5;1000*0,002;ЛОЖЬ) | | | | |
| 6 | | P(k=0) | 0.13534 | =ПУАССОН(0;1000*0,002;ЛОЖЬ) | | | | |
| 7 | | P(k≥0) = 1 - P(k=0) | 0.865 | =1-C6 | | | | |
| 8 | | | | | | | | |

Рис. 2.5

Пусть имеется некоторая последовательность событий, наступающих в случайные моменты времени (будем называть это потоком событий). ***Интенсивность потока*** (среднее число событий, появляющихся в единицу времени) равна ***λ***. Пусть этот поток событий – ***простейший*** (пуассоновский), т.е. обладает тремя свойствами:

1) вероятность появления ***k*** событий за определённый промежуток времени зависит только от длины этого промежутка, но не от точки отсчёта, другими словами, интенсивность потока есть постоянная величина (***свойство стационарности***);

2) вероятность появления ***k*** событий в любом промежутке времени не зависит от того, появлялись события в прошлом или нет (***свойство «отсутствия последействия»***);

3) появление более одного события за малый промежуток времени практически невозможно (***свойство ординарности***).

Вероятность того, что за промежуток времени *t* событие произойдёт ***k*** раз, равна $P(X=k)=P_t(k)=(\lambda t)^k\frac{e^{-\lambda t}}{k!}$

| **X** | 0 | 1 | 2 | … | k | … |
|---|---|---|---|---|---|---|
| **P** | $(\lambda t)^0\frac{e^{-\lambda t}}{0!}$ | $(\lambda t)^1\frac{e^{-\lambda t}}{1!}$ | $(\lambda t)^2\frac{e^{-\lambda t}}{2!}$ | … | $(\lambda t)^k\frac{e^{-\lambda t}}{k!}$ | … |

Числовые характеристики: $M(X)=D(X)=\lambda t$.

*Пример 2.4.* Среднее число вызовов, поступающих на АТС за 1 мин, равно двум. Найти вероятность того, что за 4 мин. поступит:

***а)*** три вызова;

***б)*** менее трёх вызовов;

***в)*** не менее трёх вызовов.

Поток вызовов – простейший.

*Решение.* Используем формулу Пуассона. $\lambda = 2$, $t = 4$.

$$P(0) = 8^0/0!\cdot e^{-8} = e^{-8} \approx 0{,}000335$$
$$P(1) = 8^1/1!\cdot e^{-8} = 8e^{-8} \approx 0{,}002684$$
$$P(2) = 8^2/2!\cdot e^{-8} = 32e^{-8} \approx 0{,}010735$$
$$P(3) = 8^3/3!\cdot e^{-8} = 85{,}33e^{-8} \approx 0{,}028626$$

| **X** | 0 | 1 | 2 | 3 | … |
|---|---|---|---|---|---|
| **P** | 0,000335 | 0,002684 | 0,010735 | 0,028626 | … |

***а)*** $P(k=3) = 0{,}028626$

***б)*** $P(k<3)=P(k\leq 2) = P(0) + P(1) + P(2) = 0{,}013754$

***в)*** $P(k\geq 3) = 1 - P(k<3) = 1 - 0{,}013754 = 0{,}986246$

**Решение с помощью табличного процессора.**

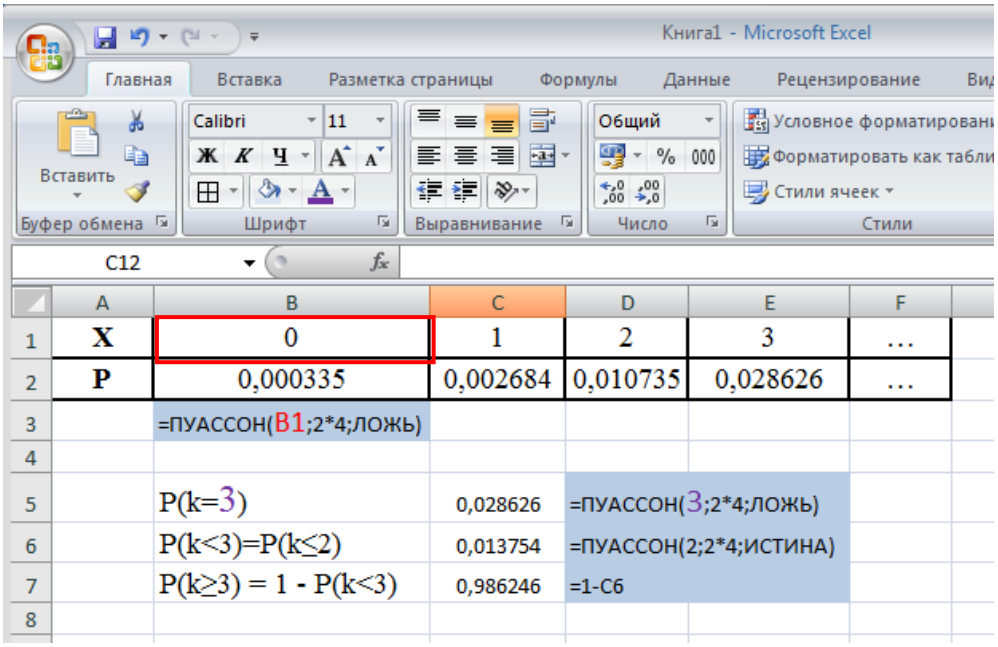



Рис. 2.6

➢Указание:

*1)* В случаях, когда находится P(k), нужно использовать функцию

$$ПУАССОН(k, np, ложь) = P_n(k) = (np)^k \frac{e^{-np}}{k!}$$

*2)* В случаях, когда находится P(≤k), нужно использовать функцию

$$ПУАССОН(k, np, истина) = P_n(\leq k) = P(0) + P(1) + \ldots + P(k)$$

## IV. Геометрическое распределение

Производится серия испытаний. Случайная величина X – количество испытаний до появления первого успеха (например, бросание мяча в корзину до первого попадания). Закон распределения имеет вид:

| **X** | 1 | 2 | 3 | … | k | … |
|---|---|---|---|---|---|---|
| **P** | $p$ | $qp$ | $q^2 p$ | … | $q^{k-1} p$ | … |

Если количество испытаний не ограничено, т.е. если случайная величина может принимать значения 1, 2, ..., ∞, то ***математическое ожидание*** и ***дисперсию*** геометрического распределения можно найти по формулам $M(X) = 1/p$, $D(X) = q/p^2$.

*Пример 2.5.* Из орудия производится стрельба по цели до первого попадания. Вероятность попадания в цель p = 0,6 при каждом выстреле. С.в. X - число возможных выстрелов до первого попадания.

***а)*** Составить ряд распределения и найти числовые характеристики.

***б)*** Найти математическое ожидание и дисперсию для случая, если стрелок намеревается произвести не более трёх выстрелов.

*Решение.*

а) Случайная величина может принимать значения 1, 2, 3, 4,..., ∞

$$P(1) = p = 0{,}6$$
$$P(2) = qp = 0{,}4 \cdot 0{,}6 = 0{,}24$$
$$P(3) = q^2p = 0{,}4^2 \cdot 0{,}6 = 0{,}096$$
$$\dots$$
$$P(k) = q^{k-1}p = 0{,}4^{k-1} \cdot 0{,}6$$
$$\dots$$

Ряд распределения:

| **X** | 1 | 2 | 3 | ... | k | ... |
|---|---|---|---|---|---|---|
| **P** | 0,6 | 0,24 | 0,096 | ... | $0{,}4^{k-1} \cdot 0{,}6$ | ... |

Контроль: $\Sigma p_i = 0{,}6/(1-0{,}4) = 1$ (сумма геометрической прогрессии)

Числовые характеристики.

$M(X) = 1/p = 1/0{,}6 \approx 1{,}667$

$D(x) = q/p^2 = 0{,}4/0{,}36 \approx 1{,}111$

$\sigma(X) = \sqrt{D(X)} \approx 1{,}054$

б) Случайная величина может принимать значения 1, 2, 3.

$P(1) = p = 0{,}6$(попал при первом выстреле);

$P(2) = qp = 0{,}4 \cdot 0{,}6 = 0{,}24$(не попал при первом выстреле **и** попал при втором выстреле);

$P(3) = q^2p + q^3 = 0{,}4^2 \cdot 0{,}6 + 0{,}4^3 = 0{,}16$ (попал при 3 выстреле **или** не попал все 3 раза).

Ряд распределения:

| **X** | 1 | 2 | 3 |
|---|---|---|---|
| **P** | 0,6 | 0,24 | 0,16 |

Контроль: $\Sigma p_i = 0,6 + 0,24 + 0,16 = 1$

Числовые характеристики.

$M(X) = 1 \cdot 0,6 + 2 \cdot 0,24 + 3 \cdot 0,16 = 1,56$

$D(X) = 1^2 \cdot 0,6 + 2^2 \cdot 0,24 + 3^2 \cdot 0,16 - 1,56^2 = 0,5664$

$\sigma(X) \approx 0,752$

## V. Гипергеометрическое распределение

Имеется ***N*** объектов. Из них ***n*** объектов обладают требуемым свойством. Из общего количества отбирается ***m*** объектов. Случайная величина ***X*** – число объектов из ***m*** отобранных, обладающих требуемым свойством. Для вычисления вероятностей используются биномиальные коэффициенты (число сочетаний $C_n^m = \frac{n!}{m!(n-m)!}$).

Закон распределения имеет вид:

| $\mathbf{X_i}$ | 0 | 1 | 2 | … | k | … | m |
|---|---|---|---|---|---|---|---|
| $\mathbf{P_i}$ | $\frac{C_n^0 C_{N-n}^m}{C_N^m}$ | $\frac{C_n^1 C_{N-n}^{m-1}}{C_N^m}$ | $\frac{C_n^2 C_{N-n}^{m-2}}{C_N^m}$ | … | $\frac{C_n^k C_{N-n}^{m-k}}{C_N^m}$ | … | $\frac{C_n^m C_{N-n}^0}{C_N^m}$ |

*Пример 2.6.* Среди 20 книг, стоящих на полке, 8 книг по математической статистике. Случайная величина X - число книг по математике из четырёх случайно взятых с этой полки книг. Составить ряд распределения, найти функцию распределения, построить её график и найти все числовые характеристики.

*Решение.*

Случайная величина X может принимать значения *0, 1, 2, 3, 4*.

$$P(0) = \frac{C_{12}^4 \cdot C_8^0}{C_{20C}^4} = \frac{\frac{12!}{4!\cdot(12-4)!}\cdot 1}{\frac{20!}{4!\cdot(20-4)!}} = \frac{\frac{12!}{4!\cdot 8!}}{\frac{20!}{4!\cdot 16!}} = \frac{12!\cdot 4!\cdot 16!}{4!\cdot 8!\cdot 20!} = \frac{8!\cdot 9\cdot 10\cdot 11\cdot 12\cdot 4!\cdot 16!}{4!\cdot 8!\cdot 16!\cdot 17\cdot 18\cdot 19\cdot 20} = \frac{9\cdot 10\cdot 11\cdot 12}{17\cdot 18\cdot 19\cdot 20} \approx 0,102167$$

$$P(1) = \frac{C_{12}^3 \cdot C_8^1}{C_{20}^4} = \frac{\frac{12!}{3!\cdot(12-3)!}\cdot\frac{8!}{1!\cdot(8-1)!}}{\frac{20!}{4!\cdot(20-4)!}} = \frac{\frac{12!}{3!\cdot 9!}\cdot\frac{8!}{1!\cdot 7!}}{\frac{20!}{4!\cdot 16!}} \approx 0,363261$$

$$P(2) = \frac{C_{12}^2 \cdot C_8^2}{C_{20}^4} = \frac{\frac{12!}{2!\cdot(12-2)!}\cdot\frac{8!}{2!\cdot(8-2)!}}{\frac{20!}{4!\cdot(20-4)!}} = \frac{\frac{12!}{4!\cdot 10!}\cdot\frac{8!}{1!\cdot 6!}}{\frac{20!}{4!\cdot 16!}} \approx 0,381424$$

$$P(3)=\frac{C_{12}^1\cdot C_8^3}{C_{20}^4}=\frac{\frac{12!}{1!\cdot(12-1)!}\cdot\frac{8!}{3!\cdot(8-3)!}}{\frac{20!}{4!\cdot(20-4)!}}=\frac{\frac{12!}{1!\cdot 11!}\cdot\frac{8!}{3!\cdot 5!}}{\frac{20!}{4!\cdot 16!}}\approx 0{,}138700$$

$$P(3)=\frac{C_{12}^0\cdot C_8^4}{C_{20}^4}=\frac{\frac{12!}{0!\cdot(12-0)!}\cdot\frac{8!}{4!\cdot(8-4)!}}{\frac{20!}{4!\cdot(20-4)!}}=\frac{\frac{12!}{0!\cdot 12!}\cdot\frac{8!}{4!\cdot 4!}}{\frac{20!}{4!\cdot 16!}}\approx 0{,}014448$$

Ряд распределения:

| $x_i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $p_i$ | 0,102167 | 0,363261 | 0,381424 | 0,138700 | 0,014448 |

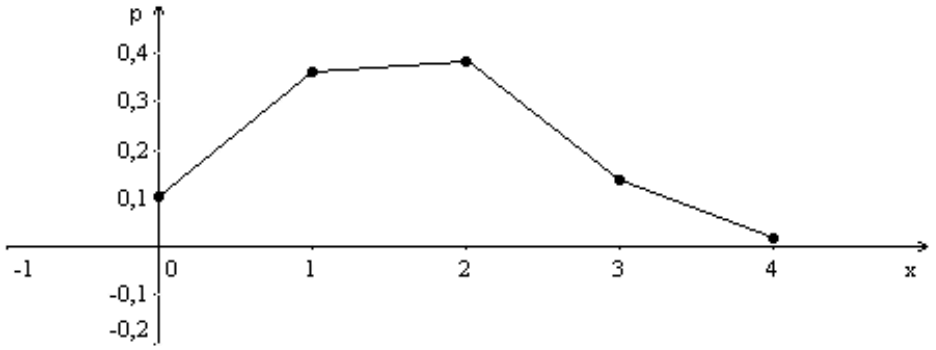



Рис. 2.7

Посчитаем числовые характеристики.

*M(X) = 0·0,1022 + 1·0,3633 + 2·0,38143 + 3·0,13873 + 4·0,0145 = 1,6*

*D(X) = $0^2$·0,1022 + $1^2$·0,3633 + $2^2$·0,38143 + $3^2$·0,13873 + $4^2$·0,0145 – $1{,}6^2 \approx 0{,}81$*

*σ(X) ≈ 0,90*

**Решение с помощью табличного процессора (рис. 2.8).**

➢Указание. Имеется ***N*** объектов. Из них ***n*** объектов обладают требуемым свойством. Из общего количества отбирается ***m*** объектов. Для вычисления вероятности того, что из ***m*** отобранных объектов ***k*** окажутся обладающими требуемыми свойствами, нужно использовать следующую формулу:

$$ГИПЕРГЕОМЕТ(k,m,N,n)=\frac{C_n^k C_{N-n}^{m-k}}{C_N^m}$$

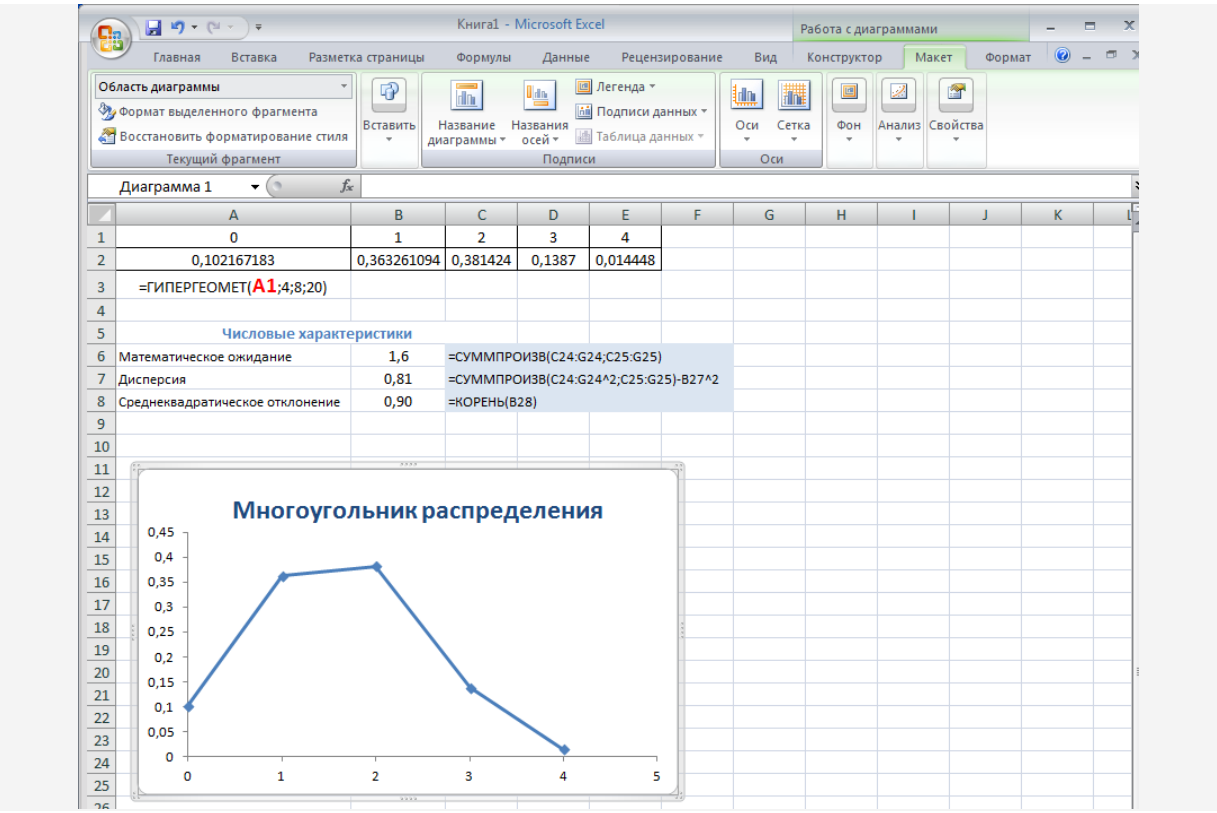



Рис. 2.8

## Задание для самостоятельного выполнения

**Задание 1.** Дискретная случайная величина X **распределена по равномерному закону**. Составить ряд распределения, построить многоугольник распределения, посчитать числовые характеристики.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ***1)*** | *X={-1,0,1}* | ***2)*** | *X={-1,0,1,2,3}* | ***3)*** | *X={-2,-1,1,2}* | ***4)*** | *X={-4,0,1,3,7}* | ***5)*** | *X={-6,-5,-4,-2}* |
| ***6)*** | *X={-2,0,7}* | ***7)*** | *X={-4,0,4,8}* | ***8)*** | *X={2,4,7}* | ***9)*** | *X={-4,-3,-2,-1}* | ***10)*** | *X={-3,0,2,3,5}* |
| ***11)*** | *X={-10,3,4}* | ***12)*** | *X={1,3,4,7}* | ***13)*** | *X={5,6,7,8}* | ***14)*** | *X={1,2,8,9}* | ***15)*** | *X={-1,3,5,6}* |

**Задание 2.** Случайная величина **распределена по биномиальному закону *Bi(n,p)***. Составить ряд распределения, построить многоугольник распределения, посчитать числовые характеристики.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ***1)*** | *Bi(3; 0,1)* | ***2)*** | *Bi(6; 0,5)* | ***3)*** | *Bi(5; 0,8)* | ***4)*** | *Bi(4; 0,75)* | ***5)*** | *Bi(7; 0,45)* |
| ***6)*** | *Bi(4; 0,2)* | ***7)*** | *Bi(7; 0,6)* | ***8)*** | *Bi(4; 0,9)* | ***9)*** | *Bi(5; 0,65)* | ***10)*** | *Bi(4; 0,35)* |
| ***11)*** | *Bi(5; 0,3)* | ***12)*** | *Bi(6; 0,7)* | ***13)*** | *Bi(3; 0,85)* | ***14)*** | *Bi(6; 0,55)* | ***15)*** | *Bi(5; 0,25)* |

**Задание 3.** В системе, состоящей из 9 равнонадежных занумерованными числами *(1,2,…,9)* приборов, отказал один прибор. Для его обнаружения приборы проверяют в порядке нумерации. Чему равно среднее число приборов, которое будет проверено?

**Задание 4.** Из десяти ключей в связке только один подходит к данному замку. Сколько в среднем придется перебрать ключей прежде, чем замок будет открыт?

**Задание 5.** Блок электронного устройства содержит 100 одинаковых элементов. Вероятность отказа каждого элемента в течение времени T равна 0,002. Элементы работают независимо. Найти вероятность того, что за время T откажет не более двух элементов.

**Задание 6.** Производятся последовательные независимые испытания трех приборов на надежность. Каждый следующий прибор испытывается только в том случае, если предыдущий оказался надежным. Вероятность выдержать испытание для каждого прибора равна 0,9. Составить закон распределения случайной величины X – числа испытанных приборов.

**Задание 7.** Баскетболист бросает мяч в корзину с вероятностью попадания при каждом броске 0,8. За каждое попадание он получает 10 очков, а в случае промаха очки ему не начисляют. Составить закон распределения случайной величины X – числа очков, полученных баскетболистом за 3 броска. Найти M(X), D(X), а также вероятность того, что он получит более 10 очков.

**Задание 8.** В коробке 9 фломастеров, из которых 2 фломастера уже не пишут. Наудачу берут 3 фломастера. Случайная величина X – число пишущих фломастеров среди взятых. Составить закон распределения случайной величины.

**Задание 9.** На стеллаже библиотеки в случайном порядке расставлено 6 учебников, причем 4 из них в переплете. Библиотекарь берет наудачу 4 учебника. Случайная величина X – число учебников в переплете среди взятых. Составить закон распределения случайной величины.

**Задание 10.** Учебник издан тиражом 50000 экземпляров. Вероятность того, что учебник сброшюрован неправильно, равна 0,0002. Найти вероятность того, что тираж содержит:

а) четыре бракованные книги,

б) менее двух бракованных книг.

## Контрольные вопросы

1. Что называется случайной величиной?
2. Какие случайные величины называются дискретными?
3. Что такое закон распределения дискретных случайных величин?
4. Что такое ряд распределения?
5. Что такое многоугольник распределения?
6. Что такое функция распределения?
7. Что такое математическое ожидание?
8. Что такое дисперсия?
9. Что такое среднеквадратическое отклонение?

10. Охарактеризовать следующие законы распределения дискретных случайных величин:

– равномерное распределение на конечном множестве;

– биномиальное распределение;

– распределение Пуассона;

– геометрическое распределение;

– гипергеометрическое распределение.

## Лабораторная работа № 3
## Законы распределения непрерывных случайных величин

**Цель лабораторной работы:** изучить основные понятия, связанные с законами распределения непрерывных случайных величин, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

*Определение.* **Случайная величина –** это величина, которая принимает определенное значение в результате опыта.

*Определение.* **Дискретной случайной величиной** называется такая величина, которая в результате опыта может принимать определенные значения с определенной вероятностью, образующие счетное множество (множество, элементы которого могут быть занумерованы).

*Определение.* **Непрерывной случайной величиной** называется такая величина, которая может принимать любые значения из некоторого конечного или бесконечного промежутка.

*Определение.* **Законом распределения дискретной случайной величины** называют соответствие между возможными значениями случайной величины и их вероятностями.

### Способы задания непрерывных случайных величин

*Определение.* **Функцией распределения** называют функцию F(x), определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее *x*.

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

Функцию распределения также называют **интегральной функцией**.

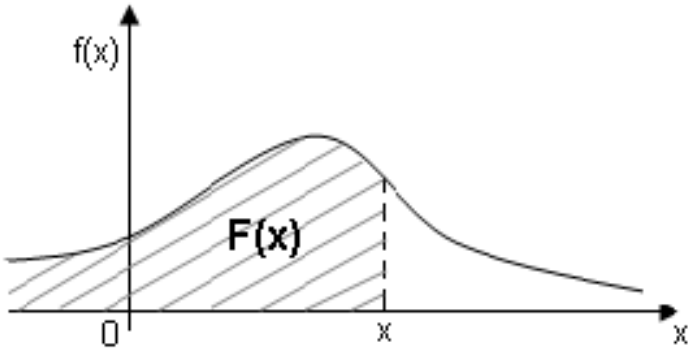



Рис. 3.1

*Определение.* **Плотностью распределения** вероятностей непрерывной случайной величины X называется функция *f(x)* – первая производная от функции распределения F(x).

$$f(x) = F'(x).$$

Зная плотность распределения, можно вычислить вероятность того, что некоторая случайная величина X примет значение, принадлежащее заданному интервалу.

*Теорема.* Вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу (a, b), равна определенному интегралу от плотности распределения, взятому в пределах от a до b.

$$P(a < X < b) = \int_a^b f(x)dx$$

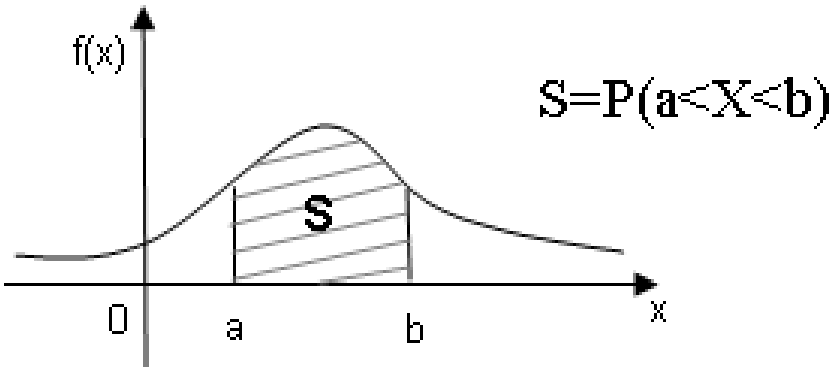



Рис. 3.2

**Числовые характеристики**

1) ***Математическое ожидание M (X)*** непрерывной случайной величины X определяется равенством:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx$$

при условии, что этот интеграл сходится абсолютно.

2) ***Дисперсия D(X)*** непрерывной случайной величины X определяется равенством:

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - [M(X)]^2$$

*3)* ***Среднее квадратическое отклонение σ(X)*** непрерывной случайной величины определяется равенством:

$$\sigma(X) = \sqrt{D(X)}$$

## Основные распределения непрерывных случайных величин

### I. Равномерное распределение $\xi \in R(a,b)$

Для того чтобы случайная величина подчинялась закону равномерного распределения, необходимо, чтобы ее значения лежали внутри некоторого определенного интервала и внутри этого интервала значения этой случайной величины были бы равновероятны.

*Определение.* Непрерывная случайная величина имеет **равномерное распределение на отрезке [a, b]**, если на этом отрезке плотность распределения случайной величины постоянна, а вне его равна нулю.

$$f(x) = \begin{cases} 0, & x < a \\ \dfrac{1}{b-a}, & a \le x \le b \\ 0, & x > b \end{cases}$$

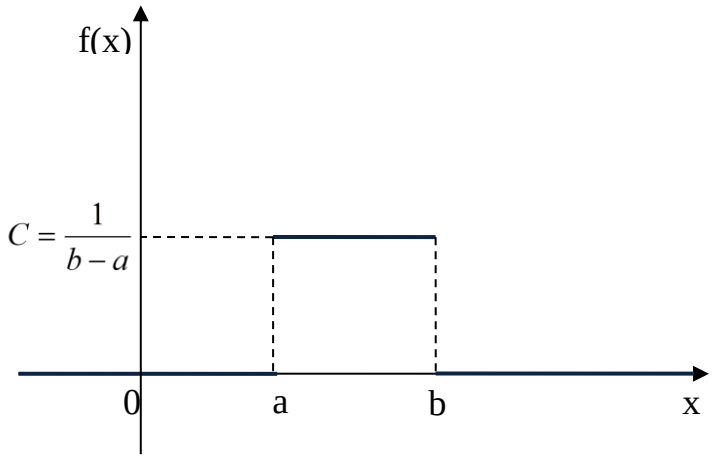



Рис. 3.3

Найдем функцию распределения *F(x)* на отрезке [*a*,*b*].

$$F(x) = \begin{cases} 0, & при \quad x < a \\ \dfrac{x-a}{b-a}, & при \quad a \le x \le b \\ 1, & при \quad x > b \end{cases}$$

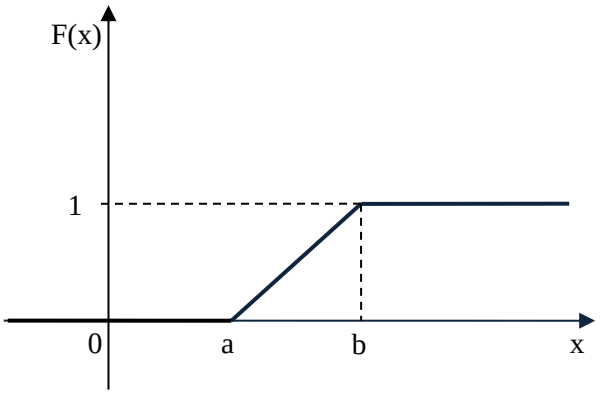



Рис. 3.4

Определим математическое ожидание и дисперсию случайной величины, подчиненной равномерному закону распределения.

$$M(x) = \frac{a+b}{2}. \qquad D(x) = \frac{(b-a)^2}{12}. \qquad \sigma(x) = \sqrt{D(x)} = \frac{b-a}{2\sqrt{3}}.$$

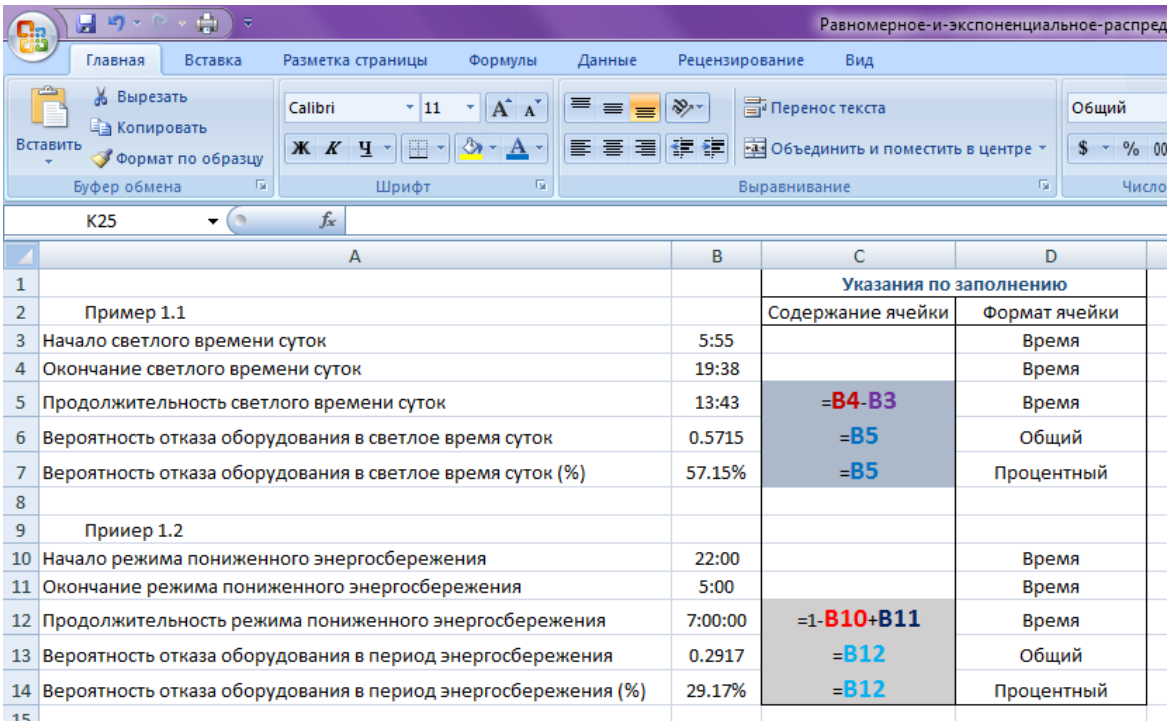
Вероятность попадания случайной величины в заданный интервал:

$$P(\alpha < X < \beta) = \frac{\beta - \alpha}{b-a}.$$

*Пример 3.1.* Предположим, что моменты отказов устройства для контроля за чистотой воздуха **равномерно распределены** в течение суток.

1. В некий день светлое время суток наступает в 5:55 и заканчивается в 19:38. Какова вероятность того, что отказ оборудования устройства произойдет в течение светлого времени суток?
2. Допустим, что с 22:00 до 5:00 устройство переходит в режим пониженного энергопотребления. Какова вероятность того, что отказ произойдет в указанный период времени?
3. Предположим, что в состав устройства входит процессор, каждый час осуществляющий проверку работоспособности оборудования. Какова вероятность того, что отказ будет обнаружен не позднее, чем через 10 мин?
4. Предположим, что в состав устройства входит процессор, каждый час осуществляющий проверку работоспособности оборудования. Какова вероятность того, что отказ будет обнаружен не раньше, чем через 40 мин?

**Решение с помощью табличного процессора.**




| | A | B | C | D |
|---|---|---|---|---|
| 1 | | | Указания по заполнению | |
| 2 | Пример 1.1 | | Содержание ячейки | Формат ячейки |
| 3 | Начало светлого времени суток | 5:55 | | Время |
| 4 | Окончание светлого времени суток | 19:38 | | Время |
| 5 | Продолжительность светлого времени суток | 13:43 | =B4-B3 | Время |
| 6 | Вероятность отказа оборудования в светлое время суток | 0.5715 | =B5 | Общий |
| 7 | Вероятность отказа оборудования в светлое время суток (%) | 57.15% | =B5 | Процентный |
| 8 | | | | |
| 9 | Прииер 1.2 | | | |
| 10 | Начало режима пониженного энергосбережения | 22:00 | | Время |
| 11 | Окончание режима пониженного энергосбережения | 5:00 | | Время |
| 12 | Продолжительность режима пониженного энергосбережения | 7:00:00 | =1-B10+B11 | Время |
| 13 | Вероятность отказа оборудования в период энергосбережения | 0.2917 | =B12 | Общий |
| 14 | Вероятность отказа оборудования в период энергосбережения (%) | 29.17% | =B12 | Процентный |

Рис. 3.5

1. Поскольку в условии задачи сказано, что моменты отказов устройства равномерно распределены в течение суток, вероятность отказа в светлое время суток – есть доля этого времени суток.

P (отказа в светлое время суток) = 19:38 – 5:55 = 57,2%.

Если представить разность окончания и начала светлого времени суток в процентном формате, то получим ответ – 57,2%. Хитрость заключается в том, что в Microsoft Excel сутки – это единица, один час – 1/24, таким образом интервал времени меньше суток будет составлять процентную часть этих суток.

2. P (отказа с 22:00 до 5:00) = 2:99 + 5:00 = 29,17%.

3. P(обнаружения отказа не позднее, чем через 10 мин) = 10/60 = 16,7%.

4. P (обнаружения отказа не раньше, чем через 40 мин) = (60–40)/60 = 3,3%.

## II. Нормальное распределение $\xi \in N(a,\sigma^2)$

*Определение.* Непрерывная случайная величина называется распределенной по **нормальному закону** (распределение Гаусса) с параметрами $M\xi = a$ и $D\xi = \sigma^2$, если ее плотность распределения равна

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Через $\xi \in N(a,\sigma^2)$ обозначается множество всех случайных величин, распределенных по нормальному закону с параметрами $a$ и $\sigma^2$.

Функция распределения нормально распределенной случайной величины равна

$$F_{\xi}(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt\,.$$

Такой интеграл нельзя вычислить аналитически, и потому для функции $F(x)$ составлены таблицы (*см. Приложение 1*). Функция $F(x)$ связана с функцией Лапласа

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{\tilde{o}}e^{-\frac{t^2}{2}}dt\,,$$

следующим соотношением $F_{\xi}(x)=\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right)$.

Поэтому вероятность попадания нормально распределенной случайной величины $\xi\in N(a,\sigma^2)$ на интервал $(c_1,c_2)$ можно вычислять по формуле

$$P\{c_1\le\xi<c_2\}=\Phi\left(\frac{c_2-a}{\sigma}\right)-\Phi\left(\frac{c_1-a}{\sigma}\right).$$

*Определение.* Когда $a=0$ и $\sigma^2=1$ **нормальное распределение** называется **стандартным**, и класс таких распределений обозначается $\xi\in N(0,1)$.

В этом случае плотность стандартного распределения равна

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

а функция распределения

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

*Определение.* Неотрицательная случайная величина ξ называется **логарифмически нормально распределенной**, если ее логарифм $\eta=ln\xi$ подчинен нормальному закону. Математическое ожидание и дисперсия логарифмически нормально распределенной случайной величины равны $M\xi=ae^{-\sigma^2}$ и $D\xi=a^2e^{\sigma^2}(e^{\sigma^2}-1)$.

Пример 3.2. Время загрузки Web-страницы **распределено нормально**, причем его математическое ожидание равно *a* = 7 с, а стандартное отклонение σ = 2 с.

1. Определите вероятность того, что время загрузки превысит 9 с.
2. Определите вероятность того, что время загрузки лежит в интервале 7–9 с.

3. Определите вероятность того, что время загрузки лежит в интервале 5–9 с.

4. Найдите значение переменной X, соответствующей интегральной вероятности, равной 0,1. Сколько секунд длится загрузка Web-страницы в 10% случаев?

**Решение с помощью табличного процессора.**

1. Вероятность того, что время загрузки не превысит 9 с, равна 0,8413, следовательно, искомая вероятность равна 1 – 0,8413 = 0,1587.

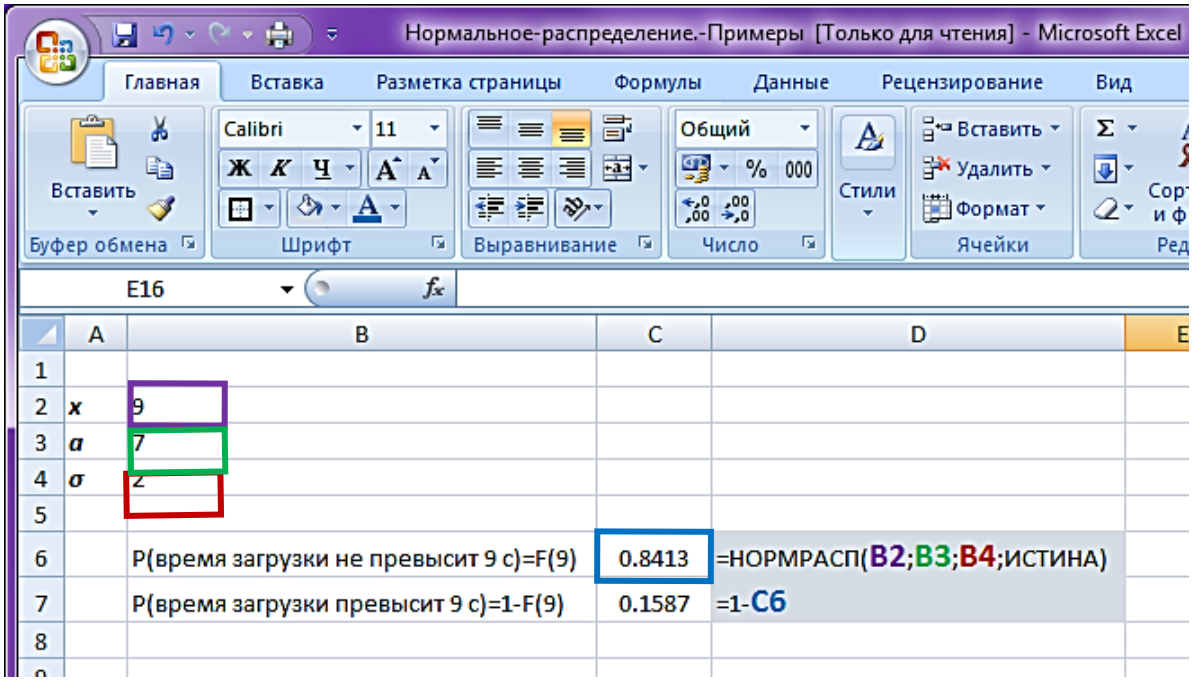



Рис. 3.6

2. $P(7 < X < 9) = P(X < 9) - P(X < 7)$.

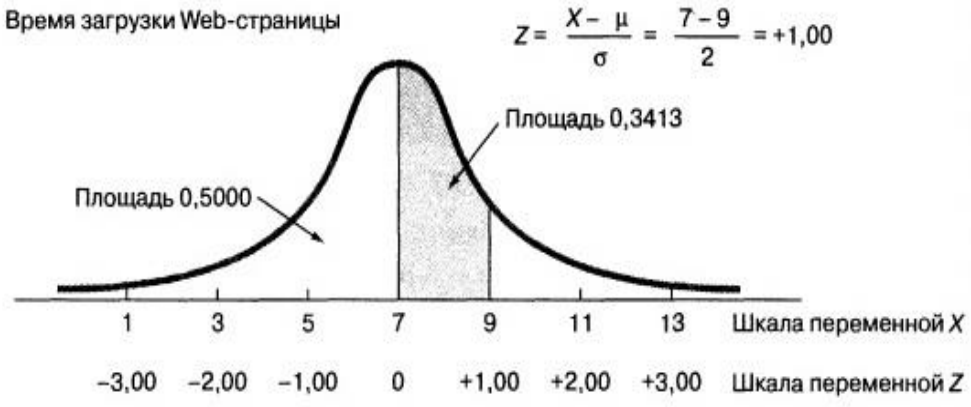



Рис. 3.7

В Excel есть функция для нестандартизированного нормального распределения.

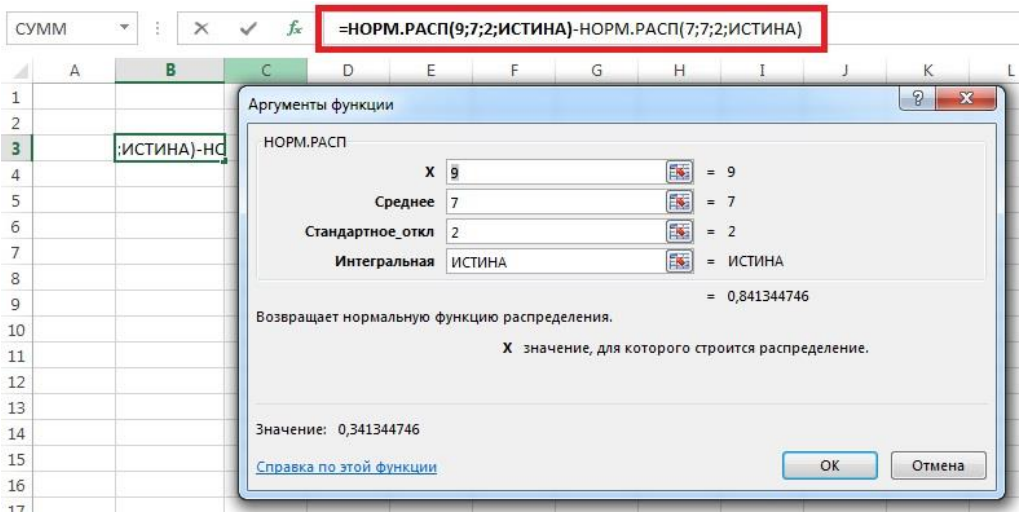



Рис. 3.8

Обратите внимание, что, поскольку математическое ожидание и медиана нормального распределения совпадают между собой, вероятность того, что загрузка продлится меньше 7 с, равна 0,5, то есть, =НОРМ.РАСП(7;7;2;ИСТИНА) = 0,5.

3. $P(5 < X < 9) = P(X < 9) – P(X < 5)$ =НОРМ.РАСП(9;7;2;ИСТИНА)-НОРМ.РАСП(5;7;2;ИСТИНА) = 0,6826.

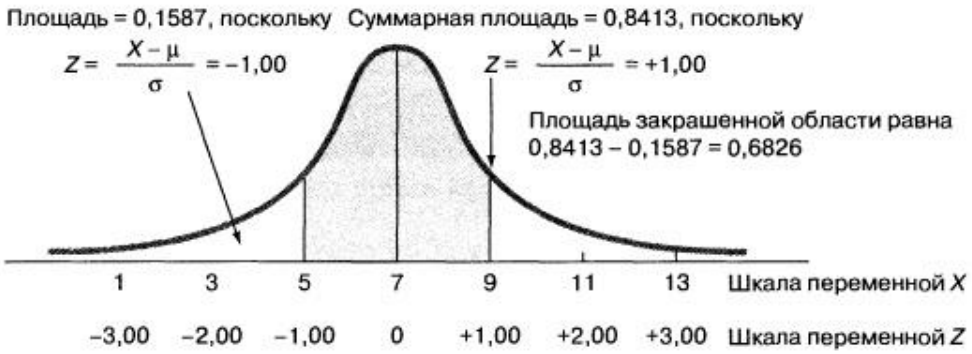



Рис. 3.9

Полученный результат довольно важен. Для любого нормального распределения вероятность того, что случайно выбранное число лежит в окрестности математического ожидания на расстоянии, не превышающем одно стандартное отклонение, равно 0,6826. В окрестности математического ожидания на расстоянии, не превышающем двух стандартных отклонений, лежит чуть более 95% нормально распределенных величин.

Выше мы вычислили вероятности, связанные с разными значениями измеренной величины.

4. Поскольку предполагается, что в 10% случаев Web-страница загружается не более чем за X с, площадь фигуры, ограниченной гауссовой кривой и осью абсцисс, равна 0,1. Для обратной задачи в Excel до версии 2007 существуют две функции =НОРМСТОБР() – возвращает обратное

значение стандартного нормального распределения, и =НОРМОБР() – возвращает обратное нормальное распределение (не стандартизированное). В версии Excel, начиная с 2010, им соответствуют функции: =НОРМ.СТ.ОБР() и =НОРМ.ОБР(). В нашем примере =НОРМ.ОБР(0,1;7;2) = 4,4 с.

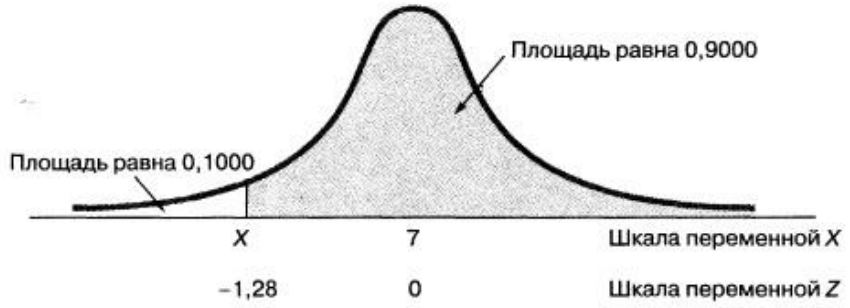



Рис. 3.10

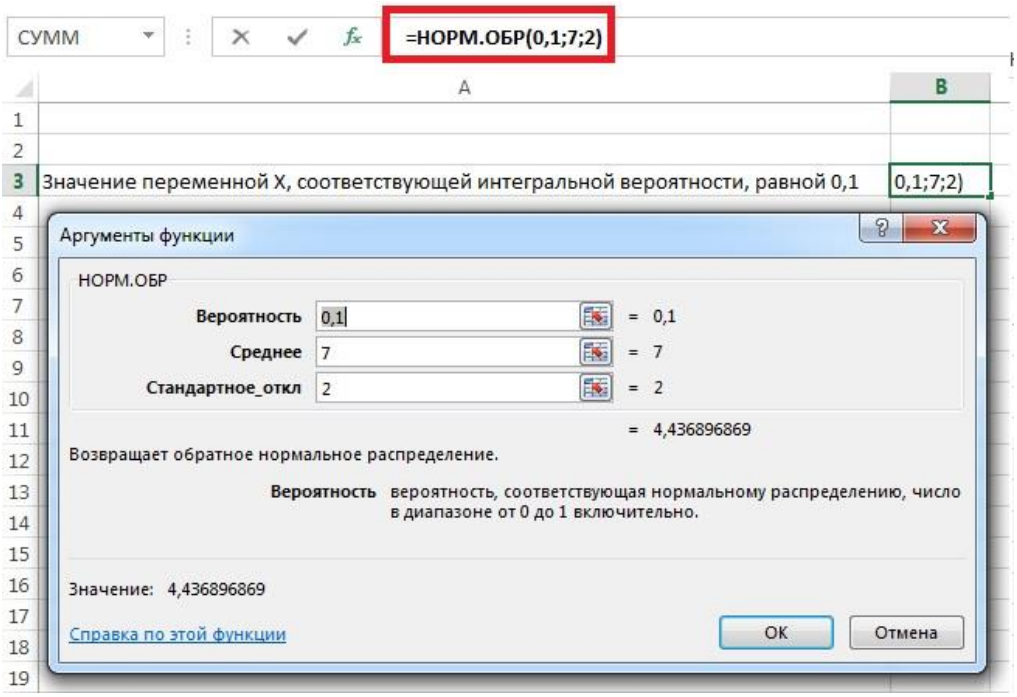



Рис. 3.11

*Пример 3.3.* Для закупки и последующей реализации мужских курток фирмой было проведено выборочное обследование мужского населения города в возрасте от 18 до 65 лет в целях определения его среднего роста. В результате было установлено, что средний рост 176 см, стандартное отклонение 6 см. Необходимо определить, какой процент общего числа закупаемых курток должны составлять куртки пятого роста (182–186 см). Предполагается, что рост мужского населения распределен **по нормальному закону**. Построить графики функции и плотности распределения.

**Решение с помощью табличного процессора.**

Сначала в MS Excel создадим последовательность случайных чисел, распределенных по нормальному закону. Для этого:

1. Вызовем из меню Сервис команду «Анализ данных»→«Генерация случайных чисел».

2. Заполним диалоговое окно.

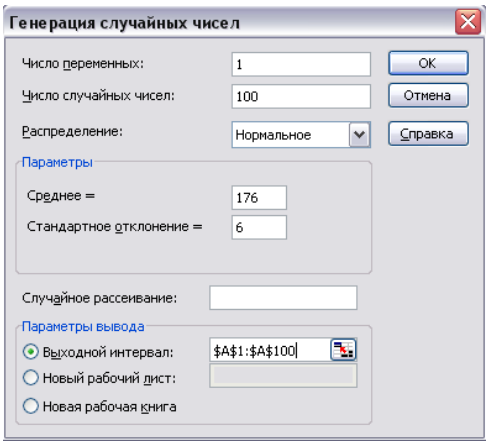



Рис. 3.12

После нажимаем кнопку OK и в указанном нами диапазоне получаем последовательность псевдослучайных чисел.

3. Упорядочиваем их по возрастанию.

4. Для подсчета значений функции плотности нормального распределения используем функцию НОРМРАСП с аргументом *интегральная* = ЛОЖЬ.

5. Значения функции распределения считаем, используя функцию НОРМРАСП с аргументом *интегральная* = ИСТИНА.

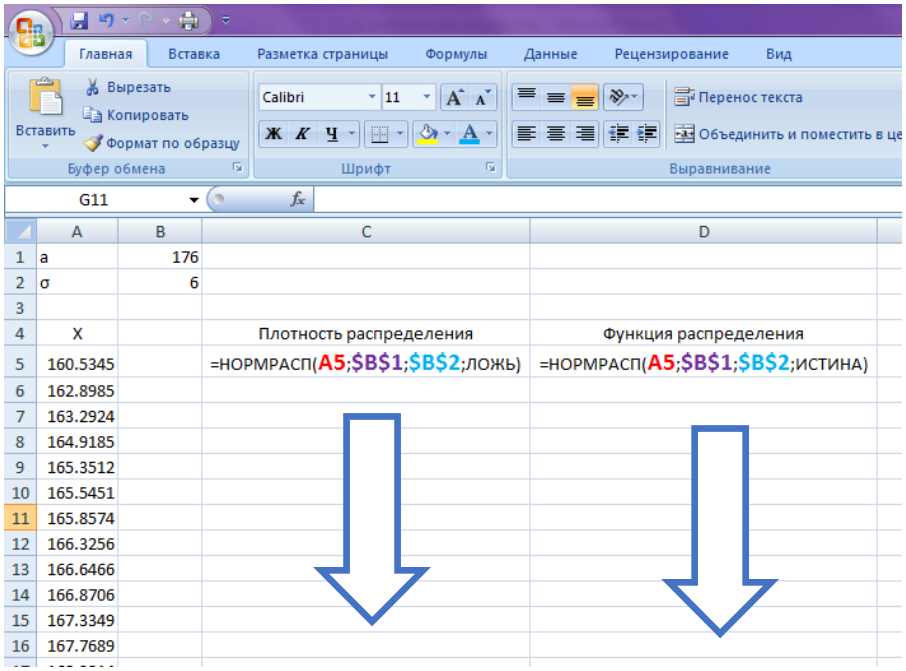
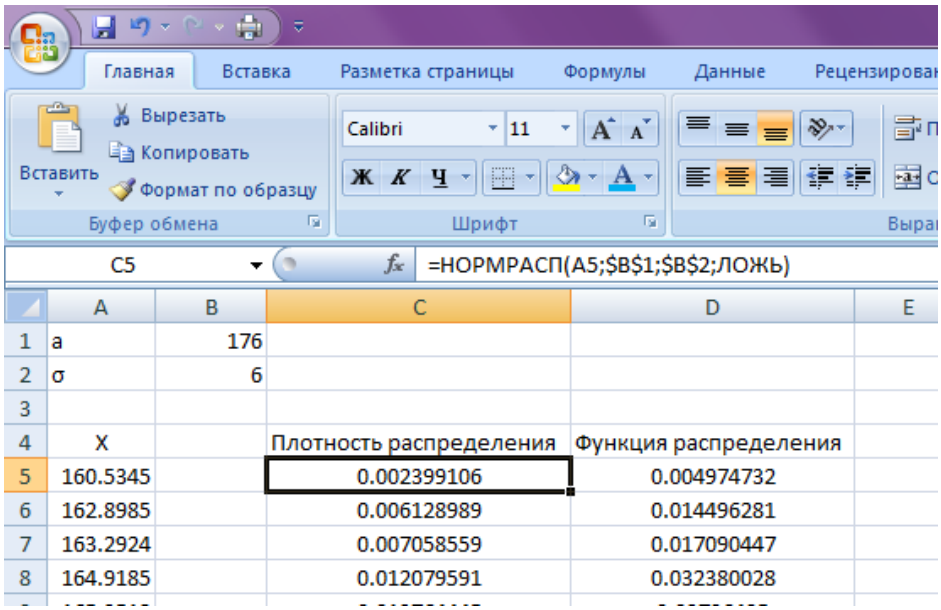



Рис. 3.13




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | a | 176 | | | |
| 2 | σ | 6 | | | |
| 3 | | | | | |
| 4 | X | | Плотность распределения | Функция распределения | |
| 5 | 160.5345 | | 0.002399106 | 0.004974732 | |
| 6 | 162.8985 | | 0.006128989 | 0.014496281 | |
| 7 | 163.2924 | | 0.007058559 | 0.017090447 | |
| 8 | 164.9185 | | 0.012079591 | 0.032380028 | |

Рис. 3.14

6. Решаем задачу по формуле:
=НОРМРАСП(186;176;6;ИСТИНА)-НОРМРАСП(182;176;6;ИСТИНА)=
=0,949675-0,84109=10,858≈11%

Вывод: куртки 5-го роста должны составлять приблизительно 11% от общего числа закупаемых курток.

7. Строим графики полученных функций, используя мастер диаграмм. Выбираем тип диаграммы – «*точечная*», вид – «*со сглаживающими линиями без маркеров*».

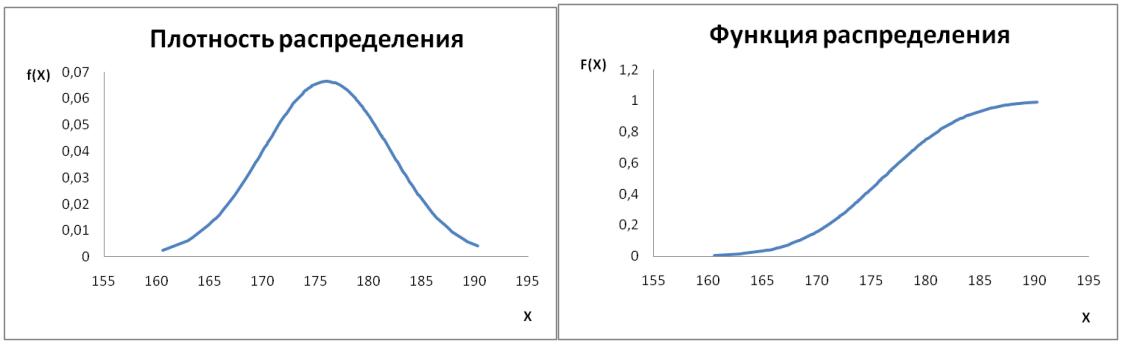



Рис. 3.15

## III. Экспоненциальное распределение $\xi \in E(\lambda)$

*Определение.* Непрерывная случайная величина X имеет **показательный (экспоненциальный) закон распределения** с параметром λ>0, если функция плотности распределения вероятностей имеет вид:

$$f(x) = \begin{cases} 0 & при\ x < 0, \\ \lambda e^{-\lambda x} & при\ x \geq 0. \end{cases}$$

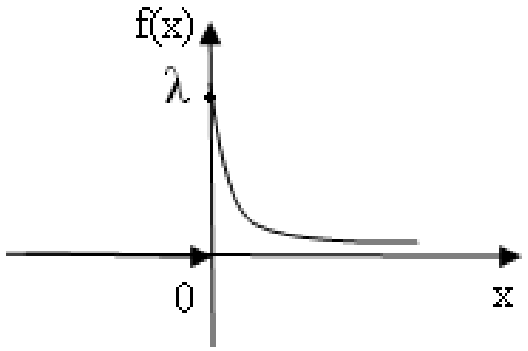



Рис. 3.16

Функция распределения случайной величины X, распределенной по показательному закону, задается формулой:

$$F(x)=\begin{cases} 0 & при\ x<0, \\ 1-e^{-\lambda x} & при\ x\geq 0. \end{cases}$$

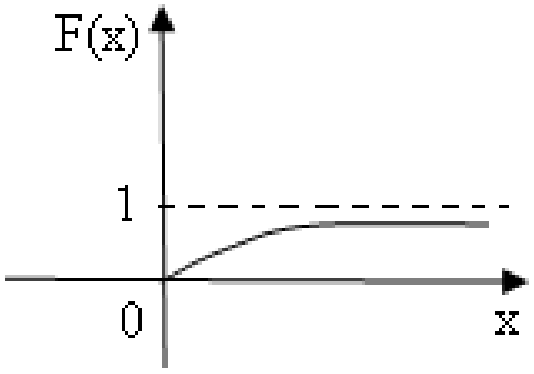



Рис. 3.17

Математическое ожидание, дисперсия и среднее квадратическое отклонение показательного распределения соответственно равны:

$$M(x)=\frac{1}{\lambda}. \qquad D(x)=\frac{1}{\lambda^2}. \qquad \sigma(x)=\sqrt{D(x)}=\frac{1}{\lambda}.$$

Таким образом, математическое ожидание и среднее квадратическое отклонение показательного распределения равны между собой.

Вероятность попадания X в интервал (a;b) вычисляется по формуле:

$$P(a<X<b)=e^{-\lambda a}-e^{-\lambda b}$$

*Пример 3.4.* **(Экспоненциальное распределение)** В отделение банка приходят 20 клиентов в час. Предположим, что в банк уже пришел один клиент. Какова вероятность того, что следующий клиент придет в течение 6 мин?

*Решение.* В данном случае λ = 20, X= 0,1 (6 мин = 0,1 ч). Используя формулу функции распределения

$$F(x)=\begin{cases}0 & при\ x<0,\\ 1-e^{-\lambda x} & при\ x\ge 0.\end{cases}$$, получаем:

P(время прихода второго клиента < 0,1) = 1 – $e^{-20*0,1}$ = 0,8647.

Таким образом, вероятность, что следующий клиент придет в течение 6 мин, равна 86,47%. Экспоненциальное распределение можно вычислить с помощью функции Excel "=ЭКСПРАСП(x, λ, ИСТИНА)".

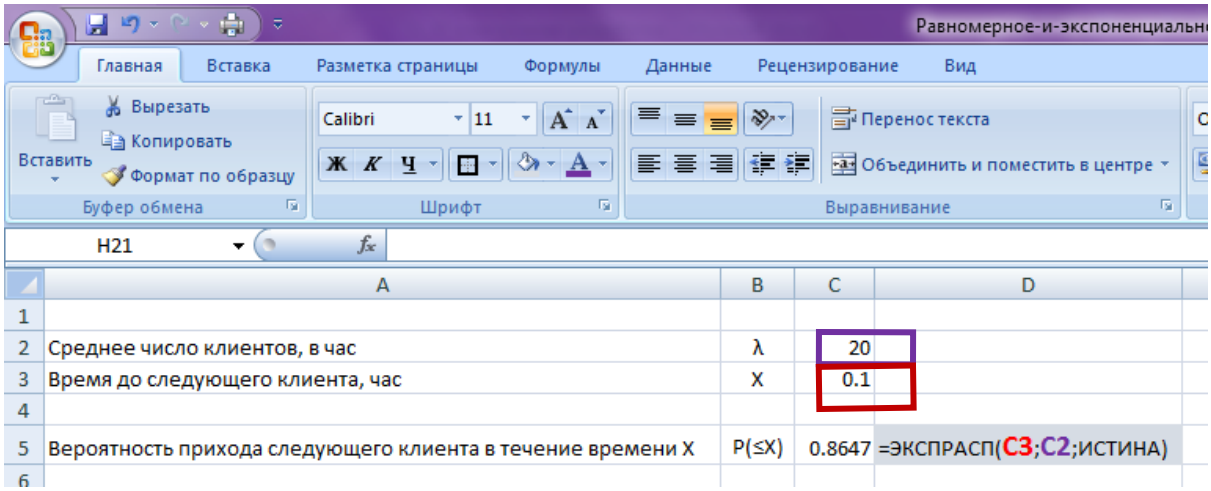



Рис. 3.18

➢Указание.

1) Плотность распределения

$$f(x)=\begin{cases}0 & при\ x<0,\\ \lambda e^{-\lambda x} & при\ x\ge 0.\end{cases}$$ = ЭКСПРАСП(*x*, *λ*, ЛОЖЬ)

2) Функция распределения

$$F(x)=\begin{cases}0 & при\ x<0,\\ 1-e^{-\lambda x} & при\ x\ge 0.\end{cases}$$ = ЭКСПРАСП(*x*, *λ*, ИСТИНА)

## Задание для самостоятельного выполнения

**Задание 1.** Сгенерировать 30 случайных чисел, **равномерно распределенных** на отрезке **[a, b]**.

**Задание 2.** При фасовке творога автомат отмеряет порции по ***a*** г. Количество творога распределено по **нормальному закону *N(a,σ²)*.** Допуск равен ***a***±5 г. Найти вероятность того, что случайно выбранная пачка творога:

а) имеет массу в пределах допуска;

б) имеет массу больше допустимой;

в) имеет массу меньше допустимой;

г) имеет массу в интервале от ***a***-1 до ***a***+1.

| Номер варианта | Равномерное распределение | Нормальное распределение | | Экспоненциальное распределение |
|---|---|---|---|---|
| | [a, b] | $a$ | $\sigma$ | $\lambda$ |
| **1** | [ 0,1] | 110 | 1 | 10 |
| **2** | [ 0,2] | 120 | 2 | 11 |
| **3** | [ 0,3] | 130 | 3 | 12 |
| **4** | [ 0,4] | 140 | 4 | 13 |
| **5** | [ 0,5] | 150 | 1 | 14 |
| **6** | [-1,1] | 160 | 2 | 15 |
| **7** | [-2,2] | 170 | 3 | 16 |
| **8** | [-3,3] | 180 | 4 | 17 |
| **9** | [-4,4] | 190 | 1 | 18 |
| **10** | [-5,5] | 200 | 2 | 19 |
| **11** | [-1,0] | 210 | 3 | 20 |
| **12** | [-2,0] | 220 | 4 | 21 |
| **13** | [-3,0] | 230 | 1 | 22 |
| **14** | [-4,0] | 240 | 2 | 23 |
| **15** | [-5,0] | 250 | 3 | 24 |
| **16** | [ 1,4] | 260 | 4 | 25 |
| **17** | [ 1,2] | 270 | 1 | 26 |
| **18** | [ 1,3] | 280 | 2 | 27 |
| **19** | [ 1,4] | 290 | 3 | 28 |
| **20** | [ 1,5] | 300 | 4 | 29 |
| **21** | [-3,1] | 310 | 1 | 30 |
| **22** | [-6,2] | 320 | 2 | 31 |
| **23** | [-2,3] | 330 | 3 | 32 |
| **24** | [-4,4] | 340 | 4 | 33 |
| **25** | [-3,5] | 350 | 1 | 34 |
| **26** | [-1,7] | 360 | 2 | 35 |
| **27** | [-2,1] | 370 | 3 | 36 |
| **28** | [ 3,5] | 380 | 4 | 37 |
| **29** | [ 4,6] | 390 | 1 | 38 |
| **30** | [-5,1] | 400 | 2 | 39 |

**Задание 3.** (**Экспоненциальное распределение**) У проселочной дороги уставший пешеход ожидает попутную машину. В среднем по дороге в нужном направлении проходит $\lambda$ машин в час. Найдите вероятность того, что, пропустив одну попутку, пешеход будет ждать следующего не более 10 минут.

**Задание 4.** Механические часы положили в шкаф, и через несколько дней они остановились. Найти вероятность того, что часовая стрелка находится между 4 и 6 часами.

**Задание 5.** Рыбак, забросив удочку, в среднем ждет поклевки 4 минуты. Найдите вероятность того, что рыбак, забросив удочку, будет ждать поклевки не менее 5 минут.

## Контрольные вопросы

1. Что называется случайной величиной?
2. Какие случайные величины называются дискретными?
3. Что такое функция распределения?
4. Что такое плотность распределения?
5. Что такое математическое ожидание?
6. Что такое дисперсия?
7. Что такое среднеквадратическое отклонение?
8. Охарактеризовать следующие законы распределения непрерывных случайных величин:

– равномерное распределение;
– нормальное распределение;
– экспоненциальное распределение.

# 2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ДАННЫХ

## Лабораторная работа № 4
## Элементы корреляционного анализа

**Цель лабораторной работы:** изучить основные понятия, связанные с корреляционным анализом, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

*Определение.* **Ковариацией** случайных величин $\boldsymbol{x}$ и $\boldsymbol{y}$ называется величина $cov(x,y)$, вычисляемая по формуле

$$cov(x,y) = M[(x-\overline{x})(y-\overline{y})] = \overline{xy} - \overline{x}\cdot\overline{y}.$$

Отличие этой величины от нуля указывает на зависимость $\boldsymbol{x}$ и $\boldsymbol{y}$. Ковариация зависит от размерности случайных величин $\boldsymbol{x}$ и $\boldsymbol{y}$ и поэтому удобнее использовать следующую характеристику.

*Определение.* Для измерения тесноты связи используют безразмерный **линейный коэффициент корреляции** $r_{xy} = \frac{cov(x,y)}{\sigma_x\sigma_y}$, где $\sigma_x$ и $\sigma_y$ – среднеквадратические отклонения.

Линейный коэффициент корреляции принимает значения от –1 до +1. Положительные значения $r_{xy}$ свидетельствуют о прямой зависимости случайных величин, а отрицательные – об обратной; модуль характеризует силу связи, которая может быть оценена, например, по шкале Чеддока:

$0{,}1 < |r_{xy}| < 0{,}3$: слабая;
$0{,}3 < |r_{xy}| < 0{,}5$: умеренная;
$0{,}5 < |r_{xy}| < 0{,}7$: заметная;
$0{,}7 < |r_{xy}| < 0{,}9$: высокая;
$0{,}9 < |r_{xy}| < 1$ : весьма высокая.

*Пример 4.1.* Найти коэффициенты ковариации и корреляции случайных величин $x_i$ и $y_i$ и охарактеризовать вид связи между ними.

| $\boldsymbol{Y_i}$ | 4 | 5 | 6 | 3 | 1 | 5 |
|---|---|---|---|---|---|---|
| $\boldsymbol{x_i}$ | 1 | 2 | 3 | 4 | 5 | 6 |

**Решение:**

| | $y_i$ | $x_i$ | $x_i y_i$ | $y_i^2$ | $x_i^2$ |
|---|---|---|---|---|---|
| | 4 | 1 | 4 | 16 | 1 |
| | 5 | 2 | 10 | 25 | 4 |
| | 6 | 3 | 18 | 36 | 9 |
| | 3 | 4 | 12 | 9 | 16 |
| | 1 | 5 | 5 | 1 | 25 |
| | 5 | 6 | 30 | 25 | 36 |
| ***Итого*** | 24 | 21 | 79 | 112 | 91 |
| ***Среднее*** | $\overline{y} = \frac{\sum y_i}{n} = 4$ | $\overline{x} = \frac{\sum x_i}{n} = 3{,}5$ | $\overline{xy} = \frac{\sum xy}{n} = \frac{79}{6} \approx 13{,}17$ | $\overline{y^2} = \frac{\sum y^2}{n} = \frac{112}{6} \approx 18{,}67$ | $\overline{x^2} = \frac{\sum x^2}{n} = \frac{91}{6} \approx 15{,}17$ |

$$cov(x,y) = M[(x-\overline{x})(y-\overline{y})] = \overline{xy} - \overline{x}\cdot\overline{y} \approx 13{,}2 - 3{,}5\cdot 4 = 13{,}2 - 14 = -0{,}83$$

$$\sigma_x^2 = \overline{x^2} - (\overline{x})^2 \approx 15{,}17 - 3{,}5^2 \approx 2{,}92 \quad \sigma_x \approx 1{,}71$$

$$\sigma_y^2 = \overline{y^2} - (\overline{y})^2 \approx 18{,}67 - 4^2 \approx 2{,}67 \quad \sigma_y \approx 1{,}63$$

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y} \approx \frac{-0{,}83}{1{,}71\cdot 1{,}63} \approx -0{,}3$$

**Вывод:** Умеренная обратная взаимосвязь случайных величин ***x*** и ***y*.**

*Пример 4.2.* Вычислить коэффициенты ковариации cov(x,y) и корреляции R двух случайных величин.

| $y_i$ | $x_i$ |
|---|---|
| 1 | 30 |
| 2 | 70 |
| 4 | 150 |
| 3 | 100 |
| 5 | 170 |
| 3 | 100 |
| 4 | 150 |

**Решение с помощью табличного процессора.**

Для расчета используем следующие последовательности пунктов меню:

1) *«Функция» → «Статистические»→ «КОВАР»*.
2) *«Функция» → «Статистические» → «КОРРЕЛ»*.

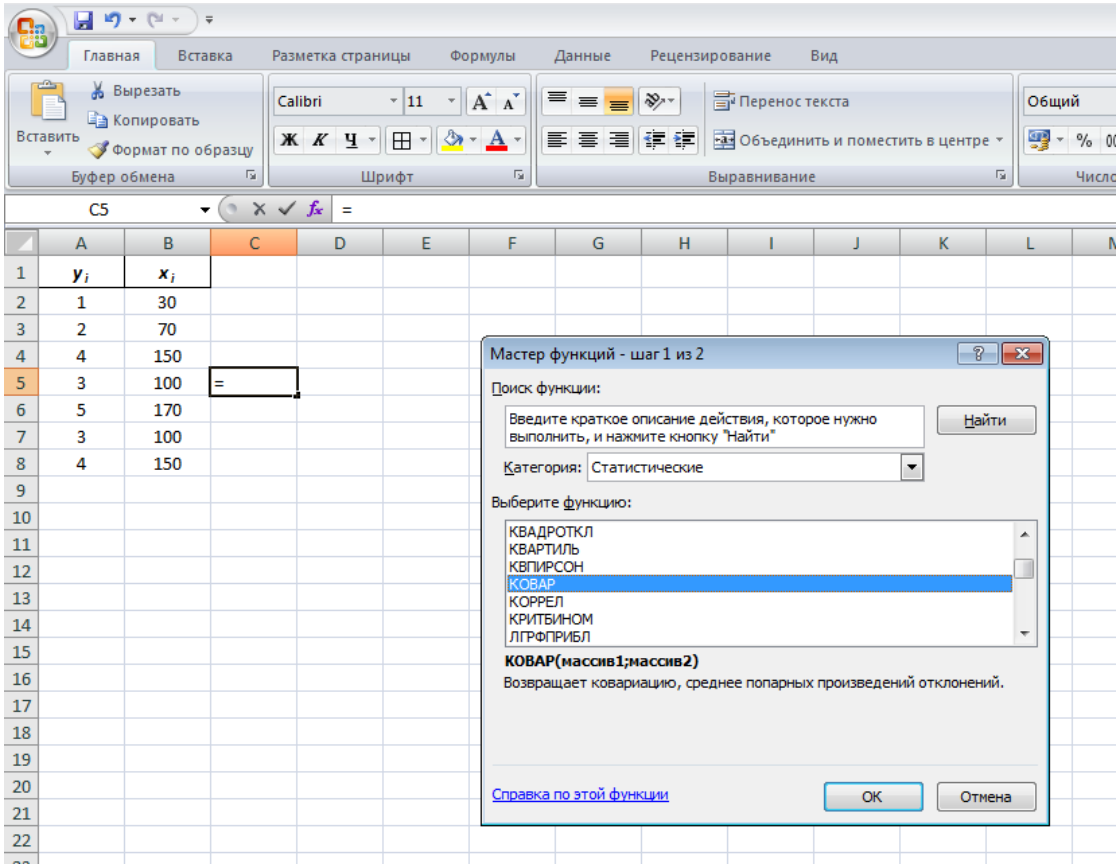

Главная
Вставка
Разметка страницы
Формулы
Данные
Рецензирование
Вид
Вырезать
Копировать
Формат по образцу
Вставить
Буфер обмена
Calibri
11
Шрифт
Перенос текста
Объединить и поместить в центре
Выравнивание
Общий
C5
=
Мастер функций - шаг 1 из 2
Поиск функции:
Введите краткое описание действия, которое нужно выполнить, и нажмите кнопку "Найти"
Найти
Категория: Статистические
Выберите функцию:
КВАДРОТКЛ
КВАРТИЛЬ
КВПИРСОН
КОВАР
КОРРЕЛ
КРИТБИНОМ
ЛГРФПРИБЛ
КОВАР(массив1;массив2)
Возвращает ковариацию, среднее попарных произведений отклонений.
Справка по этой функции
OK
Отмена

Рис. 4.1

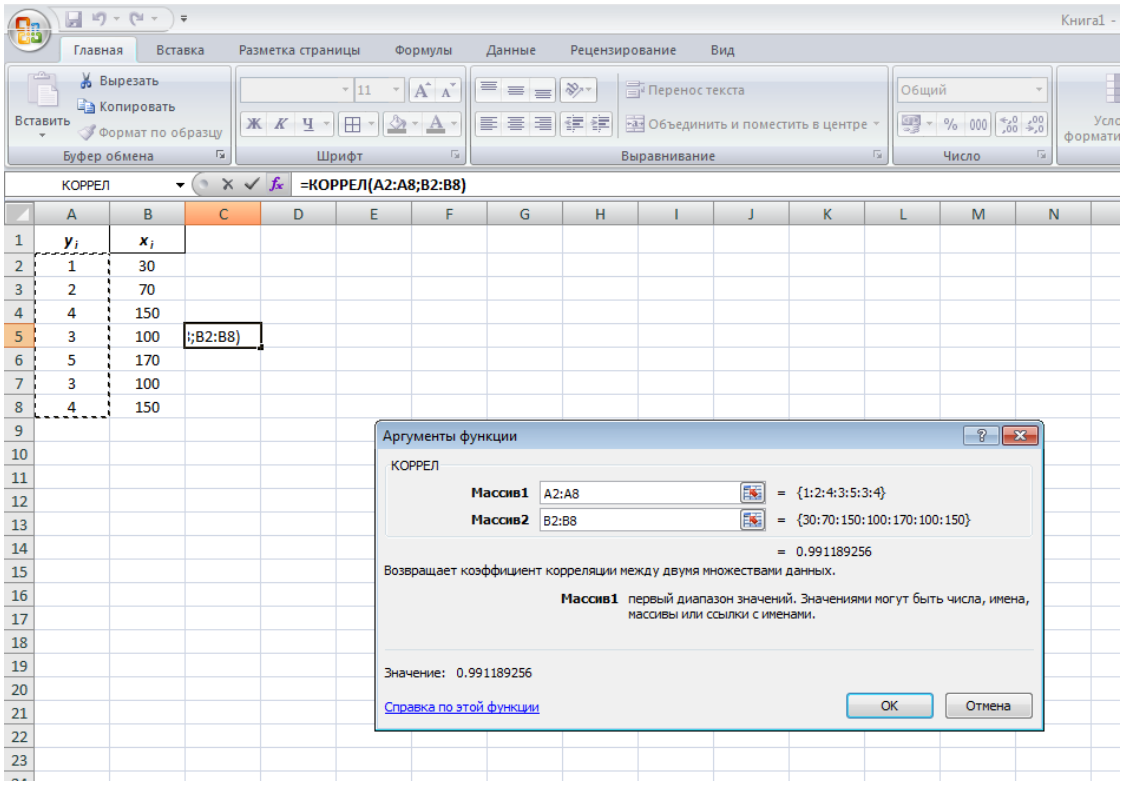
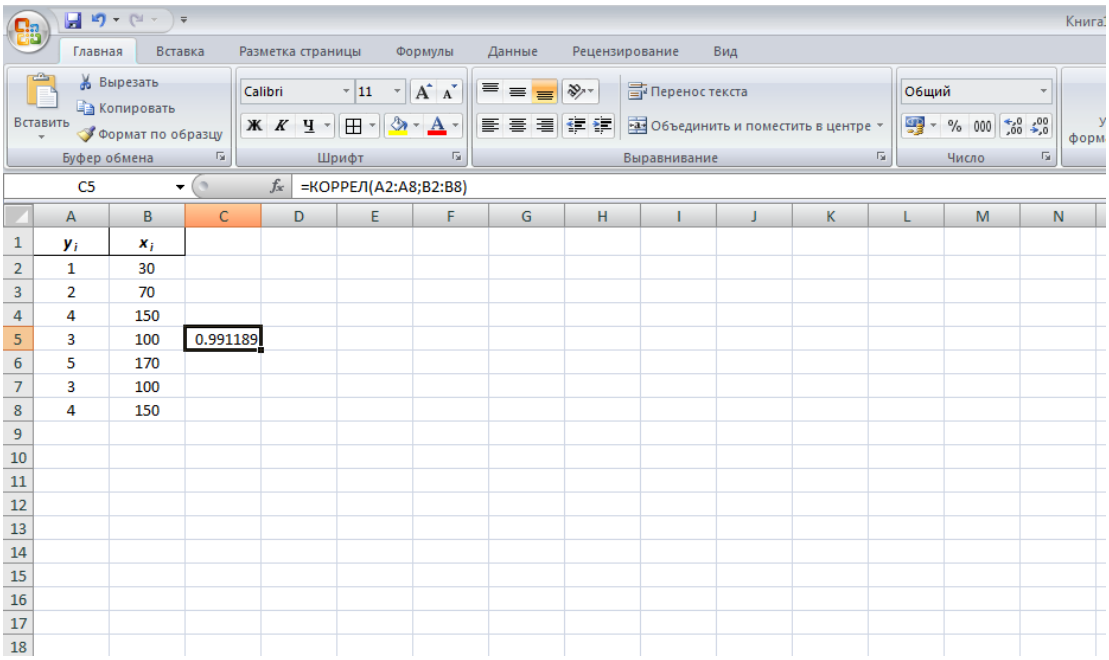

Книга1
Главная
Вставка
Разметка страницы
Формулы
Данные
Рецензирование
Вид
КОРРЕЛ
=КОРРЕЛ(A2:A8;B2:B8)
;B2:B8)
Аргументы функции
КОРРЕЛ
Массив1 A2:A8 = {1:2:4:3:5:3:4}
Массив2 B2:B8 = {30:70:150:100:170:100:150}
= 0.991189256
Возвращает коэффициент корреляции между двумя множествами данных.
Массив1 первый диапазон значений. Значениями могут быть числа, имена, массивы или ссылки с именами.
Значение: 0.991189256
Справка по этой функции
OK
Отмена

Рис. 4.2




| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $y_i$ | $x_i$ | | | | | | | | | | | | |
| 2 | 1 | 30 | | | | | | | | | | | | |
| 3 | 2 | 70 | | | | | | | | | | | | |
| 4 | 4 | 150 | | | | | | | | | | | | |
| 5 | 3 | 100 | 0.991189 | | | | | | | | | | | |
| 6 | 5 | 170 | | | | | | | | | | | | |
| 7 | 3 | 100 | | | | | | | | | | | | |
| 8 | 4 | 150 | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | |
| 11 | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | | | | | |
| 13 | | | | | | | | | | | | | | |
| 14 | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | |

Рис. 4.3

Связи между признаками могут быть слабыми и сильными (тесными). Их критерии оцениваются, например, по шкале Чеддока:

$0{,}1 < |r_{xy}| < 0{,}3$: слабая;
$0{,}3 < |r_{xy}| < 0{,}5$: умеренная;
$0{,}5 < |r_{xy}| < 0{,}7$: заметная;
$0{,}7 < |r_{xy}| < 0{,}9$: высокая;
$0{,}9 < |r_{xy}| < 1$: весьма высокая;

**Вывод:** Связь между случайными величинами *x* и *y* весьма высокая.

### Задание для самостоятельного выполнения

**Задание.** Найти коэффициенты ковариации и корреляции случайных величин $x_i$ и $y_i$ и охарактеризовать вид связи между ними (см. стр. 55)

### Контрольные вопросы

1. Что такое коэффицент ковариация?
2. Что такое линейный коэффицент корреляции?
3. Что такое коэффицент детерминации?
4. Что значит прямая и обратная взаимосвязь?
5. Что значит сильная или слабая взаимосвязь?

Вариант 1

| *X* | *Y* |
|---|---|
| 21 | 20 |
| 33 | 22 |
| 36 | 15 |
| 45 | 55 |
| 54 | 39 |
| 60 | 60 |
| 69 | 30 |
| 75 | 60 |
| 84 | 45 |
| 90 | 70 |
| 99 | 80 |
| 105 | 90 |
| 114 | 79 |
| 120 | 100 |
| 129 | 189 |
| 135 | 200 |
| 144 | 199 |
| 150 | 204 |
| 159 | 220 |
| 165 | 214 |

Вариант 2

| *X* | *Y* |
|---|---|
| 24 | 13 |
| 35 | 12 |
| 13 | 14 |
| 43 | 27 |
| 38 | 27 |
| 41 | 17 |
| 45 | 30 |
| 48 | 30 |
| 52 | 45 |
| 55 | 32 |
| 59 | 31 |
| 62 | 30 |
| 66 | 63 |
| 69 | 50 |
| 73 | 65 |
| 76 | 76 |
| 80 | 81 |
| 83 | 85 |
| 87 | 90 |
| 90 | 94 |

Вариант 3

| *X* | *Y* |
|---|---|
| 115 | 20 |
| 105 | 22 |
| 103 | 60 |
| 96 | 80 |
| 90 | 103 |
| 84 | 70 |
| 78 | 120 |
| 72 | 120 |
| 66 | 100 |
| 60 | 105 |
| 54 | 130 |
| 48 | 110 |
| 42 | 135 |
| 36 | 127 |
| 30 | 140 |
| 24 | 129 |
| 18 | 150 |
| 12 | 148 |
| 6 | 152 |
| 1 | 155 |

Вариант 4

| *X* | *Y* |
|---|---|
| 6 | 40 |
| 12 | 45 |
| 21 | 50 |
| 27 | 70 |
| 36 | 80 |
| 42 | 78 |
| 51 | 71 |
| 57 | 90 |
| 66 | 95 |
| 72 | 90 |
| 81 | 90 |
| 87 | 89 |
| 96 | 95 |
| 102 | 100 |
| 111 | 110 |
| 117 | 129 |
| 126 | 150 |
| 132 | 148 |
| 141 | 139 |
| 147 | 140 |

Вариант 5

| *X* | *Y* |
|---|---|
| 20 | 15 |
| 25 | 7 |
| 35 | 15 |
| 40 | 10 |
| 50 | 10 |
| 55 | 15 |
| 65 | 15 |
| 70 | 40 |
| 80 | 50 |
| 85 | 50 |
| 95 | 50 |
| 100 | 90 |
| 110 | 88 |
| 115 | 90 |
| 125 | 120 |
| 130 | 150 |
| 140 | 130 |
| 145 | 150 |
| 155 | 177 |
| 160 | 220 |

Вариант 6

| *X* | *Y* |
|---|---|
| 156 | 40 |
| 130 | 45 |
| 138 | 41 |
| 124 | 54 |
| 114 | 50 |
| 106 | 50 |
| 96 | 81 |
| 88 | 90 |
| 78 | 99 |
| 70 | 79 |
| 60 | 117 |
| 52 | 130 |
| 42 | 128 |
| 34 | 165 |
| 24 | 153 |
| 16 | 170 |
| 6 | 200 |
| 2 | 250 |
| 4 | 220 |
| 14 | 198 |

Вариант 7

| *X* | *Y* |
|---|---|
| 13 | 23 |
| 39 | 15 |
| 52 | 9 |
| 78 | 8 |
| 104 | 7 |
| 117 | 6 |
| 143 | 5 |
| 156 | 4 |
| 182 | 3 |
| 208 | 2 |
| 234 | 1 |
| 247 | 1 |
| 273 | 1 |
| 299 | 1 |
| 312 | 1 |
| 338 | 1 |
| 364 | 1 |
| 390 | 1 |
| 403 | 1 |
| 429 | 1 |

Вариант 8

| *X* | *Y* |
|---|---|
| 700 | 250 |
| 460 | 220 |
| 740 | 180 |
| 680 | 100 |
| 700 | 65 |
| 720 | 50 |
| 740 | 45 |
| 760 | 40 |
| 780 | 35 |
| 800 | 30 |
| 820 | 25 |
| 840 | 20 |
| 860 | 10 |
| 880 | 9 |
| 900 | 8 |
| 920 | 7 |
| 940 | 6 |
| 960 | 5 |
| 980 | 4 |
| 1000 | 3 |

Вариант 9

| *X* | *Y* |
|---|---|
| 130 | 15 |
| 127 | 29 |
| 120 | 15 |
| 116 | 50 |
| 111 | 70 |
| 106 | 40 |
| 101 | 80 |
| 96 | 90 |
| 91 | 50 |
| 86 | 88 |
| 81 | 110 |
| 76 | 120 |
| 71 | 110 |
| 66 | 110 |
| 61 | 120 |
| 56 | 190 |
| 51 | 130 |
| 46 | 150 |
| 41 | 177 |
| 36 | 186 |

Вариант 10

| *X* | *Y* |
|---|---|
| 42 | 114 |
| 66 | 120 |
| 72 | 104 |
| 90 | 99 |
| 108 | 100 |
| 120 | 50 |
| 138 | 55 |
| 150 | 60 |
| 168 | 40 |
| 180 | 35 |
| 198 | 45 |
| 210 | 40 |
| 228 | 35 |
| 240 | 25 |
| 258 | 15 |
| 270 | 10 |
| 288 | 15 |
| 300 | 33 |
| 318 | 29 |
| 330 | 22 |

## Лабораторная работа № 5
## Парная регрессия

**Цель лабораторной работы:** изучить основные понятия, связанные с парной регрессией, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

При изучении взаимосвязи случайных величин оказывается, что одному значению случайной величины *x* может соответствовать целое множество значений другой случайной величины *y*, каждое со своей вероятностью – условный закон распределения. Эту взаимосвязь описывают формулой парной регрессии $y=f(x)+\varepsilon$, где $\varepsilon$ – характеризует погрешность, т. е. отклонение фактического значения случайной величины *y* от теоретического $\hat{\boldsymbol{y}} = \boldsymbol{f}(\boldsymbol{x})$

В зависимости от вида функции *f(x)* выделяют линейную регрессию $\hat{y} = a + bx$ и нелинейные:

1) гиперболическое – $\hat{Y} = a + \frac{b_1}{X_1} + \frac{b_2}{X_2} + \cdots + \frac{b_m}{X_m}$;
2) полулогарифмическое – $\hat{Y} = a + b_1 \cdot ln\, X_1 + b_2 \cdot ln\, X_2 + \cdots + b_m \cdot ln\, X_m$;
3) квадратичное – $\hat{Y} = a + b_1 \cdot (X_1)^2 + b_2 \cdot (X_2)^2 + \cdots + b_m \cdot (X_m)^2$;
4) показательное – $\hat{Y} = a \cdot (b_1)^{X_1} \cdot (b_2)^{X_2} \ldots (b_m)^{X_m}$;
5) степенное – $\hat{Y} = a \cdot (X_1)^{b_1} \cdot (X_2)^{b_2} \ldots (X_m)^{b_m}$.

Если изучается зависимость переменной *y* от нескольких независимых случайных величин (переменных), то уравнение $y=f(x_1, \ldots, x_m)+\varepsilon$ называют ***уравнением множественной регрессии***.

Параметры парной линейной и нелинейной регрессии могут быть оценены с помощью формул, приведенных в следующей таблице.

| | | |
|---|---|---|
| Линейная регрессия<br>$\hat{Y} = a + b \cdot X$ | $b = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\overline{X^2} - (\bar{X})^2}$<br>$a = \bar{Y} - b \cdot \bar{X}$ | $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$<br>$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \overline{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i$ |
| Полиномиальная (квадратичная) регрессия<br>$\hat{Y} = a + b \cdot X^2$ | $b = \frac{\overline{X^2Y} - \overline{X^2} \cdot \bar{Y}}{\overline{X^4} - \left(\overline{X^2}\right)^2}$<br>$a = \bar{Y} - b \cdot \overline{X^2}$ | $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$<br>$\overline{X^4} = \frac{1}{n}\sum_{i=1}^{n} x_i^4, \overline{X^2Y} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \cdot y_i$ |

| | | |
|---|---|---|
| Гиперболическая регрессия<br>$\hat{Y} = a + \frac{b}{X}$ | $b = \frac{\overline{\left(\frac{Y}{X}\right)} - \overline{\left(\frac{1}{X}\right)} \cdot \bar{Y}}{\overline{\left(\frac{1}{X^2}\right)} - \overline{\left(\frac{1}{X}\right)}^2}$<br>$a = \bar{Y} - b \cdot \overline{\left(\frac{1}{X}\right)}$ | $\overline{\left(\frac{1}{X}\right)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$,<br>$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ ,<br>$\overline{\left(\frac{1}{X^2}\right)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i^2}$, $\overline{\left(\frac{Y}{X}\right)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\cdot y_i$ |

Показатели эффективности для парной регрессии находятся по следующим формулам.

*Коэффициент корреляции*
$$r_{YX} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{cov(X,Y)}{(\sigma_X)^2} \cdot \frac{\sigma_X}{\sigma_Y} = b\frac{\sigma_X}{\sigma_Y}$$

*Коэффициент детерминации*
$$r_{YX}^2 = (r_{YX})^2$$

*Скорректированный индекс корреляции*
$$\tilde{r}_{YX} = \sqrt{max\left(0; 1 - (1 - r_{YX}^2) \cdot \frac{n-1}{n-2}\right)}$$

*Скорректированный индекс детерминации*
$$\tilde{r}^2_{XY} = \max\left(0; 1-(1-r^2_{YX}) \cdot \frac{n-1}{n-2}\right)$$

*Значения t-критерия Стьюдента для коэффициента корреляции* $r_{YX}$
$$t_{\text{расч}} = \frac{|r_{YX}|}{\sqrt{1 - r_{YX}^2}}\sqrt{n-2}$$

Показатели эффективности или параметры уравнения считаются значимыми, если расчетные значения превышают табличные с заданным уровнем значимости α (ошибка первого рода).

*Пример 5.1.* Найти уравнение парной линейной регрессии, выражающей зависимость затрат на производство от выпуска продукции и проверить тесноту связи между этими показателями с помощью коэффициентов корреляции $r_{xy}$ и детерминации $R^2 = (r_{xy})^2$

| $x_i$ | 1 | 2 | 4 | 3 | 5 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| $y_i$ | 30 | 70 | 150 | 100 | 170 | 100 | 150 |

| № предприятия | Выпуск продукции тыс. ед. $x_i$ | Затраты на производство млн.руб. $y_i$ | $x_i y_i$ | $x_i^2$ | $y_i^2$ | Теоретические значения y $\hat{y} = a + bx$ | Остатки |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 30 | 30 | 1 | 900 | 31,1 | -1,1 |
| 2 | 2 | 70 | 140 | 4 | 4900 | 67,9 | 2,1 |
| 3 | 4 | 150 | 600 | 16 | 22500 | 141,6 | 8,4 |
| 4 | 3 | 100 | 300 | 9 | 10000 | 104,7 | -4,7 |
| 5 | 5 | 170 | 850 | 25 | 28900 | 178,4 | -8,4 |
| 6 | 3 | 100 | 300 | 9 | 10000 | 104,7 | -4,7 |
| 7 | 4 | 150 | 600 | 16 | 22500 | 141,6 | 8,4 |
| **Итого** | **22** | **770** | **2820** | **80** | **99700** | **770** | **0** |
| **Среднее** | $\bar{x} = \frac{\sum x_i}{n} \approx \approx 3,14$ | $\bar{y} = \frac{\sum y_i}{n} = = 110$ | $\overline{xy} = \frac{\sum xy}{n} \approx \approx 402,86$ | $\overline{x^2} = \frac{\sum x^2}{n} \approx \approx 11,43$ | $\overline{y^2} = \frac{\sum y^2}{n} \approx \approx 14243$ | $\bar{y} = \frac{\sum y_i}{n} \approx \approx 110$ | **0** |

$$\hat{y} = a + bx$$

$$b = \frac{cov(x,y)}{\sigma_x^2} = \frac{\overline{yx} - \overline{yx}}{\overline{x^2} - (\overline{x})^2} \approx \frac{402,8 - 3,14 \cdot 110}{11,43 - 3,14^2} \approx \frac{57,7}{1,5704} \approx 36.84$$

$$a = \overline{y} - b\overline{x} = 110 - 36,84 \cdot 3,14 = 5,79$$

$$\hat{y} = a + bx = -5,79 + 36,84x$$

$$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{57,7}{\sqrt{11,43 - 3,14^2} \cdot \sqrt{14243 - 110^2}} = \frac{57,7}{1,25316 \cdot 1,4639} = 31,45269$$

$$R^2 = \left(r_{xy}\right)^2 = 0,98$$

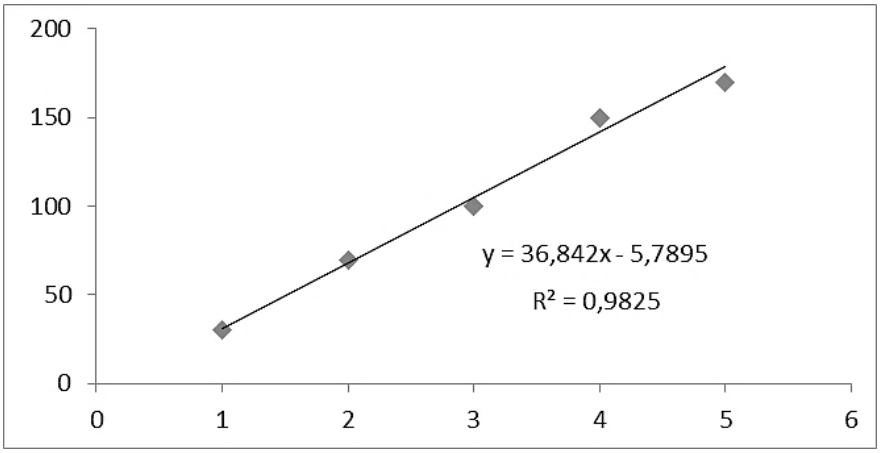



Рис. 5.1

*Пример 5.2.* Найти наилучшее уравнение регрессии.

Решение. Построить графики, найти уравнения и значения величины достоверности аппроксимации ($R^2$).

1) линейной регрессии; 2) экспоненциальная; 3) логарифмическая; 4) полиномиальная (степени 2); 5) степенная

**Решение с помощью табличного процессора.**

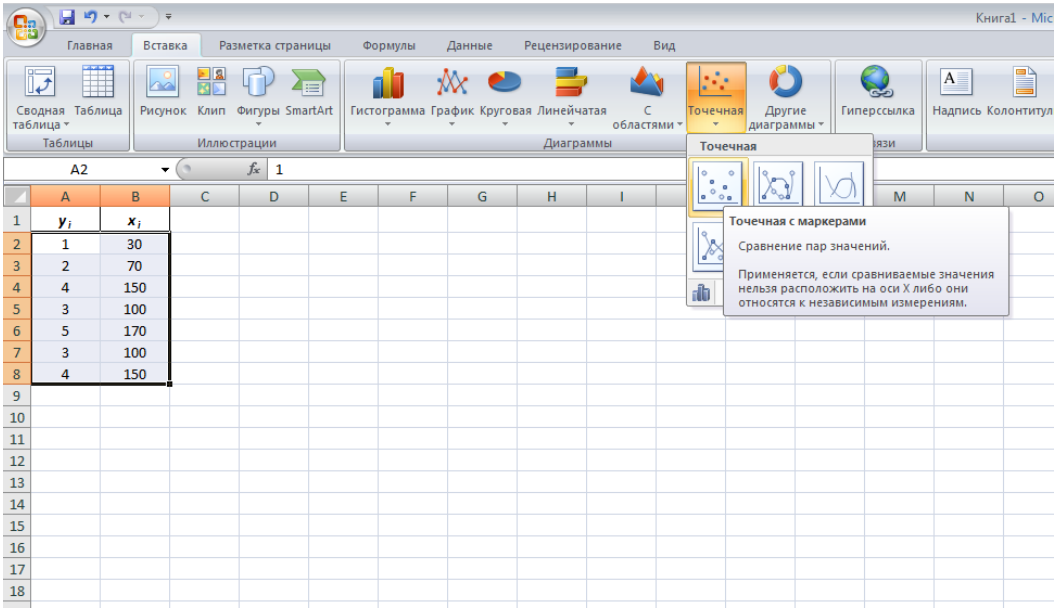



Рис. 5.2

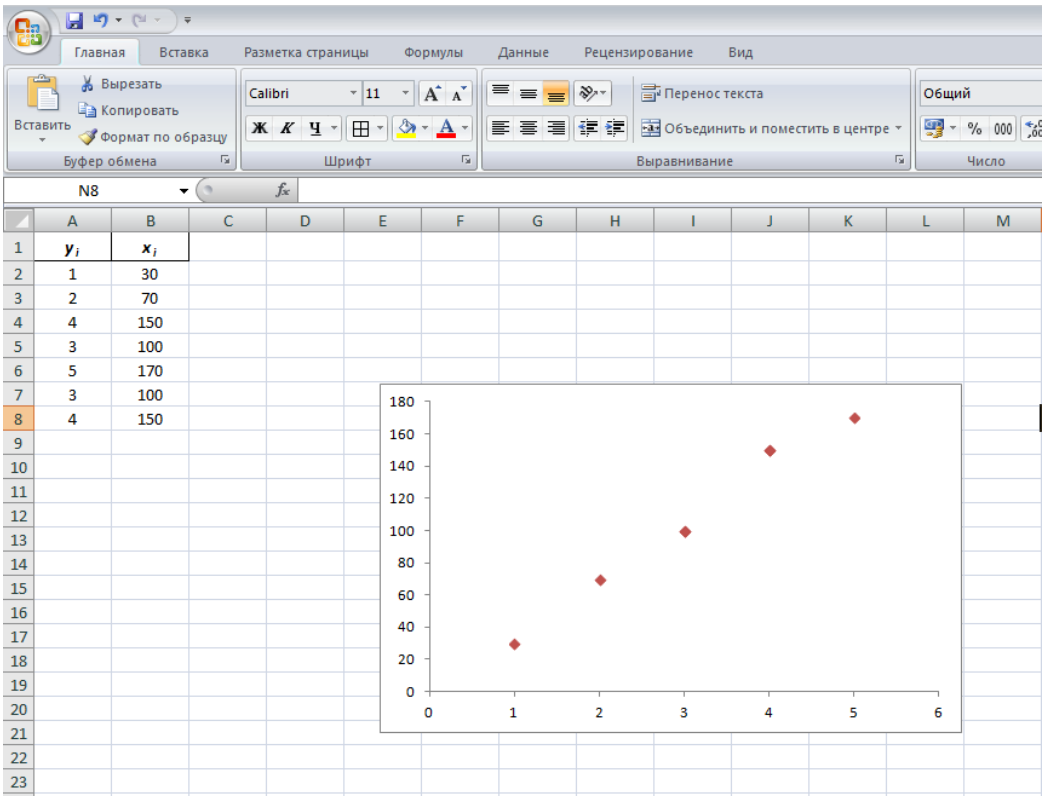
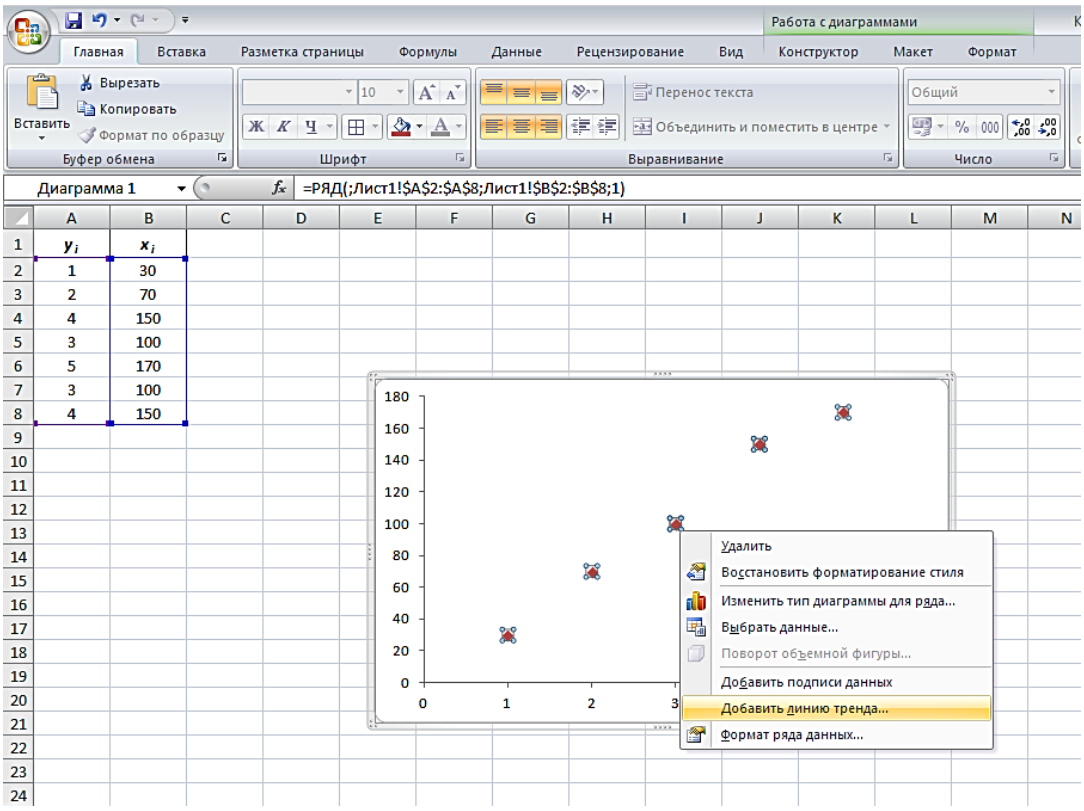



Рис. 5.3




| $y_i$ | $x_i$ |
|---|---|
| 1 | 30 |
| 2 | 70 |
| 4 | 150 |
| 3 | 100 |
| 5 | 170 |
| 3 | 100 |
| 4 | 150 |

Рис. 5.4

1) Линейной регрессии ( $\hat{y} = a + bx$ ):

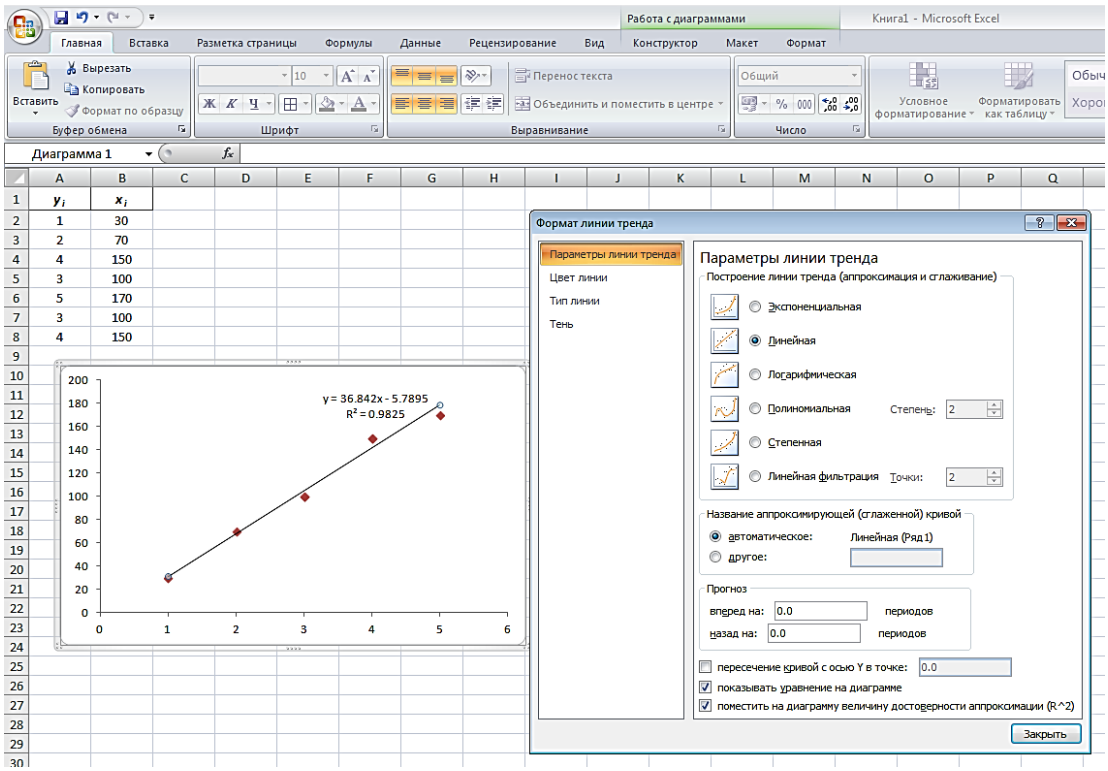



Рис. 5.5

2) Экспоненциальная ( $\hat{y} = a \cdot e^{bx}$ ):

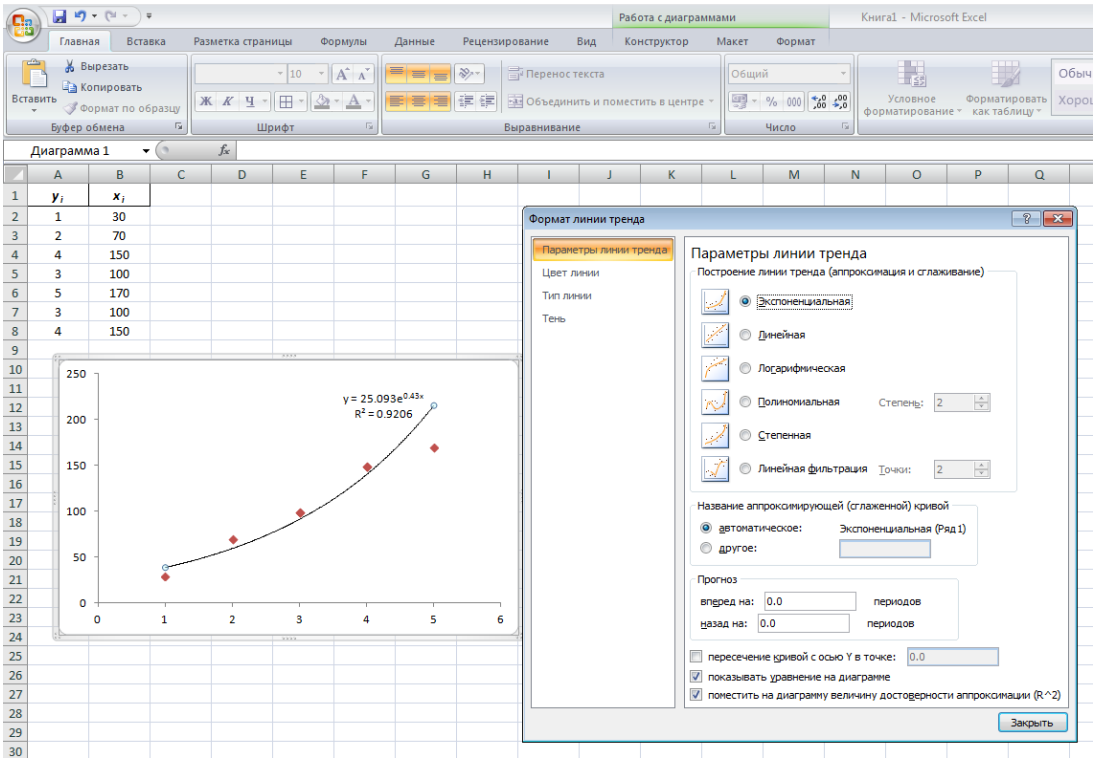



Рис. 5.6

3) логарифмическая ( $\hat{y} = a + b \ln x$ ):

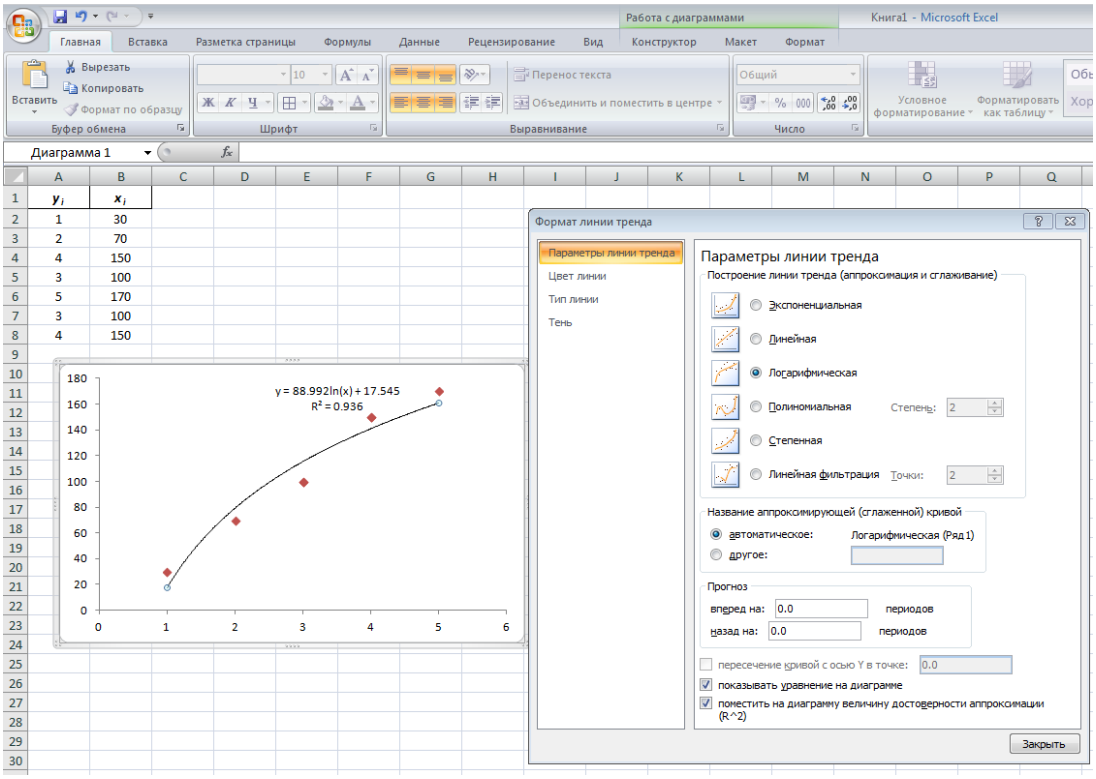



Рис. 5.7

4) полиномиальная степени 2 ( $\hat{y} = a + bx + cx^2$ ):

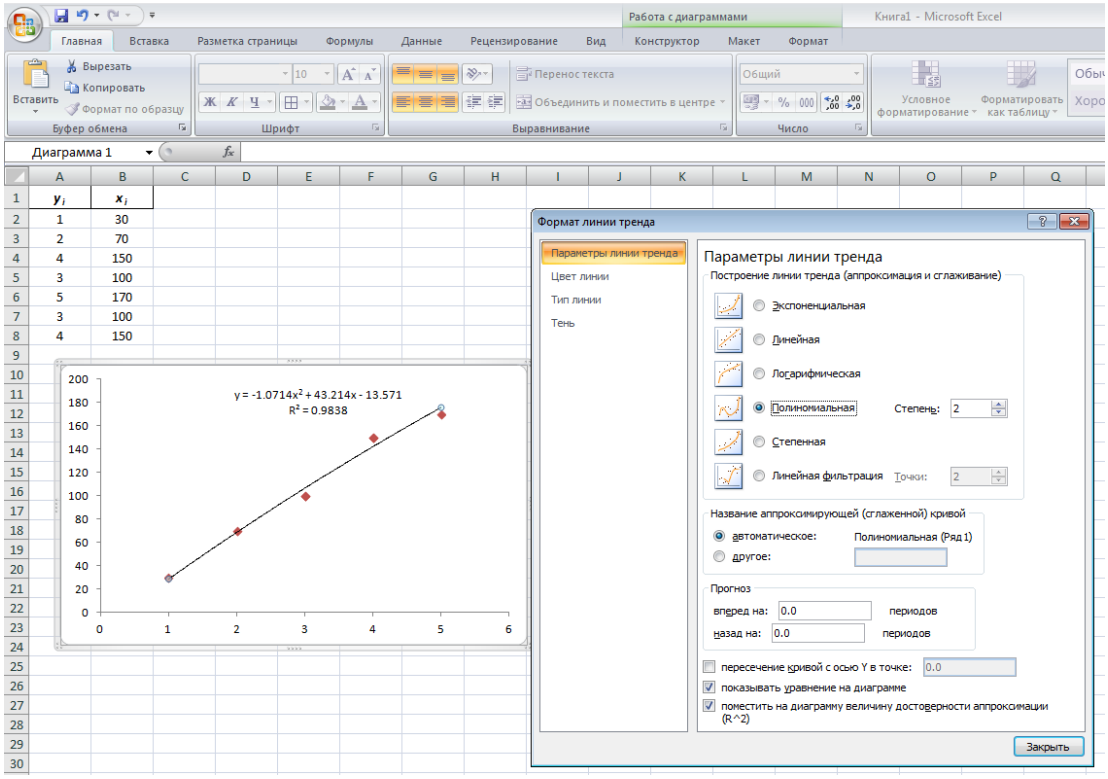



Рис. 5.8

5) степенная ( $\hat{y} = ax^b$ ):

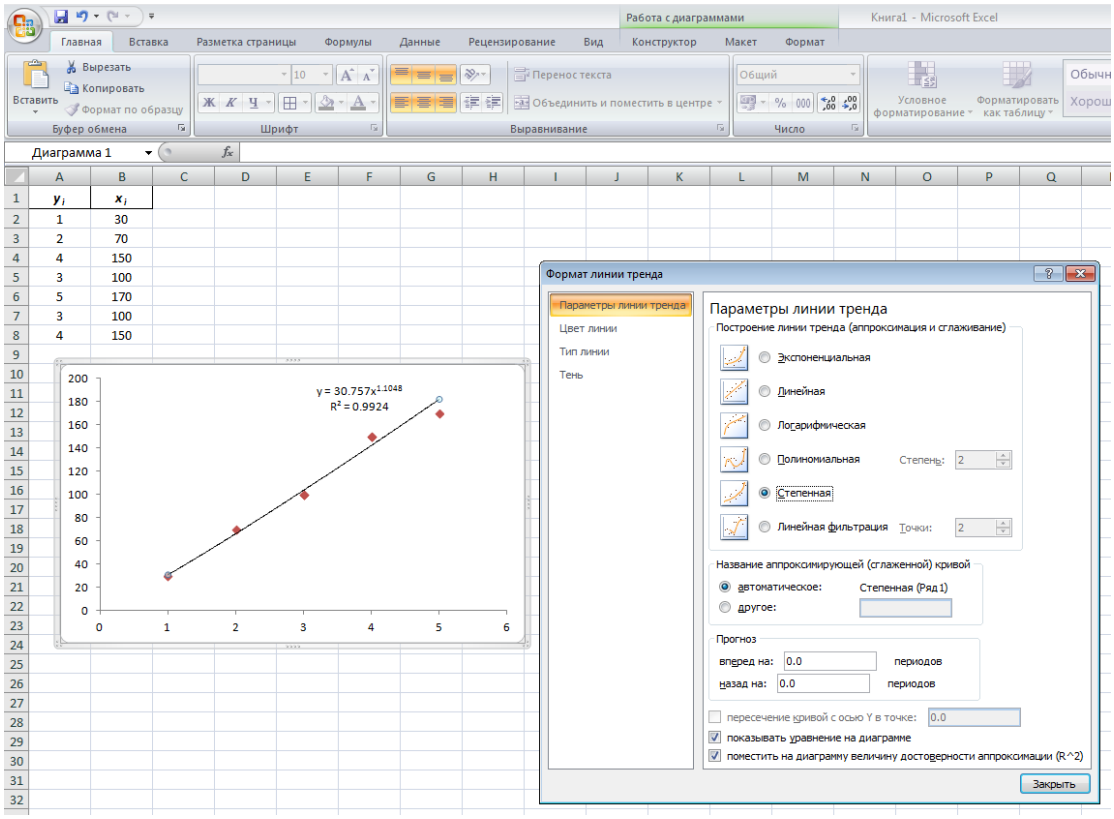



Рис. 5.9

## Выбор уравнения регрессии

На основании полученных результатов можно сделать вывод, что наиболее подходящими для описания взаимосвязи между зависимой переменной *y* и независимой переменной *x* являются **степенная** функция, поскольку эта функция имеет наиболее близкое к единице значение показателя $\mathbf{R^2}$.

| Название | Уравнение | $R^2$ |
|---|---|---|
| линейной | $\hat{y} = 36{,}842x - 5{,}7895$ | 0,9825 |
| экспоненциальная | $\hat{y} = 25{,}093e^{0{,}43x}$ | 0,9206 |
| логарифмическая | $\hat{y} = 88{,}992\ln x - 17{,}545$ | 0,9360 |
| полиномиальная | $\hat{y} = -1{,}0714x^2 + 43{,}214x - 13{,}571$ | 0,9838 |
| степенная | $\hat{y} = 30{,}757x^{1{,}1048}$ | **0,9924** |

## Задание для самостоятельного выполнения

**Задание.** Найти наилучшее уравнение парной регрессии.

Вариант 1

| *X* | *Y* |
|---|---|
| 21 | 20 |
| 33 | 22 |
| 36 | 15 |
| 45 | 55 |
| 54 | 39 |
| 60 | 60 |
| 69 | 30 |
| 75 | 60 |
| 84 | 45 |
| 90 | 70 |
| 99 | 80 |
| 105 | 90 |
| 114 | 79 |
| 120 | 100 |
| 129 | 189 |
| 135 | 200 |
| 144 | 199 |
| 150 | 204 |
| 159 | 220 |
| 165 | 214 |

Вариант 2

| *X* | *Y* |
|---|---|
| 24 | 13 |
| 35 | 12 |
| 13 | 14 |
| 43 | 27 |
| 38 | 27 |
| 41 | 17 |
| 45 | 30 |
| 48 | 30 |
| 52 | 45 |
| 55 | 32 |
| 59 | 31 |
| 62 | 30 |
| 66 | 63 |
| 69 | 50 |
| 73 | 65 |
| 76 | 76 |
| 80 | 81 |
| 83 | 85 |
| 87 | 90 |
| 90 | 94 |

Вариант 3

| *X* | *Y* |
|---|---|
| 115 | 20 |
| 105 | 22 |
| 103 | 60 |
| 96 | 80 |
| 90 | 103 |
| 84 | 70 |
| 78 | 120 |
| 72 | 120 |
| 66 | 100 |
| 60 | 105 |
| 54 | 130 |
| 48 | 110 |
| 42 | 135 |
| 36 | 127 |
| 30 | 140 |
| 24 | 129 |
| 18 | 150 |
| 12 | 148 |
| 6 | 152 |
| 1 | 155 |

Вариант 4

| *X* | *Y* |
|---|---|
| 6 | 40 |
| 12 | 45 |
| 21 | 50 |
| 27 | 70 |
| 36 | 80 |
| 42 | 78 |
| 51 | 71 |
| 57 | 90 |
| 66 | 95 |
| 72 | 90 |
| 81 | 90 |
| 87 | 89 |
| 96 | 95 |
| 102 | 100 |
| 111 | 110 |
| 117 | 129 |
| 126 | 150 |
| 132 | 148 |
| 141 | 139 |
| 147 | 140 |

Вариант 5

| *X* | *Y* |
|---|---|
| 20 | 15 |
| 25 | 7 |
| 35 | 15 |
| 40 | 10 |
| 50 | 10 |
| 55 | 15 |
| 65 | 15 |
| 70 | 40 |
| 80 | 50 |
| 85 | 50 |
| 95 | 50 |
| 100 | 90 |
| 110 | 88 |
| 115 | 90 |
| 125 | 120 |
| 130 | 150 |
| 140 | 130 |
| 145 | 150 |
| 155 | 177 |
| 160 | 220 |

Вариант 6

| *X* | *Y* |
|---|---|
| 156 | 40 |
| 130 | 45 |
| 138 | 41 |
| 124 | 54 |
| 114 | 50 |
| 106 | 50 |
| 96 | 81 |
| 88 | 90 |
| 78 | 99 |
| 70 | 79 |
| 60 | 117 |
| 52 | 130 |
| 42 | 128 |
| 34 | 165 |
| 24 | 153 |
| 16 | 170 |
| 6 | 200 |
| 2 | 250 |
| 4 | 220 |
| 14 | 198 |

Вариант 7

| *X* | *Y* |
|---|---|
| 13 | 23 |
| 39 | 15 |
| 52 | 9 |
| 78 | 8 |
| 104 | 7 |
| 117 | 6 |
| 143 | 5 |
| 156 | 4 |
| 182 | 3 |
| 208 | 2 |
| 234 | 1 |
| 247 | 1 |
| 273 | 1 |
| 299 | 1 |
| 312 | 1 |
| 338 | 1 |
| 364 | 1 |
| 390 | 1 |
| 403 | 1 |
| 429 | 1 |

Вариант 8

| *X* | *Y* |
|---|---|
| 700 | 250 |
| 460 | 220 |
| 740 | 180 |
| 680 | 100 |
| 700 | 65 |
| 720 | 50 |
| 740 | 45 |
| 760 | 40 |
| 780 | 35 |
| 800 | 30 |
| 820 | 25 |
| 840 | 20 |
| 860 | 10 |
| 880 | 9 |
| 900 | 8 |
| 920 | 7 |
| 940 | 6 |
| 960 | 5 |
| 980 | 4 |
| 1000 | 3 |

Вариант 9

| *X* | *Y* |
|---|---|
| 130 | 15 |
| 127 | 29 |
| 120 | 15 |
| 116 | 50 |
| 111 | 70 |
| 106 | 40 |
| 101 | 80 |
| 96 | 90 |
| 91 | 50 |
| 86 | 88 |
| 81 | 110 |
| 76 | 120 |
| 71 | 110 |
| 66 | 110 |
| 61 | 120 |
| 56 | 190 |
| 51 | 130 |
| 46 | 150 |
| 41 | 177 |
| 36 | 186 |

Вариант 10

| *X* | *Y* |
|---|---|
| 42 | 114 |
| 66 | 120 |
| 72 | 104 |
| 90 | 99 |
| 108 | 100 |
| 120 | 50 |
| 138 | 55 |
| 150 | 60 |
| 168 | 40 |
| 180 | 35 |
| 198 | 45 |
| 210 | 40 |
| 228 | 35 |
| 240 | 25 |
| 258 | 15 |
| 270 | 10 |
| 288 | 15 |
| 300 | 33 |
| 318 | 29 |
| 330 | 22 |

## Контрольные вопросы

1. Что такое регрессия?
2. Что такое линейная регрессия?
3. Что такое нелинейная регрессия?
4. Какие существуют виды нелинейной регрессии?
5. Что такое парная регрессия?
6. Что такое множественная регрессия?

## Лабораторная работа № 6
## Множественная регрессия

**Цель лабораторной работы:** изучить основные понятия, связанные с множественной регрессией, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

Задачей построения модели множественной регрессии является установление функциональной зависимости между одной зависимой переменной $Y$, называемой результатом, и несколькими независимыми $X_1, X_2, \ldots, X_m$, называемыми факторами в интересах прогноза значений результата, а также оценки влияния отдельных факторов.

Связь между переменными проявляется как некоторая закономерность лишь в среднем на основе выборки совместных наблюдений, т.е. наборов значений зависимого и независимых показателей, относящихся к одному объекту:

$$\begin{gathered}\left(y^1, x_1^1, x_2^1, \ldots, x_m^1\right), \\ \left(y^2, x_1^2, x_2^2, \ldots, x_m^2\right), \\ \ldots, \\ \left(y^n, x_1^n, x_2^n, \ldots, x_m^n\right).\end{gathered}$$

Здесь и везде далее использованы обозначения

$n$ – количество наблюдений,

$m$ – количество факторов.

Задача состоит в том, чтобы по данным наблюдениям подобрать аналитически описываемую функцию $\hat{Y} = f(X_1, X_2, \ldots, X_m)$, которая в некотором смысле наилучшим образом описывает результаты наблюдений.

Для каждого $i$-го наблюдения можно записать:

$$y^i = \hat{y}^i + \varepsilon^i,$$

где $y^i$ – фактическое значение результата,

$\hat{y}^i$ – теоретическое значение результата,

$\varepsilon^i$ – остаточная величина, показывающая отклонение фактического значения результата от теоретического.

Таким образом, в общем виде уравнение связи этих показателей имеет вид

$$Y = f(X_1, X_2, \ldots, X_m) + \varepsilon = \hat{Y} + E,$$

где $Y = (y^1, y^2, \ldots, y^n)$,

$\hat{Y} = (\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^n)$,

$E = (\varepsilon^1, \varepsilon^2, \ldots, \varepsilon^n)$.

Точность уравнения тем выше, чем меньше значения E для всей совокупности наблюдений. На их точность влияют следующие факторы:

– неправильный выбор функции $f$;

– нерепрезентативность выборки наблюдений;

– неточность измерения значений переменных $X_1, X_2, \ldots, X_m$ и $Y$.

Частный случай парной регрессии, когда $m = 1$, был рассмотрен ранее.

*Определение.* Если $\hat{Y} = a + b_1 X_1 + b_2 X_2 + \ldots + b_m X_m$, то регрессия называется **линейной**, а иначе – **нелинейной**.

Нелинейные уравнения, как и в случае парной регрессии, разделяются на два класса. К первому классу относятся уравнения, линеаризуемые с помощью замены переменных.

Среди таких уравнений обычно рассматривают:

1) гиперболическое – $\hat{Y} = a + \frac{b_1}{X_1} + \frac{b_2}{X_2} + \ldots + \frac{b_m}{X_m}$;

2) полулогарифмическое – $\hat{Y} = a + b_1 \ln X_1 + b_2 \ln X_2 + \ldots + b_m \ln X_m$;

3) квадратичное – $\hat{Y} = a + b_1 (X_1)^2 + b_2 (X_2)^2 + \ldots + b_m (X_m)^2$;

4) кубическое – $\hat{Y} = a + b_1 (X_1)^3 + b_2 (X_2)^3 + \ldots + b_m (X_m)^3$.

Ко второму классу относят уравнения, которые линеаризуются с помощью нелинейных функциональных преобразований. Среди них наиболее часто рассматривают

1) показательное – $\hat{Y} = a (b_1)^{X_1} (b_2)^{X_2} \ldots (b_m)^{X_m}$;

2) степенное – $\hat{Y} = a (X_1)^{b_1} (X_2)^{b_2} \ldots (X_m)^{b_m}$.

Наиболее важным является линейное уравнение множественной регрессии. Его параметры могут быть найдены с использованием метода наименьших квадратов.

Параметры $b_1, b_2, \ldots, b_m$ в уравнении называются **коэффициентами чистой регрессии** и имеют следующий содержательный смысл:

коэффициент регрессии $b_i$ показывает, на сколько единиц в среднем изменится результат $Y$, если фактор $X_i$ изменится на одну единицу при неизменных значениях остальных факторов.

Для численной оценки вероятностных зависимостей случайных величин в таком случае используются

– индекс корреляции,

– индекс детерминации,

– скорректированный индекс корреляции,

– скорректированный индекс детерминации,

– средняя ошибка аппроксимации.

Обратимся к изучению этих оценок.

На основе этих наблюдений, как показано выше, общая дисперсия случайной величины $Y$ определяется по формуле:

$$\sigma^2_{Y\ общ} = \frac{\sum_{i=1}^{n}(y_i - \overline{Y})^2}{n}.$$

Определим также дисперсию, объясненную дисперсией

$$\sigma^2_{Y\ регр} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{Y})^2}{n},$$

и остаточную дисперсию

$$\sigma^2_{Y\ ост} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}.$$

Можно доказать, что

$$\sigma^2_{Y\ общ} = \sigma^2_{Y\ регр} + \sigma^2_{Y\ ост}.$$

**Индекс корреляции** определяется по формуле:

$$r_{YX_1X_2\ldots X_m} = r = \sqrt{1 - \frac{\sigma^2_{Y\ ост}}{\sigma^2_{Y\ общ}}} = \sqrt{\frac{\sigma^2_{Y\ регр}}{\sigma^2_{Y\ общ}}},$$

**индекс детерминации** –

$$r^2_{YX_1X_2\ldots X_m} = r^2 = 1 - \frac{\sigma^2_{Y\ ост}}{\sigma^2_{Y\ общ}} = \frac{\sigma^2_{Y\ регр}}{\sigma^2_{Y\ общ}}.$$

*Пример 6.1.* С помощью инструмента анализа данных *«Регрессия»* найти уравнение множественной регрессии.

**Решение с помощью табличного процессора.**

1) В главном меню последовательно выбрать *«Сервис» → «Анализ данных» → «Регрессия» → «ОК».*

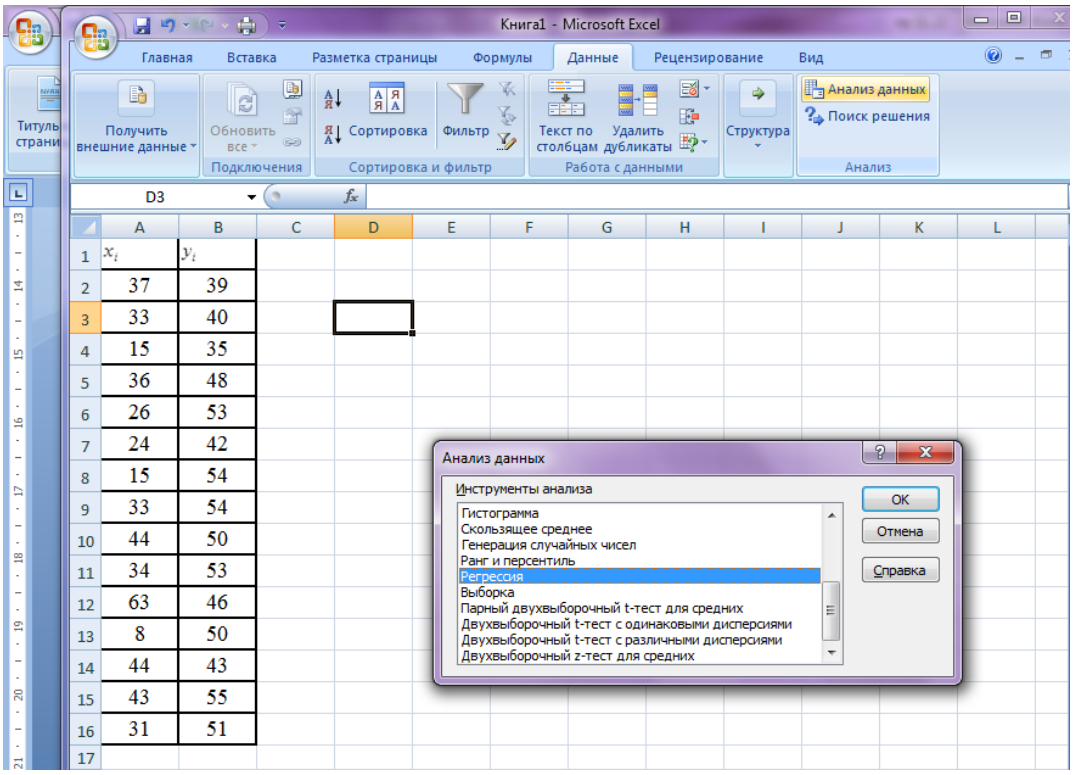



Рис. 6.1

2) после вызова режима *«Регрессия»* на экране появляется диалоговое окно, в котором задаются следующие параметры:

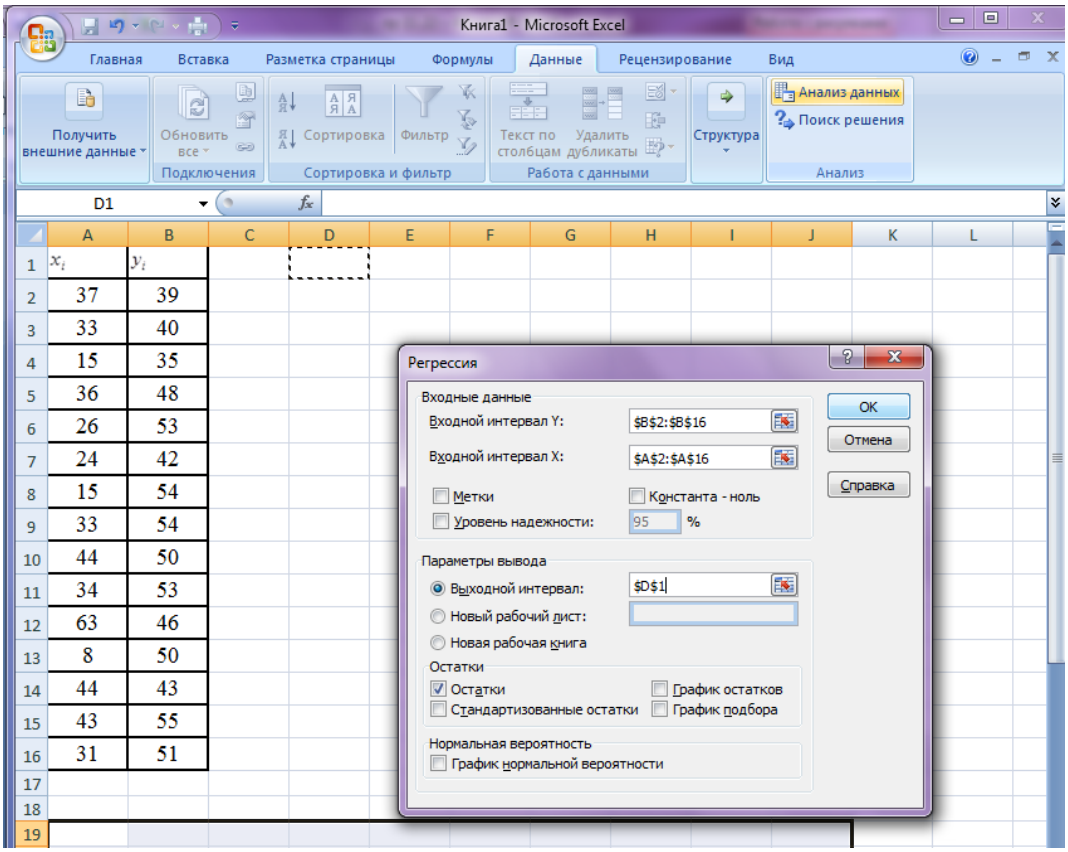



Рис. 6.2

➢Указание.

***Входной интервал Y*** – диапазон (столбец), содержащий данные со значениями зависимой переменной;

***Входной интервал X*** – диапазон (столбцы), содержащий данные со значениями независимых переменных;

***Метки*** – флажок, который указывает, содержат ли первые элементы отмеченных диапазонов названия переменных (столбцов) или нет;

***Константа ноль*** – флажок указывает на наличие или отсутствие свободного члена в уравнении ($\beta_0$);

***Выходной интервал*** – достаточно указать левую верхнюю ячейку будущего диапазона, в котором будет сохранен отчет построения модели;

***Новый рабочий лист*** – можно задать произвольное имя нового листа, в котором будет сохранен отчет.

***Остатки*** – вывод теоретических значений зависимой переменной и остатков, т. е. разности между фактическими и теоретическими значениями зависимой переменной.

Если необходимо получить значения и графики остатков, установите соответствующие флажки в диалоговом окне. Нажмите на кнопку **ОК**.

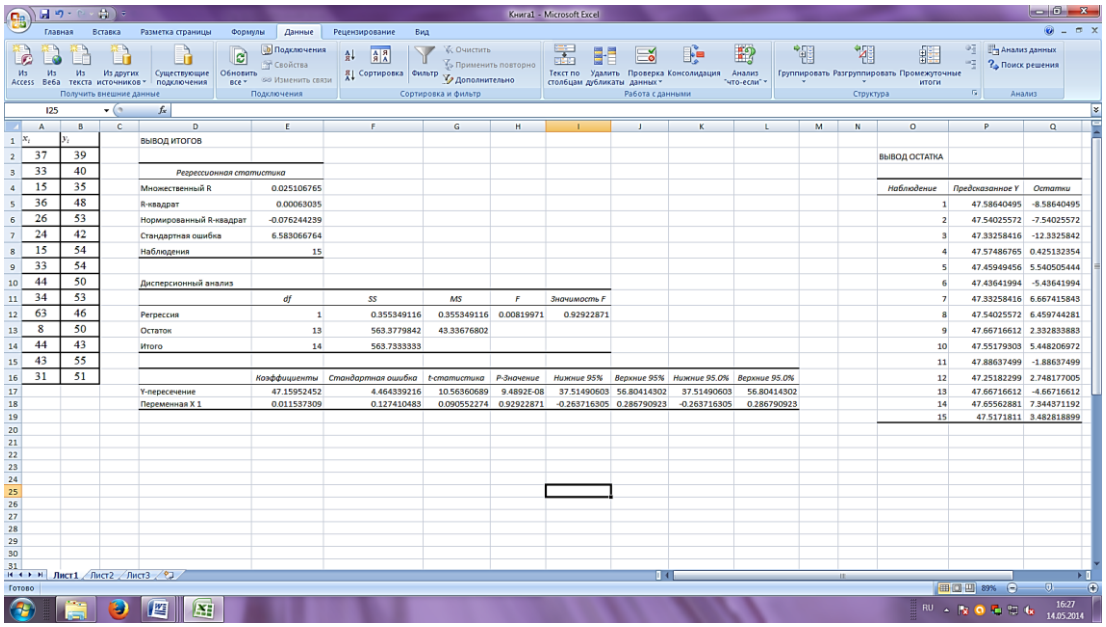



Рис. 6.3

➢Указание.

**Множественный R** – модуль коэффициента корреляции, характеризует тесноту связи.

**Значимость F –** позволяет проверить значимость уравнения регрессии (если < 0,05, то уравнение значимо).

**P-значение -** позволяет проверить значимость каждой переменной (если < 0,05, то переменная значима).

**Коэффициенты** (коэффициенты уравнения регрессии):

$\hat{y}_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + ... + a_m x_{im}$, $i = 1, n$.

Коэффициент ***Y***-пересечения: $a_0$

Коэффициент $x_1$: $a_1$

Коэффициент $x_2$: $a_2$ и т.д.

**Нижние 95%:** нижняя граница доверительного интервала переменной с надежностью 95%.

**Верхнее 95%:** верхняя граница доверительного интервала переменной с надежностью 95%.

**Предсказанное y:** теоретическое значение $\hat{y}$

**Остатки** $\varepsilon = y - \hat{y}$

*Пример 6.2.* Определить зависимость количества нарушений от количества осужденных, количества повторно осужденных и количества осужденных, моложе 30 лет.

**Решение с помощью табличного процессора.**

| № | Количество нарушений | Количество осужденных | Количество повторно осужденных | Количество лиц, моложе 30 лет |
|---|---|---|---|---|
| | Y | X1 | X2 | X3 |
| 1. | 54 | 5000 | 108 | 310 |
| 2. | 65 | 3120 | 131 | 380 |
| 3. | 85 | 4050 | 150 | 500 |
| 4. | 68 | 3260 | 137 | 400 |
| 5. | 95 | 4550 | 189 | 550 |
| 6. | 53 | 2540 | 105 | 310 |
| 7. | 67 | 3210 | 133 | 390 |
| 8. | 61 | 2900 | 121 | 350 |
| 9. | 59 | 2830 | 153 | 340 |
| 10. | 74 | 3550 | 137 | 430 |
| 11. | 58 | 2780 | 115 | 340 |
| 12. | 43 | 2100 | 115 | 250 |
| 13. | 76 | 3500 | 147 | 440 |
| 14. | 77 | 3650 | 155 | 470 |
| 15. | 72 | 3470 | 147 | 420 |
| 16. | 81 | 3780 | 175 | 540 |
| 17. | 56 | 2550 | 173 | 320 |
| 18. | 74 | 3450 | 133 | 430 |
| 19. | 50 | 2410 | 134 | 210 |
| 20. | 65 | 3120 | 125 | 380 |
| 21. | 55 | 2640 | 114 | 320 |
| 22. | 45 | 2150 | 120 | 260 |
| 23. | 85 | 4060 | 170 | 450 |
| 24. | 97 | 4300 | 190 | 220 |

Рис. 6.4

*Выполнение задания*

Предполагается, что нет тесной связи между факторами.

Построим уравнение множественной регрессии вида $\hat{Y} = a + \sum_{i=1}^{n} b_i \cdot X_i^2$.

Возведем в квадрат значения всех факторов.

| G | H | I | J |
|---|---|---|---|
| **Y** | **X1^2** | **X2^2** | **X3^2** |
| 54 | 25000000 | 11664 | 96100 |
| 65 | 9734400 | 17161 | 144400 |
| 85 | 16402500 | 22500 | 250000 |
| 68 | 10627600 | 18769 | 160000 |
| 95 | 20702500 | 35721 | 302500 |
| 53 | 6451600 | 11025 | 96100 |
| 67 | 10304100 | 17689 | 152100 |
| 61 | 8410000 | 14641 | 122500 |
| 59 | 8008900 | 23409 | 115600 |
| 74 | 12602500 | 18769 | 184900 |
| 58 | 7728400 | 13225 | 115600 |
| 43 | 4410000 | 13225 | 62500 |
| 76 | 12250000 | 21609 | 193600 |
| 77 | 13322500 | 24025 | 220900 |
| 72 | 12040900 | 21609 | 176400 |
| 81 | 14288400 | 30625 | 291600 |
| 56 | 6502500 | 29929 | 102400 |
| 74 | 11902500 | 17689 | 184900 |
| 50 | 5808100 | 17956 | 44100 |
| 65 | 9734400 | 15625 | 144400 |
| 55 | 6969600 | 12996 | 102400 |
| 45 | 4622500 | 14400 | 67600 |
| 85 | 16483600 | 28900 | 202500 |
| 97 | 18490000 | 36100 | 48400 |

Рис. 6.5

Проведем регрессионный анализ с полученными значениями.

ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,900301376 |
| R-квадрат | 0,810542567 |
| Нормированный R-квадрат | 0,782123952 |
| Стандартная ошибка | 6,881459918 |
| Наблюдения | 24 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 3 | 4051,868521 | 1350,62284 | 28,52153667 | 2,00571E-07 |
| Остаток | 20 | 947,0898121 | 47,3544906 | | |
| Итого | 23 | 4998,958333 | | | |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | 27,67787563 | 4,550566374 | 6,082292478 | 6,04922E-06 | 18,18556051 | 37,17019075 | 18,18556051 | 37,17019075 |
| X1^2 | 9,05098E-07 | 3,33628E-07 | 2,712893613 | 0,013394406 | 2,09162E-07 | 1,60103E-06 | 2,09162E-07 | 1,60103E-06 |
| X2^2 | 0,000963272 | 0,000234063 | 4,115439685 | 0,000537077 | 0,000475025 | 0,001451518 | 0,000475025 | 0,001451518 |
| X3^2 | 6,49257E-05 | 2,44832E-05 | 2,65184385 | 0,015304809 | 1,38546E-05 | 0,000115997 | 1,38546E-05 | 0,000115997 |

Рис. 6.6

Анализ результатов показал, что все эти факторы являются значимыми. Уравнение нелинейной множественной регрессии примет вид

$\hat{Y} = 27,67787563 + (9,05098E\text{-}07)X_1^2 + 0,000963272X_2^2 + (6,49257E\text{-}05)X_3^2$,

$R^2 = 0,810542567$.

Аналогичным образом построим остальные уравнения нелинейной множественной регрессии.

Построим $\hat{Y} = a + \sum_{i=1}^{n} b_i \cdot X_i^3$.

| Y | X1^3 | X2^3 | X3^3 |
|---|---|---|---|
| 54 | 125000000000 | 1259712 | 29791000 |
| 65 | 30371328000 | 2248091 | 54872000 |
| 85 | 66430125000 | 3375000 | 125000000 |
| 68 | 34645976000 | 2571353 | 64000000 |
| 95 | 94196375000 | 6751269 | 166375000 |
| 53 | 16387064000 | 1157625 | 29791000 |
| 67 | 33076161000 | 2352637 | 59319000 |
| 61 | 24389000000 | 1771561 | 42875000 |
| 59 | 22665187000 | 3581577 | 39304000 |
| 74 | 44738875000 | 2571353 | 79507000 |
| 58 | 21484952000 | 1520875 | 39304000 |
| 43 | 9261000000 | 1520875 | 15625000 |
| 76 | 42875000000 | 3176523 | 85184000 |
| 77 | 48627125000 | 3723875 | 103823000 |
| 72 | 41781923000 | 3176523 | 74088000 |
| 81 | 54010152000 | 5359375 | 157464000 |
| 56 | 16581375000 | 5177717 | 32768000 |
| 74 | 41063625000 | 2352637 | 79507000 |
| 50 | 13997521000 | 2406104 | 9261000 |
| 65 | 30371328000 | 1953125 | 54872000 |
| 55 | 18399744000 | 1481544 | 32768000 |
| 45 | 9938375000 | 1728000 | 17576000 |
| 85 | 66923416000 | 4913000 | 91125000 |
| 97 | 79507000000 | 6859000 | 10648000 |

Рис. 6.7

Проведем регрессионный анализ с полученными значениями.

ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,880478617 |
| R-квадрат | 0,775242595 |
| Нормированный R-квадрат | 0,741528985 |
| Стандартная ошибка | 7,495174783 |
| Наблюдения | 24 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 3 | 3875,405433 | 1291,801811 | 22,99494417 | 1,08567E-06 |
| Остаток | 20 | 1123,552901 | 56,17764503 | | |
| Итого | 23 | 4998,958333 | | | |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | 41,5063599 | 3,485516624 | 11,90823754 | 1,55688E-10 | 34,23569963 | 48,77702017 | 34,23569963 | 48,77702017 |
| X1^3 | 1,2834E-10 | 6,41967E-11 | 1,999168994 | 0,059362147 | -5,57192E-12 | 2,62252E-10 | -5,57192E-12 | 2,62252E-10 |
| X2^3 | 4,22509E-06 | 1,13757E-06 | 3,714144186 | 0,001371291 | 1,85216E-06 | 6,59801E-06 | 1,85216E-06 | 6,59801E-06 |
| X3^3 | 1,22972E-07 | 4,37947E-08 | 2,80792686 | 0,010864865 | 3,16182E-08 | 2,14327E-07 | 3,16182E-08 | 2,14327E-07 |

Рис. 6.8

Анализ показал, что коэффициент при X1^3 не значим. Исключим данный параметр и проведем повторно регрессионный анализ.

ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,854592571 |
| R-квадрат | 0,730328462 |
| Нормированный R-квадрат | 0,704645458 |
| Стандартная ошибка | 8,012124098 |
| Наблюдения | 24 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 2 | 3650,88155 | 1825,440775 | 28,43625581 | 1,05625E-06 |
| Остаток | 21 | 1348,076784 | 64,19413255 | | |
| Итого | 23 | 4998,958333 | | | |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | 43,18888794 | 3,615668806 | 11,94492368 | 7,92599E-11 | 35,66969304 | 50,70808285 | 35,66969304 | 50,70808285 |
| X2^3 | 4,88871E-06 | 1,1631E-06 | 4,20317456 | 0,000399721 | 2,46991E-06 | 7,30751E-06 | 2,46991E-06 | 7,30751E-06 |
| X3^3 | 1,48271E-07 | 4,48183E-08 | 3,308268621 | 0,003344617 | 5,50663E-08 | 2,41476E-07 | 5,50663E-08 | 2,41476E-07 |

Рис. 6.9

Уравнение нелинейной множественной регрессии имеет вид $\hat{Y} = 43{,}18888794 + (4{,}88871\text{E-}06)X_2^3 + (1{,}48271\text{E-}07)X_3^3$, $R^2 = 0{,}730328462$.

Построим $\hat{Y} = a + b_1 \cdot X_1^2 + b_2 \cdot X_2^3 + b_i \cdot \log_{10}(X_3)$.

ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,900596262 |
| R-квадрат | 0,811073628 |
| Нормированный R-квадрат | 0,782734672 |
| Стандартная ошибка | 6,871808576 |
| Наблюдения | 24 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 3 | 4054,523271 | 1351,507757 | 28,6204486 | 1,95077E-07 |
| Остаток | 20 | 944,4350621 | 47,2217531 | | |
| Итого | 23 | 4998,958333 | | | |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | -62,71019858 | 34,43610412 | -1,821059617 | 0,0835929 | -134,542653 | 9,122255886 | -134,542653 | 9,122255886 |
| X1^2 | 8,74616E-07 | 3,34513E-07 | 2,614595192 | 0,01659425 | 1,76834E-07 | 1,5724E-06 | 1,76834E-07 | 1,5724E-06 |
| X2^3 | 4,67473E-06 | 9,9254E-07 | 4,709862974 | 0,00013431 | 2,60433E-06 | 6,74513E-06 | 2,60433E-06 | 6,74513E-06 |
| log10(X3) | 41,33104695 | 13,99487188 | 2,953299416 | 0,00785869 | 12,13825576 | 70,52383814 | 12,13825576 | 70,52383814 |

Рис. 6.10

Уравнение нелинейной множественной регрессии имеет вид $\hat{Y} = -62{,}710198 + (8{,}74616\text{E-}07)X_1^2 + (4{,}67473\text{E-}06)X_2^3 + (41{,}33104695)\log_{10} X_3$, $R^2 = 0{,}811073628$.

Сравнительный анализ всех полученных уравнений показал, что преимуществом обладает последнее, обладающее наибольшим индексом детерминации.

## Задание для самостоятельного выполнения

Вариант 1

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 110 | 13000 | 154 | 473 | 324 |
| 134 | 12400 | 147 | 594 | 393 |
| 156 | 14600 | 175 | 682 | 450 |
| 172 | 16800 | 189 | 781 | 411 |
| 134 | 14400 | 175 | 671 | 567 |
| 180 | 16200 | 189 | 935 | 315 |
| 90 | 11000 | 133 | 385 | 399 |
| 134 | 14600 | 175 | 627 | 363 |
| 132 | 14000 | 168 | 605 | 459 |
| 86 | 9000 | 119 | 319 | 411 |
| 46 | 4200 | 70 | 132 | 345 |
| 46 | 5200 | 70 | 154 | 345 |
| 112 | 12800 | 154 | 605 | 441 |
| 180 | 17800 | 210 | 847 | 465 |
| 112 | 13200 | 154 | 495 | 441 |
| 134 | 13600 | 161 | 627 | 525 |
| 136 | 15600 | 189 | 605 | 519 |
| 64 | 8000 | 105 | 231 | 399 |
| 24 | 3400 | 49 | 121 | 402 |
| 68 | 5000 | 77 | 231 | 375 |
| 110 | 5280 | 63 | 594 | 342 |
| 90 | 4300 | 56 | 341 | 360 |
| 170 | 8120 | 105 | 847 | 510 |
| 194 | 8600 | 119 | 1012 | 570 |

Вариант 2

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 275 | 620 | 364 | 86 | 108 |
| 335 | 754 | 403 | 108 | 131 |
| 390 | 875 | 455 | 124 | 150 |
| 430 | 650 | 286 | 142 | 137 |
| 110 | 224 | 143 | 20 | 12 |
| 440 | 772 | 455 | 98 | 80 |
| 160 | 384 | 221 | 38 | 29 |
| 390 | 936 | 533 | 80 | 68 |
| 100 | 312 | 273 | 44 | 20 |
| 50 | 225 | 221 | 62 | 5 |
| 280 | 550 | 286 | 46 | 50 |
| 275 | 653 | 325 | 60 | 45 |
| 320 | 768 | 403 | 58 | 54 |
| 150 | 570 | 221 | 32 | 29 |
| 60 | 431 | 286 | 22 | 9 |
| 75 | 180 | 195 | 20 | 8 |
| 55 | 132 | 273 | 34 | 7 |
| 345 | 828 | 429 | 80 | 61 |
| 285 | 684 | 325 | 58 | 54 |
| 215 | 516 | 247 | 48 | 40 |
| 350 | 840 | 481 | 72 | 69 |
| 195 | 472 | 273 | 48 | 31 |
| 305 | 690 | 507 | 64 | 58 |
| 110 | 430 | 520 | 40 | 12 |

Вариант 3

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 180 | 16250 | 840 | 1534 | 124 |
| 102 | 8875 | 528 | 624 | 88 |
| 168 | 15000 | 768 | 1404 | 108 |
| 90 | 8750 | 384 | 520 | 68 |
| 168 | 14750 | 720 | 1326 | 100 |
| 102 | 8875 | 792 | 858 | 84 |
| 240 | 19450 | 984 | 1950 | 140 |
| 144 | 12850 | 648 | 988 | 100 |
| 84 | 16250 | 840 | 676 | 76 |
| 174 | 14925 | 744 | 1404 | 108 |
| 204 | 17800 | 888 | 1690 | 132 |
| 114 | 9975 | 552 | 884 | 84 |
| 96 | 8500 | 408 | 780 | 76 |
| 54 | 13250 | 840 | 416 | 60 |
| 156 | 14300 | 744 | 1248 | 92 |
| 234 | 19900 | 960 | 1820 | 148 |
| 294 | 23750 | 1272 | 2392 | 164 |
| 288 | 24325 | 1224 | 2340 | 148 |
| 276 | 23850 | 1224 | 572 | 168 |
| 282 | 22375 | 1248 | 910 | 180 |
| 210 | 18925 | 888 | 936 | 132 |
| 117 | 10850 | 984 | 624 | 84 |
| 183 | 16925 | 360 | 312 | 180 |
| 120 | 18750 | 120 | 390 | 108 |

Вариант 4

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 240 | 13000 | 805 | 1593 | 93 |
| 216 | 13000 | 644 | 1377 | 66 |
| 260 | 12400 | 805 | 1620 | 81 |
| 340 | 14600 | 1035 | 2187 | 126 |
| 272 | 16800 | 828 | 1674 | 99 |
| 380 | 14400 | 1150 | 2484 | 135 |
| 212 | 16200 | 667 | 1242 | 81 |
| 268 | 11000 | 828 | 1674 | 105 |
| 244 | 14600 | 759 | 1593 | 99 |
| 236 | 14000 | 736 | 1431 | 81 |
| 296 | 9000 | 897 | 1971 | 111 |
| 232 | 4200 | 713 | 1350 | 63 |
| 172 | 5200 | 552 | 1080 | 57 |
| 304 | 12800 | 920 | 1917 | 99 |
| 320 | 17800 | 966 | 2052 | 123 |
| 288 | 13200 | 874 | 1890 | 111 |
| 372 | 13600 | 1127 | 2295 | 123 |
| 224 | 15600 | 690 | 1485 | 111 |
| 296 | 8000 | 897 | 1917 | 126 |
| 136 | 3400 | 437 | 783 | 51 |
| 216 | 5000 | 644 | 1377 | 99 |
| 156 | 8680 | 943 | 648 | 63 |
| 244 | 13540 | 345 | 324 | 135 |
| 160 | 15000 | 345 | 405 | 81 |

Вариант 5

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 305 | 1300 | 66 | 120 | 91 |
| 275 | 1140 | 58 | 280 | 82 |
| 305 | 1240 | 70 | 310 | 91 |
| 335 | 1460 | 74 | 330 | 95 |
| 285 | 1148 | 54 | 290 | 85 |
| 130 | 714 | 28 | 170 | 41 |
| 385 | 1620 | 62 | 390 | 99 |
| 115 | 534 | 34 | 150 | 35 |
| 260 | 1170 | 52 | 270 | 77 |
| 185 | 914 | 30 | 210 | 57 |
| 160 | 750 | 34 | 190 | 49 |
| 280 | 904 | 50 | 270 | 82 |
| 120 | 520 | 24 | 220 | 37 |
| 155 | 824 | 30 | 170 | 45 |
| 160 | 886 | 42 | 180 | 45 |
| 135 | 624 | 38 | 150 | 42 |
| 175 | 886 | 22 | 210 | 51 |
| 335 | 1394 | 64 | 310 | 77 |
| 220 | 800 | 44 | 270 | 65 |
| 175 | 780 | 30 | 190 | 53 |
| 130 | 546 | 28 | 150 | 41 |
| 120 | 690 | 24 | 170 | 37 |
| 385 | 1354 | 50 | 390 | 57 |
| 375 | 1500 | 30 | 400 | 50 |

Вариант 6

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 84 | 2800 | 40 | 210 | 85 |
| 322 | 1640 | 210 | 1435 | 320 |
| 623 | 3360 | 430 | 1925 | 523 |
| 315 | 1600 | 210 | 1505 | 325 |
| 532 | 2840 | 320 | 2590 | 457 |
| 623 | 3360 | 430 | 2905 | 599 |
| 469 | 2480 | 310 | 2240 | 457 |
| 238 | 1160 | 150 | 1050 | 235 |
| 469 | 2480 | 310 | 2310 | 455 |
| 609 | 3280 | 410 | 2905 | 570 |
| 686 | 3720 | 470 | 3325 | 650 |
| 224 | 1080 | 140 | 1120 | 225 |
| 315 | 1600 | 210 | 1400 | 297 |
| 546 | 2920 | 370 | 2450 | 534 |
| 623 | 3360 | 430 | 1995 | 621 |
| 455 | 2400 | 300 | 525 | 477 |
| 315 | 1600 | 210 | 1540 | 315 |
| 546 | 2920 | 370 | 2660 | 532 |
| 532 | 2840 | 310 | 2450 | 535 |
| 238 | 1160 | 150 | 1155 | 239 |
| 182 | 1092 | 140 | 525 | 355 |
| 168 | 1380 | 120 | 770 | 275 |
| 539 | 2708 | 440 | 875 | 435 |
| 525 | 3000 | 200 | 175 | 345 |

Вариант 7

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 60 | 1400 | 22 | 60 | 25 |
| 230 | 820 | 62 | 410 | 89 |
| 445 | 1680 | 110 | 550 | 155 |
| 215 | 960 | 58 | 370 | 86 |
| 115 | 520 | 30 | 150 | 46 |
| 270 | 1220 | 70 | 440 | 108 |
| 335 | 1420 | 90 | 600 | 134 |
| 485 | 1980 | 130 | 900 | 99 |
| 60 | 340 | 24 | 110 | 24 |
| 230 | 1040 | 62 | 400 | 92 |
| 325 | 1400 | 86 | 450 | 33 |
| 85 | 440 | 24 | 100 | 34 |
| 95 | 480 | 24 | 140 | 38 |
| 460 | 1940 | 118 | 840 | 184 |
| 155 | 780 | 42 | 230 | 62 |
| 380 | 1620 | 100 | 550 | 152 |
| 70 | 340 | 22 | 140 | 28 |
| 170 | 780 | 44 | 240 | 68 |
| 390 | 1640 | 104 | 500 | 22 |
| 205 | 920 | 54 | 310 | 82 |
| 190 | 800 | 50 | 110 | 76 |
| 305 | 1360 | 82 | 220 | 33 |
| 435 | 1780 | 24 | 330 | 22 |
| 375 | 1086 | 22 | 220 | 50 |

Вариант 8

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 84 | 2800 | 120 | 210 | 36 |
| 322 | 1640 | 70 | 1435 | 33 |
| 70 | 480 | 30 | 280 | 31 |
| 105 | 680 | 30 | 420 | 44 |
| 175 | 1120 | 50 | 525 | 75 |
| 161 | 880 | 50 | 735 | 69 |
| 77 | 440 | 40 | 350 | 32 |
| 98 | 640 | 30 | 385 | 43 |
| 154 | 960 | 50 | 665 | 67 |
| 210 | 1160 | 60 | 700 | 87 |
| 168 | 920 | 50 | 665 | 77 |
| 147 | 880 | 50 | 630 | 65 |
| 126 | 760 | 40 | 490 | 54 |
| 91 | 440 | 30 | 350 | 39 |
| 133 | 920 | 50 | 350 | 57 |
| 112 | 680 | 40 | 315 | 48 |
| 231 | 1480 | 70 | 875 | 98 |
| 203 | 1320 | 70 | 700 | 87 |
| 182 | 1200 | 50 | 735 | 78 |
| 147 | 1000 | 50 | 700 | 63 |
| 119 | 840 | 40 | 350 | 51 |
| 140 | 960 | 40 | 525 | 59 |
| 609 | 3560 | 170 | 1155 | 125 |
| 525 | 2172 | 110 | 770 | 122 |

Вариант 9

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 280 | 1160 | 50 | 440 | 36 |
| 115 | 400 | 22 | 210 | 33 |
| 65 | 300 | 14 | 110 | 31 |
| 270 | 1020 | 56 | 330 | 44 |
| 105 | 320 | 22 | 120 | 75 |
| 445 | 1880 | 70 | 770 | 69 |
| 135 | 620 | 30 | 250 | 32 |
| 180 | 640 | 38 | 310 | 43 |
| 455 | 1760 | 44 | 870 | 67 |
| 275 | 1280 | 44 | 450 | 87 |
| 360 | 1380 | 70 | 650 | 77 |
| 165 | 760 | 34 | 270 | 65 |
| 250 | 800 | 50 | 370 | 54 |
| 65 | 300 | 14 | 110 | 39 |
| 375 | 1640 | 44 | 550 | 57 |
| 55 | 260 | 10 | 90 | 48 |
| 90 | 500 | 18 | 150 | 98 |
| 245 | 1080 | 44 | 370 | 87 |
| 305 | 1240 | 14 | 550 | 78 |
| 220 | 1060 | 42 | 410 | 63 |
| 85 | 420 | 18 | 100 | 51 |
| 100 | 480 | 20 | 150 | 59 |
| 435 | 670 | 10 | 330 | 125 |
| 375 | 500 | 10 | 220 | 122 |

Вариант 10

| Y | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 161 | 1932 | 120 | 525 | 35 |
| 238 | 2856 | 190 | 875 | 46 |
| 378 | 4536 | 300 | 1575 | 66 |
| 147 | 1764 | 140 | 595 | 33 |
| 462 | 5544 | 380 | 1995 | 78 |
| 518 | 6216 | 360 | 2345 | 86 |
| 553 | 6636 | 420 | 2415 | 91 |
| 315 | 3780 | 260 | 1295 | 57 |
| 252 | 3024 | 220 | 1085 | 48 |
| 259 | 3108 | 240 | 1155 | 49 |
| 336 | 3180 | 230 | 1575 | 60 |
| 189 | 2268 | 130 | 875 | 39 |
| 154 | 1848 | 140 | 665 | 34 |
| 168 | 2016 | 150 | 735 | 36 |
| 651 | 7812 | 460 | 2765 | 57 |
| 644 | 7728 | 490 | 2695 | 48 |
| 637 | 7644 | 440 | 3115 | 98 |
| 322 | 3864 | 220 | 1295 | 87 |
| 329 | 3948 | 50 | 1225 | 78 |
| 217 | 2604 | 70 | 945 | 63 |
| 119 | 1428 | 90 | 420 | 51 |
| 140 | 2200 | 100 | 525 | 59 |
| 609 | 6280 | 50 | 2695 | 125 |
| 525 | 5000 | 50 | 2345 | 122 |

## Контрольные вопросы

1. Что показывает уравнение множественной регрессии?
2. Как записывается уравнение множественной линейной регрессии?
3. Как определить корреляционную связь между факторами?
4. Какое условие для включения факторов в модель регрессии?
5. Что означает значения индекса множественной корреляции?
6. Как определяется индекс детерминации?
7. Что такое уровень значимости?
8. Как осуществляется прогноз по уравнению линейной множественной регрессии?
9. Каким образом сравнивают уравнения линейной множественной регрессии?

# 3. ПРОГНОЗИРОВАНИЕ

## Лабораторная работа № 7
## Прогнозирование

**Цель лабораторной работы:** изучить основные понятия, связанные с прогнозированием, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

*Определение.* **Временной (динамический) ряд –** это совокупность значений какого-либо показателя за несколько последовательных моментов или интервалов времени.

Соответственно ряды подразделяются на:

1) **интервальные**, если значения относятся к интервалам времени (например, количество преступлений, совершенных за месяц);

2) **моментные**, если значения относятся к конкретным моментам времени (например, количество нераскрытых преступлений на 1 число каждого месяца).

При изучении временных рядов решают следующие **задачи:**

– определение характерных особенностей ряда;

– подбор статистической модели;

– прогноз будущих значений на основе прошлых наблюдений;

– исследование взаимосвязи ряда с другими рядами.

*Определение.* Значение ряда, относящееся к моменту времени $t$ обозначается $y_t$ и называется **уровнем ряда**. Последовательность всех уровней ряда будем обозначать

$$Y = (y_1, y_2, ..., y_n),$$

где $n$ – число уровней ряда.

Показатели, характеризующие уровни ряда, обычно соотносятся с некоторым уровнем $y_{баз}$, который принимается в качестве базового.

К таким показателям относят:

1) **абсолютный прирост** – величину, показывающую, на сколько уровень $y_t$ больше базового

$$y_t - y_{баз};$$

2) **темп роста** — величину, показывающую, во сколько раз уровень $y_t$ больше базового или сколько процентов он составляет от базового

$$\frac{y_t}{y_{баз}} \text{ или } \frac{y_t}{y_{баз}} \cdot 100\%;$$

3) **темп прироста** – величину, показывающую приращение уровня $y_t$ по сравнению с базовым

$$\frac{y_t - y_{баз}}{y_{баз}} \text{ или } \frac{y_t - y_{баз}}{y_{баз}} \cdot 100\%.$$

Выделяют следующие характеристики временного ряда:

1) **максимальным значением**

$y_{\max} = \max\limits_{t} y_t$;

2) **минимальным значением**

$y_{\min} = \min\limits_{t} y_t$;

3) **амплитудой изменения**

$\Delta y = y_{\max} - y_{\min}$;

4) **средним значением**

$$y_{ср} = \frac{y_1 + y_2 + ... + y_n}{n}.$$

*Определение.* Если число уровней ряда $n$ четно, то $y_{ср}$ не соотносится ни с одним уровнем. В этом случае удобно использовать **среднее хронологическое значение**, при расчете которого число уровней искусственно уменьшено на один:

$$y_{ср} = \frac{\frac{1}{2}y_1 + y_2 + ... + \frac{1}{2}y_n}{n-1}.$$

*Пример 7.1.* Изучается явление $Y$. Данные за семь лет сведены в таблицу. Определим показатели, характеризующие уровни ряда.

| | 1 год | 2 год | 3 год | 4 год | 5 год | 6 год | 7 год |
|---|---|---|---|---|---|---|---|
| $Y$ | 16,2 | 17,4 | 22,0 | 24,5 | 19,3 | 20,7 | 18,5 |

*Решение.* Для заданного временного ряда

1) максимальное значение $y_{\max} = \max\limits_{t} y_t = 24{,}5$;

2) минимальное значение $y_{\min} = \min\limits_{t} y_t = 16{,}2$;

3) амплитуда изменения $\Delta y = y_{\max} - y_{\min} = 24{,}5 - 16{,}2 = 8{,}3$;

4) среднее значение $y_{ср} = \frac{y_1 + y_2 + ... + y_n}{n} =$

$$= \frac{16{,}2 + 17{,}4 + 22{,}0 + ... + 20{,}7 + 18{,}5}{7} = 19{,}8.$$

| | 1 год | 2 год | 3 год | 4 год | 5 год | 6 год | 7 год |
|---|---|---|---|---|---|---|---|
| $Y$ | 16,2 | 17,4 | 22,0 | 24,5 | 19,3 | 20,7 | 18,5 |
| абсолютный прирост $y_t - y_{баз}$ | — | 1,2 | 5,8 | 8,3 | 3,1 | 4,5 | 2,3 |
| темп роста $\frac{y_t}{y_{баз}} \cdot 100\%$ | — | 107,40 | 135,80 | 151,23 | 119,13 | 127,77 | 114,19 |
| темп прироста $\frac{y_t - y_{баз}}{y_{баз}} \cdot 100\%$ | — | 7,40 | 35,80 | 51,23 | 19,13 | 27,77 | 14,19 |

## Модели временных рядов

На численные значения уровней ряда оказывают влияние множество факторов, которые подразделяются на три группы:

1) **постоянные факторы**, совокупный эффект действия которых сказывается постепенно и не является периодическим (эти факторы формируют тенденцию (тренд ряда);

2) **периодически проявляющиеся факторы**, которые определяют циклические (в частности, сезонные) изменения уровней ряда;

3) **факторы, которые носят непостоянный характер** (эти факторы определяют случайную составляющую уровней ряда).

Каждая из описанных групп факторов может отсутствовать.

Последовательность значений уровней ряда, полученных в результате воздействия факторов, формирующих **тенденцию**, будем обозначать – $T = (t_1, t_2, ..., t_n)$;

последовательность значений уровней ряда, полученных в результате воздействия факторов, формирующих **сезонные колебания** – $S = (s_1, s_2, ..., s_n)$;

последовательность значений уровней ряда, полученных в результате воздействия факторов, которые носят **случайный характер** – $E = (e_1, e_2, ..., e_n)$.

Таким образом, модель временного ряда представляет собой функцию трех компонент: тренда $T$, сезонной $S$, случайной $E$, т. е.

$Y = f(T, S, E)$.

Как правило, рассматривают два типа моделей:

1) **аддитивную**, представляющую собой сумму перечисленных компонент

$Y = T + S + E$;

2) **мультипликативную**, представляющую собой произведение перечисленных компонент

$Y = T \cdot S \cdot E$.

Иногда еще рассматривают **смешанную** модель $Y = T \cdot S + E$.

Основная задача исследования временного ряда состоит в определении типа его модели и количественном выражении каждой из компонент в интересах прогнозирования будущих значений ряда или исследования взаимосвязи двух и более временных рядов.

Обычно моделирование временных рядов начинается с выявления наличия сезонной компоненты и – в случае ее наличия – определения количества уровней, относящихся к одному периоду.

Для этого используется функция автокорреляции уровней ряда. Количественно ее определяют с помощью коэффициента корреляции между уровнями исходного ряда и уровнями ряда, которые сдвинуты на несколько шагов.

Обозначим

$Y_1$ – ряд, уровни которого сдвинуты на 1;

$Y_2$ – ряд, уровни которого сдвинуты на 2 и т.д.;

$Y_k$ – ряд, уровни которого сдвинуты на $k$.

Необходимо найти значения коэффициентов корреляции $r_{YY_1}, r_{YY_2}, ..., r_{YY_k}$ уровней исходного ряда $Y$ и рядов $Y_1, Y_2, ..., Y_k$. Для того, чтобы не нарушить достоверность результата, $k$ не должно превышать $\frac{n}{4}$.

Непосредственной проверкой определим

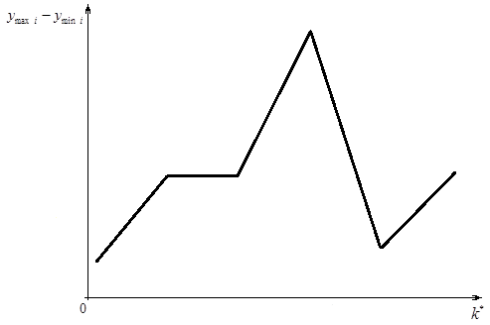
$r^*$ – максимальное по модулю значение $r_{YY_i}$,

$k^*$ – минимальный номер уровня ряда, для которого $r_{YY_i} = r^*$.

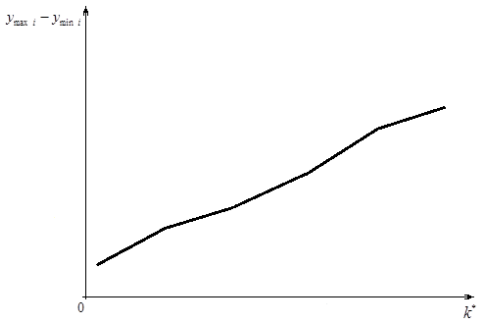
Если $r^*$ близок к 1, то это свидетельствует о наличии **автокорреляции уровней ряда.**

Если $k^* = 1$, то ряд не содержит сезонной компоненты; а если $k^* > 1$, то можно сделать вывод о наличии сезонной компоненты с периодом, включающим $k^*$ уровней ряда.

После этого необходимо найти амплитуду изменения уровней ряда для каждого периода изменения сезонной компоненты, т. е. разность между максимальным и минимальным значением. Это позволяет построить график изменения амплитуды




Аддитивная модель




Мультипликативная модель

Рис. 7.1

Проводя анализ графика разностей между максимальными и минимальными значениями за сезон, делаем вывод о том, что модель временного ряда аддитивная или мультипликативная.

Если амплитуда изменяется монотонно, т. е. либо возрастает, либо убывает, то следует строить мультипликативную модель $Y = T \cdot S \cdot E$, иначе аддитивную – $Y = T + S + E$.

*Пример 7.2.* В результате сбора данных о количестве преступлений по двум статьям УК были получены временные ряды.

Ряд А

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 8 | 9 | 0 | 5 | 6 | 24 | 4 | 6 | 17 | 52 | 5 | 17 | 38 | 80 | 9 |

Ряд Б

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 32 | | 17 | 41 | 21 | 44 | 39 | 70 | 28 | 84 | 42 | 88 | 64 | 05 | 56 | 98 | 80 |

*Решение.* С помощью корреляционного анализа выявлена сезонность для ряда А, включающая 4 уровня, ряда Б – 2 уровня.

Для нахождения разностей между максимальными и минимальными значениями за сезон построим таблицы.

Для ряда А

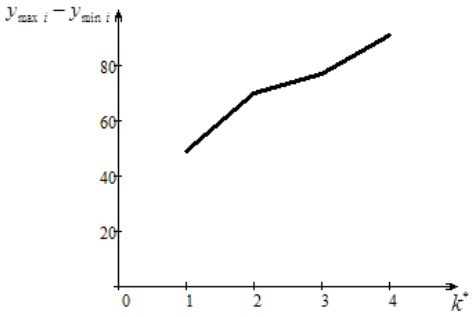
| | | | | |
|---|---|---|---|---|
| 1 уровень | 54 | 75 | 96 | 117 |
| 2 уровень | 68 | 96 | 117 | 138 |
| 3 уровень | 89 | 124 | 152 | 180 |
| 4 уровень | 40 | 54 | 75 | 89 |
| $y_{\max i} - y_{\min i}$ | 89–40=**49** | 124–54=**70** | 152–75=**77** | 180–89=**91** |

Для ряда Б

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 уровень | 132 | 141 | 144 | 170 | 184 | 188 | 205 | 198 |
| 2 уровень | 117 | 121 | 139 | 128 | 142 | 164 | 156 | 180 |
| $y_{\max i} - y_{\min i}$ | 15 | 20 | 5 | 42 | 42 | 24 | 49 | 18 |

Построим графики амплитуд изменения уровней ряда для каждого периода изменения сезонной компоненты.

Ряд А




Ряд Б

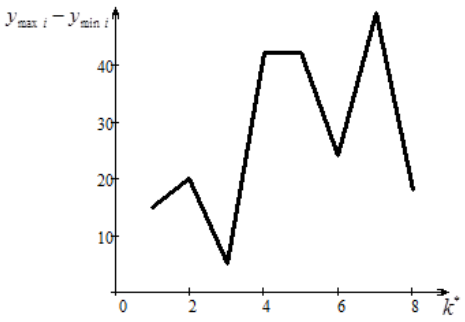



Рис. 7.2

Так как для ряда А амплитуда изменяется монотонно (возрастает), то строим мультипликативную модель ($Y = T \cdot S \cdot E$). Для ряда Б амплитуда изменяется немонотонно, поэтому строим аддитивную модель ($Y = T + S + E$).

## Прогнозирование без выявления структуры ряда. Метод усредненного цикла

Первоначально рассмотрим метод усредненного цикла, который позволяет прогнозировать значения уровней ряда без выявления сезонной и трендовой компонент.

Идея метода заключается в усреднении значений прироста уровней ряда по отношению к предыдущему уровню по всем периодам изменения сезонной компоненты. При этом для аддитивной модели рассматривается абсолютный прирост, а для мультипликативной – темп прироста.

Пусть задан временной ряд

$$\{x_{ij}\} = \{x_{11} \quad x_{12} \quad x_{13} \quad \dots \quad x_{1\,n-1} \quad x_{1n}$$
$$x_{21} \quad x_{22} \quad x_{23} \quad \dots \quad x_{2\,n-1} \quad x_{2n}$$
$$\dots$$
$$x_{m-1,1} \; x_{m-1,2} \; x_{m-1,3} \dots \; x_{m-1,\,n-1} \; x_{m-1,\,n}$$
$$x_{m1} \quad x_{m2} \quad x_{m3} \quad \dots \quad x_{m\,n-1} \quad x_{mn}\},$$

где $i = 1, 2, ..., m$, $m$ – количество циклов;

$j = 1, 2, ..., n$, $n$ – количество данных в одном цикле.

Будем считать, что ряд включает целое число периодов. Если это не так, то из рассмотрения исключаются первые уровни ряда до тех пор, пока условие не будет выполнено.

Обозначим последнее значение временного ряда $x_{mn} = y_0$, а $y_k$ $(k \in \mathrm{N})$ – прогнозируемые значения временного ряда.

Тогда прогнозируемые значения уровней временного ряда

для аддитивной модели:

$$y_1^{cp} = \frac{1}{m-1} \cdot \sum_{i=1}^{m-1} \left(x_{i+1,1} - x_{in}\right) + y_0,$$

$$y_k^{cp} = \frac{1}{m} \cdot \sum_{i=1}^{m} \left(x_{ik} - x_{i\,k-1}\right) + y_{k-1}, \qquad k = 2, 3, ..., n;$$

для мультипликативной модели:

$$y_1^{cp} = \frac{1}{m-1} \cdot \left( \sum_{i=1}^{m-1} \frac{x_{i+1,1}}{x_{in}} \right) \cdot y_0,$$

$$y_k^{cp} = \frac{1}{m} \cdot \left( \sum_{i=1}^{m} \frac{x_{ik}}{x_{i\,k-1}} \right) \cdot y_{k-1}, \qquad k = 2, 3, ..., n.$$

Определим средние и несмещенные выборочные дисперсии для аддитивной модели

$$\bar{x}_1^{cp} = \frac{1}{m-1} \cdot \sum_{i=1}^{m-1} \left(x_{i+1,1} - x_{in}\right), \; \bar{x}_k^{cp} = \frac{1}{m} \cdot \sum_{i=1}^{m} \left(x_{ik} - x_{i\,k-1}\right),$$

$$s_{01}^2 = \frac{1}{m-2} \cdot \sum_{i=1}^{m-1} \left[\left(x_{i+1,1} - x_{in}\right) - \bar{x}_1^{cp}\right], \; s_{0k}^2 = \frac{1}{m-1} \cdot \sum_{i=1}^{m} \left[\left(x_{ik} - x_{i\,k-1}\right) - \bar{x}_k^{cp}\right];$$

для мультипликативной модели временного ряда

$$\bar{x}_1^{cp} = \frac{1}{m-1} \cdot \sum_{i=1}^{m-1} \frac{x_{i+1,1}}{x_{i\,n}},\ \bar{x}_k^{cp} = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{x_{i\,k}}{x_{i\,k-1}},$$

$$s_{0\,1}^2 = \frac{1}{m-2} \cdot \sum_{i=1}^{m-1} \left[ \frac{x_{i+1,1}}{x_{i\,n}} - \bar{x}_1^{cp} \right],\ s_{0\,k}^2 = \frac{1}{m-1} \cdot \sum_{i=1}^{m} \left[ \frac{x_{i\,k}}{x_{i\,k-1}} - \bar{x}_k^{cp} \right].$$

Тогда доверительный интервал найдем из соотношения

$$\bar{x}_1^{cp} - t_{m-2,\,\alpha} \cdot s_{0\,s} < y_1^{cp} < \bar{x}_1^{cp} + t_{m-2,\,\alpha} \cdot s_{0\,s},$$

$$\bar{x}_k^{cp} - t_{m-1,\,\alpha} \cdot s_{0\,s} < y_k^{cp} < \bar{x}_k^{cp} + t_{m-1,\,\alpha} \cdot s_{0\,s},\ s = 1, 2, ..., n,$$

где $t_{s,\,\alpha}$ – значение распределения Стьюдента при уровне значимости $\alpha$.

Пример будет рассмотрен на лабораторном занятии.

## Выявление сезонной и трендовой компонент

Для нахождения сезонной компоненты может быть использован **метод скользящей средней**. Его суть заключается в следующем.

За значение $k^*$ возьмем количество уровней ряда, входящих в один период изменения сезонной компоненты. Первоначально находятся скользящие средние $y_{скол,\,i}$ по формуле

$$y_{скол,\,i} = \begin{cases} \frac{1}{2} y_{i-\frac{k^*}{2}} + y_{i-\frac{k^*}{2}+1} + ... + y_{i+\frac{k^*}{2}-1} + \frac{1}{2} y_{i+\frac{k^*}{2}}, & \text{если } k^* \text{ четно;} \\ y_{i-\frac{k^*-1}{2}} + y_{i-\frac{k^*}{2}+1} + ... + y_{i+\frac{k^*}{2}-1} + y_{i+\frac{k^*-1}{2}}, & \text{если } k^* \text{ нечетно.} \end{cases}$$

Полученные оценки после корректировки преобразуются в значения сезонной компоненты $S = (s_1, s_2, ..., s_n)$ так, чтобы

$\sum_{i=1}^{k^*} s_i = 0$ – для аддитивной модели и

$\prod_{i=1}^{k^*} s_i = 1$ – для мультипликативной модели.

После нахождения сезонной компоненты ее можно легко исключить:

$T + E = Y - S$ – для аддитивной модели,

$T \cdot E = \frac{Y}{S}$ – для мультипликативной модели.

Трендовая компонента $T = (t_1, t_2, ..., t_n)$ находится с помощью построения уравнения парной регрессии $\hat{T} = f(N)$, где в качестве независимой переменной выступают номера уровней ряда

$N = (1, 2, ..., n)$.

После нахождения трендовой компоненты находится случайная компонента по формулам:

$E = Y - S - T$ — для аддитивной модели,

$E = \frac{Y}{S \cdot T}$ — для мультипликативной модели.

В качестве оценки эффективности моделей используется величина

$$\varepsilon = \frac{(e_1)^2 + (e_2)^2 + ... + (e_n)^2}{n}.$$

Чем меньше $\varepsilon$, тем точнее модель.

## Использование фиктивных переменных при построении модели временного ряда

Иногда оказывается возможным для определения сезонной и трендовой компонент использовать фиктивные переменные.

Пусть $k$ – длина периода изменения сезонной компоненты. Введем переменные

$$Z_1 = \begin{cases} 1 - \text{для первого уровня периода,} \\ 0 - \text{для всех остальных;} \end{cases}$$

…

$$Z_{k-1} = \begin{cases} 1 - \text{для k - 1 уровня периода,} \\ 0 - \text{для всех остальных.} \end{cases}$$

Построим уравнение

$$\hat{Y} = a + b \cdot N + c_1 \cdot Z_1 + ... + c_{k-1} \cdot Z_{k-1}.$$

Если оказывается, что все переменные $Z_1, ..., Z_{k-1}$ значимы, то величины $c_1, ..., c_{k-1}, 0$ описывают сезонную компоненту, а $\hat{Y} = a + b \cdot N$ представляет собой уравнение тренда.

Достоинство метода в его простоте и в том, что он позволяет одновременно найти все компоненты ряда, недостаток – часто некоторые фиктивные переменные оказываются незначимыми в силу недостаточного объема данных. Как правило, метод успешно применяется, если $k = 2$.

С помощью фиктивных переменных можно также проверить наличие или отсутствие скачкообразного изменения тренда (**метод Гуйарати**)

Вводится фиктивная переменная

$$Z = \begin{cases} 1 - \text{для первой части уровней ряда,} \\ 0 - \text{для всех остальных.} \end{cases}$$

Строится уравнение

$$\hat{Y} = a + b \cdot N + c \cdot Z$$

Если $Z$ значимо, то имеет место скачкообразное изменение тренда и следует строить для него два уравнения:

$\hat{Y} = a + c + b \cdot N$ — для первой части уровней;

$\hat{Y} = a + b \cdot N$ — для второй части уровней.

## Исследование взаимосвязей временных рядов

Важной задачей является изучение взаимосвязей между временными рядами $X = f(T_X, S_X, E_X)$ и $Y = f(T_Y, S_Y, E_Y)$.

Показателем, характеризующим взаимосвязь, является коэффициент корреляции $r_{XY}$. Однако компоненты ряда могут сильно искажать эту оценку.

Если ряды имеют однонаправленный тренд, то значение $r_{XY}$ окажется завышенным, если же тренд разнонаправленный, то это значение будет заниженным.

То же самое можно сказать и о сезонной компоненте.

В связи с этим для исследования взаимосвязи указанные компоненты должны быть исключены и оценку взаимосвязи характеризует величина $r_{E_X E_Y}$.

*Пример 7.3.* За 6 лет работы подразделения полиции были собраны ежеквартальные данные о количестве краж, совершенных на подконтрольной территории, значения которых сведены в таблицу.

| Год | № квартала | Количество краж |
|---|---|---|
| 1 | 1 | 45 |
| | 2 | 55 |
| | 3 | 70 |
| | 4 | 50 |
| 2 | 5 | 51 |
| | 6 | 59 |
| | 7 | 75 |
| | 8 | 53 |
| 3 | 9 | 55 |
| | 10 | 63 |
| | 11 | 79 |
| | 12 | 62 |
| 4 | 13 | 60 |
| | 14 | 67 |
| | 15 | 85 |
| | 16 | 63 |
| 5 | 17 | 65 |
| | 18 | 77 |
| | 19 | 91 |
| | 20 | 69 |
| 6 | 21 | 70 |
| | 22 | 83 |
| | 23 | 94 |
| | 24 | 73 |

На основании имеющихся данных сделать прогноз оценки количества краж на год, следующий за отчетным периодом.

**Решение с помощью табличного процессора.**

Заносим данные в виде таблицы. Построим график временного ряда: во вкладке «*Вставка*» выбираем «*График*».

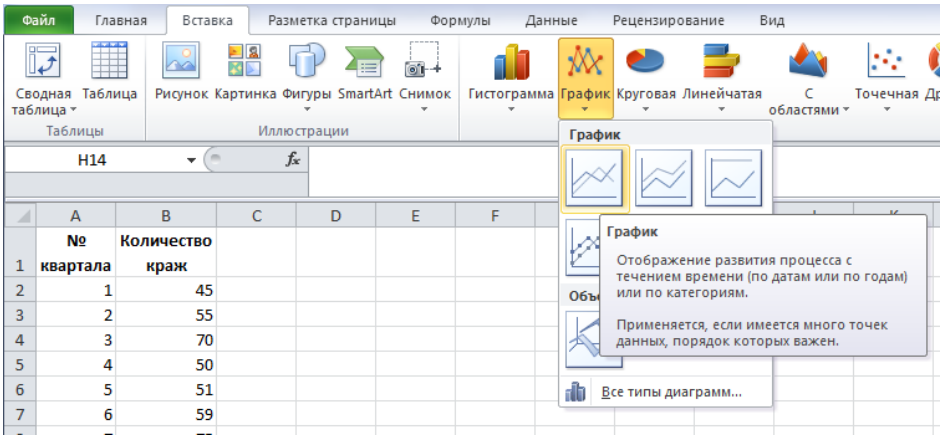



Рис. 7.3

В появившемся окне нажимаем правую клавишу мыши и нажимаем «*Выбрать данные…*».

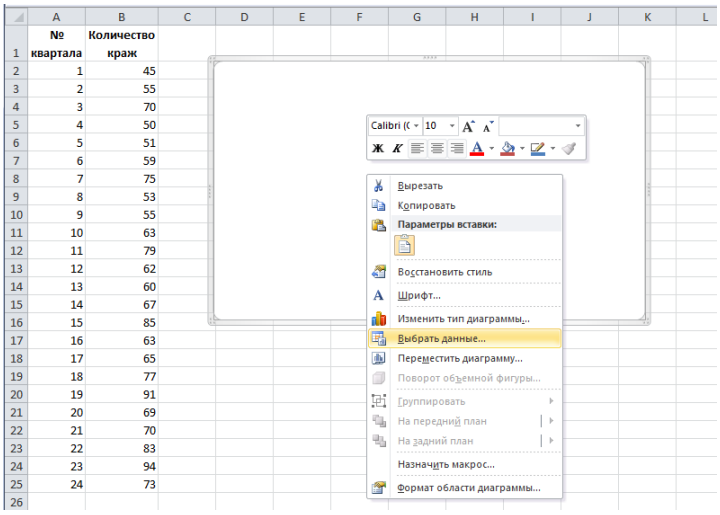



Рис. 7.4

В окне «Выбор источника данных» нажимаем «Добавить».

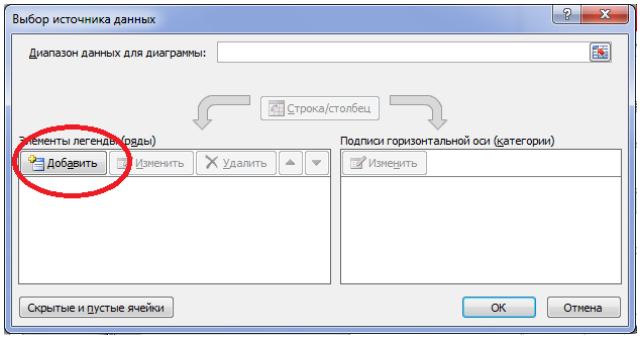



Рис. 7.5

В окне «*Изменение ряда*» в поле «*Имя ряда:*» вводим номер ячейки, содержащей название, в поле «*Значения:*» – номера ячеек с данными временного ряда. Далее «*ОК*» → «*ОК*».

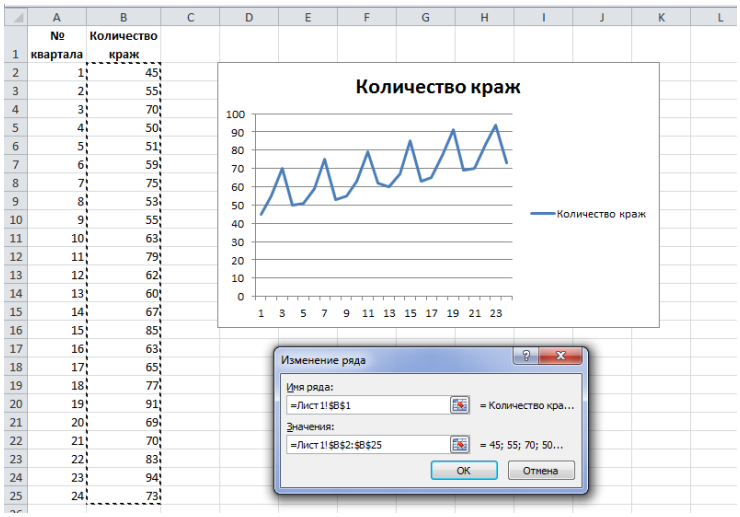



Рис. 7.6

По виду графика делаем предположение, что модель данного ряда имеет аддитивный характер.

Для определения сезонной компоненты воспользуемся методом скользящей средней.

➢Указание: Вместо метода скользящего среднего можно использовать также метод экспоненциального сглаживания. Используется для сглаживания краткосрочных колебаний во временных рядах (сезонную компоненту), чтобы облегчить определение тренда, а также для прогнозирования. В отличие от метода скользящего среднего, где прошлые наблюдения имеют одинаковый вес, экспоненциальное сглаживание присваивает им экспоненциально убывающие веса, по мере того как наблюдения становятся старше. Другими словами, последние наблюдения дают относительно больший вес при прогнозировании, чем старые наблюдения. Выберем «Данные» → «Анализ данных» → «Экспоненциальное сглаживание» (рис. 7.7).

Вернемся к выявлению сезонной компоненты с помощью метода скользящей средней. Алгоритм справедлив для данных, повторяющихся с определенной периодичностью. Очевидно, что количество краж из года в год в различные сезоны либо повышается, либо понижается, так, например, летом их значение увеличивается в связи с сезоном отпусков, и т.п. Проверим, имеют ли наши данные цикличность. Для этого копируем данные о количестве краж и вставляем их в любом свободном месте листа. Рядом продолжаем вставлять данные со сдвигом на одну ячейку вниз и т.д.

«Обрезаем» полученную таблицу снизу по первому столбцу (рис. 7.8).

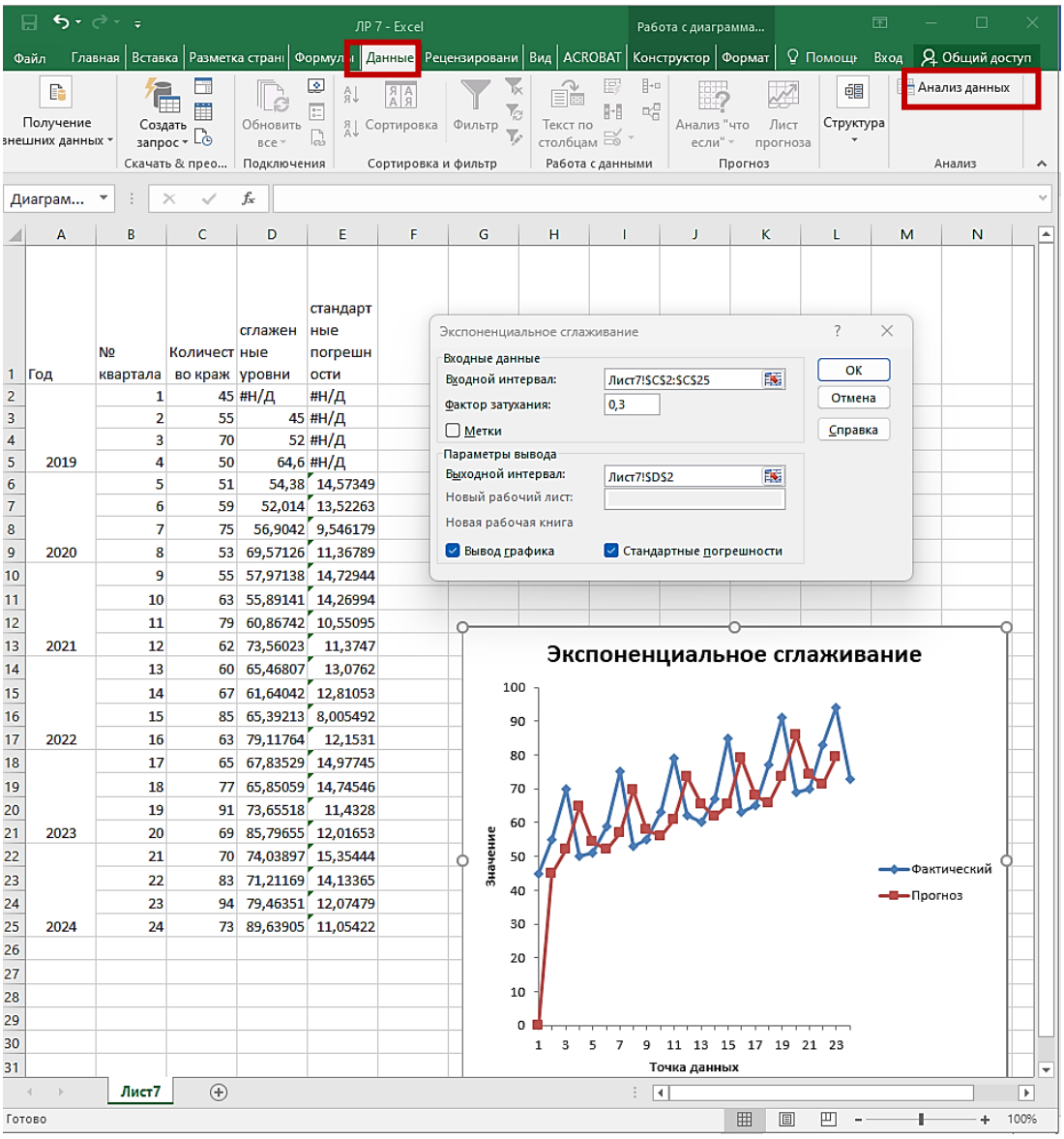



Рис. 7.7

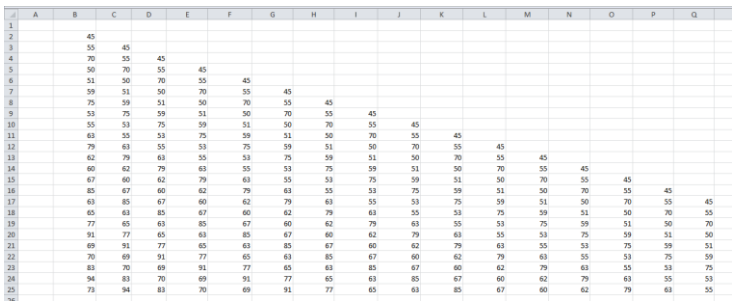
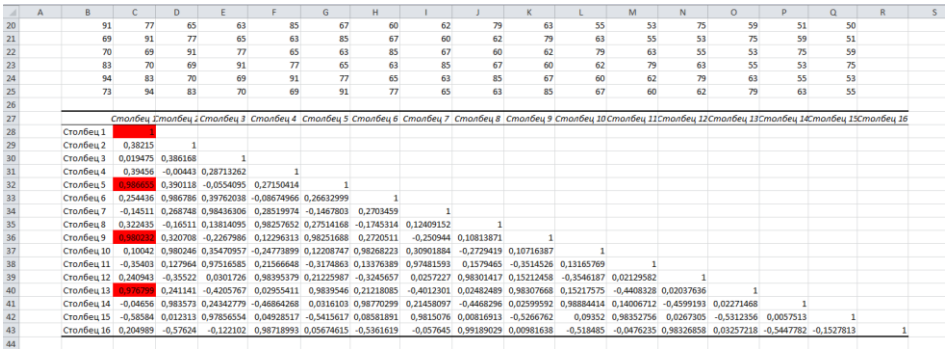


Рис. 7.8

Во вкладке «*Данные*» → «*Анализ данных*» → «*Корреляция*». В поле «*Входной интервал:*» вводим номера ячеек получившейся треугольной таблицы. Выведем расчетные значения на этот же лист, выбрав «*Выходной интервал:*» и указав ячейку (это левый верхний угол таблицы с результатами). Нажимаем «*ОК*».

Полученная таблица дает результат сравнения каждого столбца с каждым на предмет схожести. Рассмотрим значения столбца «Столбец 1». В каждой строчке первый столбец нашей таблицы сравнивался с самим собой и с последующими. Максимальное значение при сравнении 1 (коэффициент корреляции); если столбцы не похожи, то значение стремится к нулю. Очевидно, что первый столбец при сравнении с самим собой даст значение коэффициента сравнения 1. Выделим цветом те ячейки, в которых значение коэффициента сравнения близко к 1.

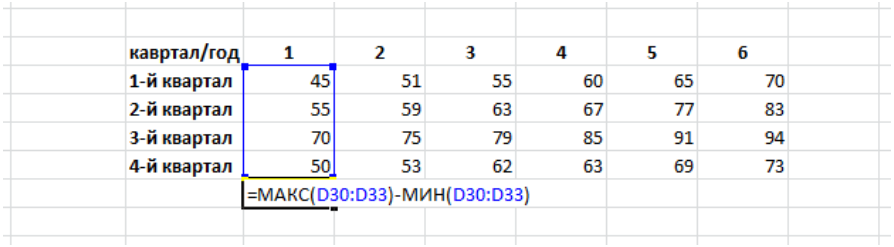


| | Столбец 1 | Столбец 2 | Столбец 3 | Столбец 4 | Столбец 5 | Столбец 6 | Столбец 7 | Столбец 8 | Столбец 9 | Столбец 10 | Столбец 11 | Столбец 12 | Столбец 13 | Столбец 14 | Столбец 15 | Столбец 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Столбец 1 | 1 | | | | | | | | | | | | | | | |
| Столбец 2 | 0,38215 | 1 | | | | | | | | | | | | | | |
| Столбец 3 | 0,019475 | 0,386168 | 1 | | | | | | | | | | | | | |
| Столбец 4 | 0,39456 | -0,00443 | 0,28713262 | 1 | | | | | | | | | | | | |
| Столбец 5 | 0,986655 | 0,390118 | -0,0554095 | 0,27150414 | 1 | | | | | | | | | | | |
| Столбец 6 | 0,254436 | 0,986786 | 0,39762038 | -0,08674966 | 0,26632999 | 1 | | | | | | | | | | |
| Столбец 7 | -0,14511 | 0,268748 | 0,98436306 | 0,28519974 | -0,1467803 | 0,2703459 | 1 | | | | | | | | | |
| Столбец 8 | 0,322435 | -0,16511 | 0,13814095 | 0,98257652 | 0,27514168 | -0,1745314 | 0,12409152 | 1 | | | | | | | | |
| Столбец 9 | 0,980232 | 0,320708 | -0,2267986 | 0,12296313 | 0,98251688 | 0,2720511 | -0,250944 | 0,10813871 | 1 | | | | | | | |
| Столбец 10 | 0,10042 | 0,980246 | 0,35470957 | -0,24773899 | 0,12208747 | 0,98268223 | 0,30901884 | -0,2729419 | 0,10716387 | 1 | | | | | | |
| Столбец 11 | -0,35403 | 0,127964 | 0,97516585 | 0,21566648 | -0,3174863 | 0,13376389 | 0,97481593 | 0,1579465 | -0,3514526 | 0,13165769 | 1 | | | | | |
| Столбец 12 | 0,240943 | -0,35522 | 0,0301726 | 0,98395379 | 0,21225987 | -0,3245657 | 0,0257227 | 0,98301417 | 0,15212458 | -0,3546187 | 0,02129582 | 1 | | | | |
| Столбец 13 | 0,976799 | 0,241141 | -0,4205767 | 0,02955411 | 0,9839546 | 0,21218085 | -0,4012301 | 0,02482489 | 0,98307668 | 0,15217575 | -0,4408328 | 0,02037636 | 1 | | | |
| Столбец 14 | -0,04656 | 0,983573 | 0,24342779 | -0,46864268 | 0,0316103 | 0,98770299 | 0,21458097 | -0,4468296 | 0,02599592 | 0,98884414 | 0,14006712 | -0,4599193 | 0,02271468 | 1 | | |
| Столбец 15 | -0,58584 | 0,012313 | 0,97856554 | 0,04928517 | -0,5415617 | 0,08581891 | 0,9815076 | 0,00816913 | -0,5266762 | 0,09352 | 0,98352756 | 0,0267305 | -0,5312356 | 0,0057513 | 1 | |
| Столбец 16 | 0,204989 | -0,57624 | -0,122102 | 0,98718993 | 0,05674615 | -0,5361619 | -0,057645 | 0,99189029 | 0,00981638 | -0,518485 | -0,0476235 | 0,98326858 | 0,03257218 | -0,5447782 | -0,1527813 | 1 |

Рис. 7.9

Видно, что ячейки, выделенные красным цветом, повторяются с периодичностью 4, то есть у такого преступления, как кражи, есть сезонность.

Проверим, что модель временного ряда является аддитивной. Представим исходные данные в виде таблицы и найдем разницу между максимальным и минимальными значениями за каждый год. Построим график разностей между максимальными и минимальными значениями за каждый год.

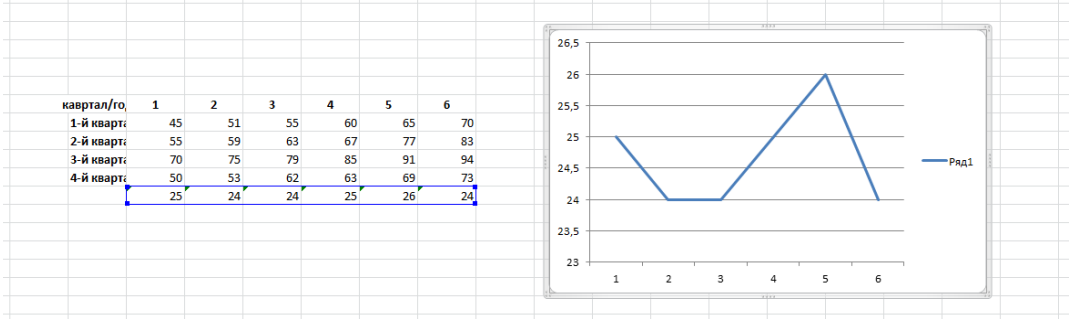


| квартал/год | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1-й квартал | 45 | 51 | 55 | 60 | 65 | 70 |
| 2-й квартал | 55 | 59 | 63 | 67 | 77 | 83 |
| 3-й квартал | 70 | 75 | 79 | 85 | 91 | 94 |
| 4-й квартал | 50 | 53 | 62 | 63 | 69 | 73 |

=МАКС(D30:D33)-МИН(D30:D33)

Рис. 7.10

В результате получим.




| квартал/го | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1-й кварта | 45 | 51 | 55 | 60 | 65 | 70 |
| 2-й кварта | 55 | 59 | 63 | 67 | 77 | 83 |
| 3-й кварта | 70 | 75 | 79 | 85 | 91 | 94 |
| 4-й кварта | 50 | 53 | 62 | 63 | 69 | 73 |
| | 25 | 24 | 24 | 25 | 26 | 24 |

Рис. 7.11

Проводя анализ графика временного ряда и графика разностей между максимальными и минимальными значениями за каждый год, делаем вывод о том, что модель временного ряда аддитивная, т. е. представима в виде

$$Y = T + S + E,$$

где $T$ – трендовая компонента;

$S$ – сезонная компонента;

$E$ – случайная компонента.

Перейдем к дальнейшим вычислениям. Рассчитаем скользящую среднюю. Обратите внимание, что значения ячеек, содержащие данные о количестве краж в 1 и 5 кварталах, поделены на 2.

| | A | B | C | D |
|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | |
| 2 | 1 | 45 | | |
| 3 | 2 | 55 | | |
| 4 | 3 | | =(B2/2+B3+B4+B5+B6/2)/4 | |
| 5 | 4 | 50 | | |
| 6 | 5 | 51 | | |
| 7 | 6 | 59 | | |
| 8 | 7 | 75 | | |
| 9 | 8 | 53 | | |
| 10 | 9 | 55 | | |
| 11 | 10 | 63 | | |
| 12 | 11 | 79 | | |
| 13 | 12 | 62 | | |
| 14 | 13 | 60 | | |
| 15 | 14 | 67 | | |
| 16 | 15 | 85 | | |
| 17 | 16 | 63 | | |
| 18 | 17 | 65 | | |
| 19 | 18 | 77 | | |
| 20 | 19 | 91 | | |
| 21 | 20 | 69 | | |
| 22 | 21 | 70 | | |
| 23 | 22 | 83 | | |
| 24 | 23 | 94 | | |
| 25 | 24 | 73 | | |
| 26 | | | | |

| | A | B | C | D |
|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | |
| 2 | 1 | 45 | | |
| 3 | 2 | 55 | | |
| 4 | 3 | 70 | 55,75 | |
| 5 | 4 | 50 | | |
| 6 | 5 | 51 | | |
| 7 | 6 | 59 | | |
| 8 | 7 | 75 | | |
| 9 | 8 | 53 | | |
| 10 | 9 | 55 | | |
| 11 | 10 | 63 | | |
| 12 | 11 | 79 | | |
| 13 | 12 | 62 | | |
| 14 | 13 | 60 | | |
| 15 | 14 | 67 | | |
| 16 | 15 | 85 | | |
| 17 | 16 | 63 | | |
| 18 | 17 | 65 | | |
| 19 | 18 | 77 | | |
| 20 | 19 | 91 | | |
| 21 | 20 | 69 | | |
| 22 | 21 | 70 | | |
| 23 | 22 | 83 | | |
| 24 | 23 | 94 | | |
| 25 | 24 | 73 | | |
| 26 | | | | |

Рис. 7.12

Проведем вычисления для нужного диапазона: подведем курсор к правому нижнему краю выделенной ячейки и растянем на весь вычисляемый ряд.

Расчетные значения, по очевидным причинам, в две первые и в две последние ячейки не вводятся (при расчете в 1, 2, 23 и 24 кварталах вычисления по данной формуле не дадут правдивого результата, значения данных ячеек определим ниже) (рис. 7.13).

Проведем оценку сезонной компоненты («**Количество краж**» минус «**Скользящая средняя**»). Продолжим вычисления на весь диапазон.

➢ У к а з а н и е : Метод скользящей средней можно также реализовать с помощью встроенного инструмента табличного процессора «*Данные*» → «*Анализ данных*» → «*Скользящее среднее*». Но стоит отметить, что у метода есть различные варианты, что может сказаться на небольшой разнице в

итоговых результатах. Выбор наиболее подходящего варианта метода в каждом случае индивидуальный (рис. 7.14).

| | A | B | C | D |
|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | |
| 2 | 1 | 45 | | |
| 3 | 2 | 55 | | |
| 4 | 3 | 70 | 55,75 | |
| 5 | 4 | 50 | 57 | |
| 6 | 5 | 51 | 58,125 | |
| 7 | 6 | 59 | 59,125 | |
| 8 | 7 | 75 | 60 | |
| 9 | 8 | 53 | 61 | |
| 10 | 9 | 55 | 62 | |
| 11 | 10 | 63 | 63,625 | |
| 12 | 11 | 79 | 65,375 | |
| 13 | 12 | 62 | 66,5 | |
| 14 | 13 | 60 | 67,75 | |
| 15 | 14 | 67 | 68,625 | |
| 16 | 15 | 85 | 69,375 | |
| 17 | 16 | 63 | 71,25 | |
| 18 | 17 | 65 | 73,25 | |
| 19 | 18 | 77 | 74,75 | |
| 20 | 19 | 91 | 76,125 | |
| 21 | 20 | 69 | 77,5 | |
| 22 | 21 | 70 | 78,625 | |
| 23 | 22 | 83 | 79,5 | |
| 24 | 23 | 94 | | |
| 25 | 24 | 73 | | |
| 26 | | | | |

Рис. 7.13

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | |
| 2 | 1 | 45 | | | |
| 3 | 2 | 55 | | | |
| 4 | 3 | 70 | 55,75 | =B4-C4 | |
| 5 | 4 | 50 | 57 | | |
| 6 | 5 | 51 | 58,125 | | |
| 7 | 6 | 59 | 59,125 | | |
| 8 | 7 | 75 | 60 | | |
| 9 | 8 | 53 | 61 | | |
| 10 | 9 | 55 | 62 | | |
| 11 | 10 | 63 | 63,625 | | |
| 12 | 11 | 79 | 65,375 | | |
| 13 | 12 | 62 | 66,5 | | |
| 14 | 13 | 60 | 67,75 | | |
| 15 | 14 | 67 | 68,625 | | |
| 16 | 15 | 85 | 69,375 | | |
| 17 | 16 | 63 | 71,25 | | |
| 18 | 17 | 65 | 73,25 | | |
| 19 | 18 | 77 | 74,75 | | |
| 20 | 19 | 91 | 76,125 | | |
| 21 | 20 | 69 | 77,5 | | |
| 22 | 21 | 70 | 78,625 | | |
| 23 | 22 | 83 | 79,5 | | |
| 24 | 23 | 94 | | | |
| 25 | 24 | 73 | | | |
| 26 | | | | | |

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | |
| 2 | 1 | 45 | | | |
| 3 | 2 | 55 | | | |
| 4 | 3 | 70 | 55,75 | 14,25 | |
| 5 | 4 | 50 | 57 | -7 | |
| 6 | 5 | 51 | 58,125 | -7,125 | |
| 7 | 6 | 59 | 59,125 | -0,125 | |
| 8 | 7 | 75 | 60 | 15 | |
| 9 | 8 | 53 | 61 | -8 | |
| 10 | 9 | 55 | 62 | -7 | |
| 11 | 10 | 63 | 63,625 | -0,625 | |
| 12 | 11 | 79 | 65,375 | 13,625 | |
| 13 | 12 | 62 | 66,5 | -4,5 | |
| 14 | 13 | 60 | 67,75 | -7,75 | |
| 15 | 14 | 67 | 68,625 | -1,625 | |
| 16 | 15 | 85 | 69,375 | 15,625 | |
| 17 | 16 | 63 | 71,25 | -8,25 | |
| 18 | 17 | 65 | 73,25 | -8,25 | |
| 19 | 18 | 77 | 74,75 | 2,25 | |
| 20 | 19 | 91 | 76,125 | 14,875 | |
| 21 | 20 | 69 | 77,5 | -8,5 | |
| 22 | 21 | 70 | 78,625 | -8,625 | |
| 23 | 22 | 83 | 79,5 | 3,5 | |
| 24 | 23 | 94 | | | |
| 25 | 24 | 73 | | | |
| 26 | | | | | |

Рис. 7.14

Далее определим сезонную компоненту. Для этого данные из столбца «**Оценка сезонной компоненты**» в любом свободном месте листа вставим таким образом, чтобы получилась таблица. Скопированные данные вставим (нажав правую кнопку мыши) через специальную вставку как значения.

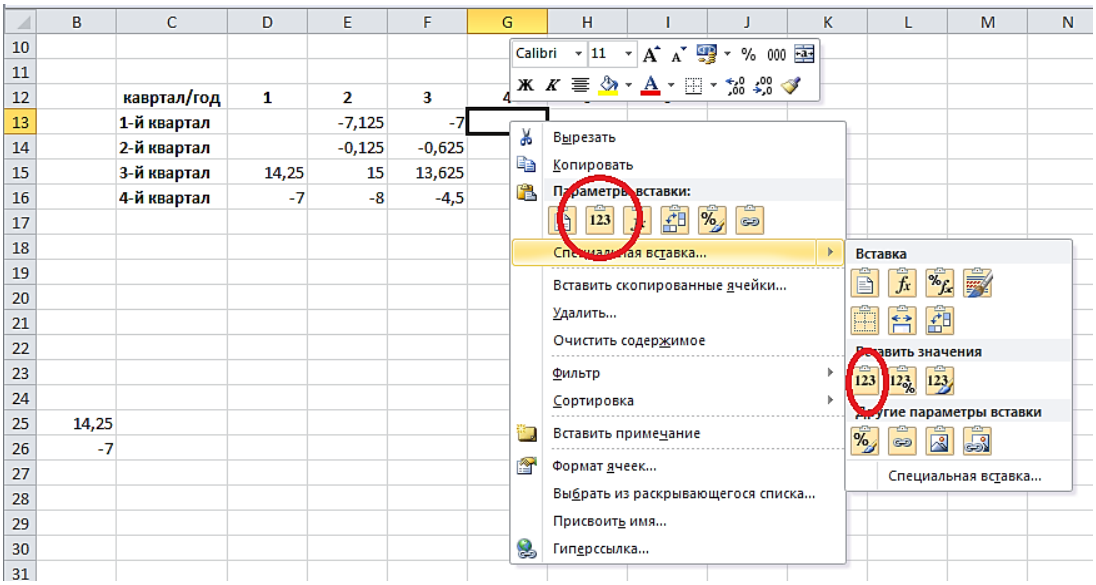



Рис. 7.15

В получившейся таблице вычислим среднее по каждому кварталу за 6 лет, используя вставку функции «*СРЗНАЧ*».

| | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | | |
| 11 | | | | | | | | | | |
| 12 | | **кварталгод** | **1** | **2** | **3** | **4** | **5** | **6** | **Среднее за квартал** | |
| 13 | | **1-й квартал** | | -7,125 | -7 | -7,75 | -8,25 | -8,625 | =СРЗНАЧ(E13:I13) | |
| 14 | | **2-й квартал** | | -0,125 | -0,625 | -1,625 | 2,25 | 3,5 | | |
| 15 | | **3-й квартал** | 14,25 | 15 | 13,625 | 15,625 | 14,875 | | | |
| 16 | | **4-й квартал** | -7 | -8 | -4,5 | -8,25 | -8,5 | | | |
| 17 | | | | | | | | | | |

Рис. 7.16

Значения первого и второго кварталов первого года в вычислениях не учитываются так же, как и значения третьего и четвертого кварталов 6 года. В результате получаем. Найдем сумму средних за кварталы.

| | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | | |
| 11 | | | | | | | | | | |
| 12 | | **кварталгод** | **1** | **2** | **3** | **4** | **5** | **6** | **Среднее за квартал** | |
| 13 | | **1-й квартал** | | -7,125 | -7 | -7,75 | -8,25 | -8,625 | -7,75 | |
| 14 | | **2-й квартал** | | -0,125 | -0,625 | -1,625 | 2,25 | 3,5 | 0,675 | |
| 15 | | **3-й квартал** | 14,25 | 15 | 13,625 | 15,625 | 14,875 | | 14,675 | |
| 16 | | **4-й квартал** | -7 | -8 | -4,5 | -8,25 | -8,5 | | -7,25 | |
| 17 | | | | | | | | **Сумма:** | =СУММ(J13:J16) | |
| 18 | | | | | | | | | | |

Рис. 7.17

Так как сумма не равна нулю (=0,35), то определим поправочный коэффициент как среднее от среднего за квартал.

| | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | | |
| 11 | | | | | | | | | | |
| 12 | | **кавртал/год** | **1** | **2** | **3** | **4** | **5** | **6** | **Среднее за квартал** | |
| 13 | | **1-й квартал** | | -7,125 | -7 | -7,75 | -8,25 | -8,625 | -7,75 | |
| 14 | | **2-й квартал** | | -0,125 | -0,625 | -1,625 | 2,25 | 3,5 | 0,675 | |
| 15 | | **3-й квартал** | 14,25 | 15 | 13,625 | 15,625 | 14,875 | | 14,675 | |
| 16 | | **4-й квартал** | -7 | -8 | -4,5 | -8,25 | -8,5 | | -7,25 | |
| 17 | | | | | | | | **Сумма:** | 0,35 | |
| 18 | | | | | | | | **Поправка:** | =СРЗНАЧ(J13:J16) | |
| 19 | | | | | | | | | | |

Рис. 7.18

Получаем поправочный коэффициент равный 0,0875.

Далее найдем исправленное среднее («**Среднее за квартал**» минус «**Поправка**»).

И посчитаем данное значение для каждого квартала. Букву номера ячейки с поправкой ограничим знаком $ для закрепления значений данной ячейки. Найдем оставшиеся значения исправленной средней. Подведем курсор к правому нижнему краю ячейки и проведем вычисления для нужного диапазона.

| | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | | | |
| 11 | | | | | | | | | | | |
| 12 | | **кавртал/год** | **1** | **2** | **3** | **4** | **5** | **6** | **Среднее за квартал** | **Исправленное среднее** | |
| 13 | | **1-й квартал** | | -7,125 | -7 | -7,75 | -8,25 | -8,625 | -7,75 | -7,8375 | |
| 14 | | **2-й квартал** | | -0,125 | -0,625 | -1,625 | 2,25 | 3,5 | 0,675 | 0,5875 | |
| 15 | | **3-й квартал** | 14,25 | 15 | 13,625 | 15,625 | 14,875 | | 14,675 | 14,5875 | |
| 16 | | **4-й квартал** | -7 | -8 | -4,5 | -8,25 | -8,5 | | -7,25 | -7,3375 | |
| 17 | | | | | | | | **Сумма:** | 0,35 | | |
| 18 | | | | | | | | **Поправка:** | 0,0875 | | |
| 19 | | | | | | | | | | | |

Рис. 7.19

Для проверки правильности наших вычислений найдем сумму исправленного среднего, которая должна равняться нулю.

| | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | | | | | | | | | | | |
| 11 | | | | | | | | | | | |
| 12 | | **кавртал/год** | **1** | **2** | **3** | **4** | **5** | **6** | **Среднее за квартал** | **Исправленное среднее** | |
| 13 | | **1-й квартал** | | -7,125 | -7 | -7,75 | -8,25 | -8,625 | -7,75 | -7,8375 | |
| 14 | | **2-й квартал** | | -0,125 | -0,625 | -1,625 | 2,25 | 3,5 | 0,675 | 0,5875 | |
| 15 | | **3-й квартал** | 14,25 | 15 | 13,625 | 15,625 | 14,875 | | 14,675 | 14,5875 | |
| 16 | | **4-й квартал** | -7 | -8 | -4,5 | -8,25 | -8,5 | | -7,25 | -7,3375 | |
| 17 | | | | | | | | **Сумма:** | 0,35 | =СУММ(K13:K16) | |
| 18 | | | | | | | | **Поправка:** | 0,0875 | | |
| 19 | | | | | | | | | | | |

Рис. 7.20

Значения исправленного среднего по каждому кварталу вносим в столбец сезонной компоненты общих вычислений (предлагается после

копирования значений исправленного среднего использовать специальную вставку как значение). В результате получим.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | |
| 2 | 1 | 45 | | | -7,8375 | |
| 3 | 2 | 55 | | | 0,5875 | |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | |
| 24 | 23 | 94 | | | 14,5875 | |
| 25 | 24 | 73 | | | -7,3375 | |
| 26 | | | | | | |

Рис. 7.21

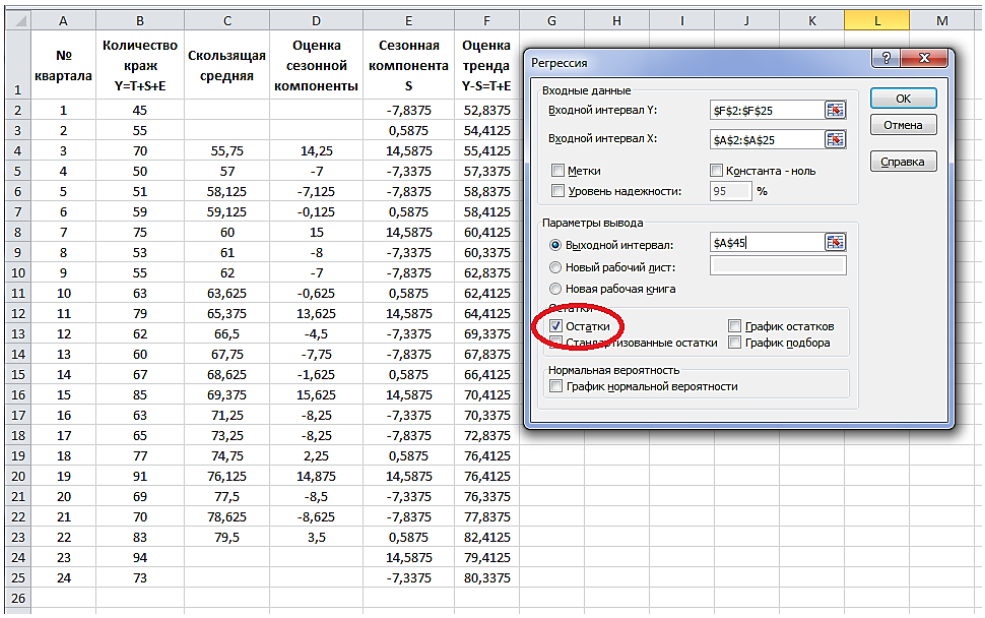
Проведем оценку тренда $Y - S = T + E$ («Количество краж» минус «Сезонная компонента»). Вычислим данное значение для каждого квартала.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 |
| 26 | | | | | | |

Рис. 7.22

Определим тренд. Для этого проведем регрессионный анализ между оценкой тренда и порядковым номером квартала.

Выберем «*Данные*» → «*Анализ данных*» → «*Регрессия*». В поле «*Входной интервал Y:*» введем номера ячеек «**Оценка тренда**». В поле «*Входной интервал X:*» введем номера ячеек «**№ квартала**». В поле «*Выходной интервал:*» выбираем удобную для размещения результатов расчета ячейку. «*Остатки*» → поставим галочку.

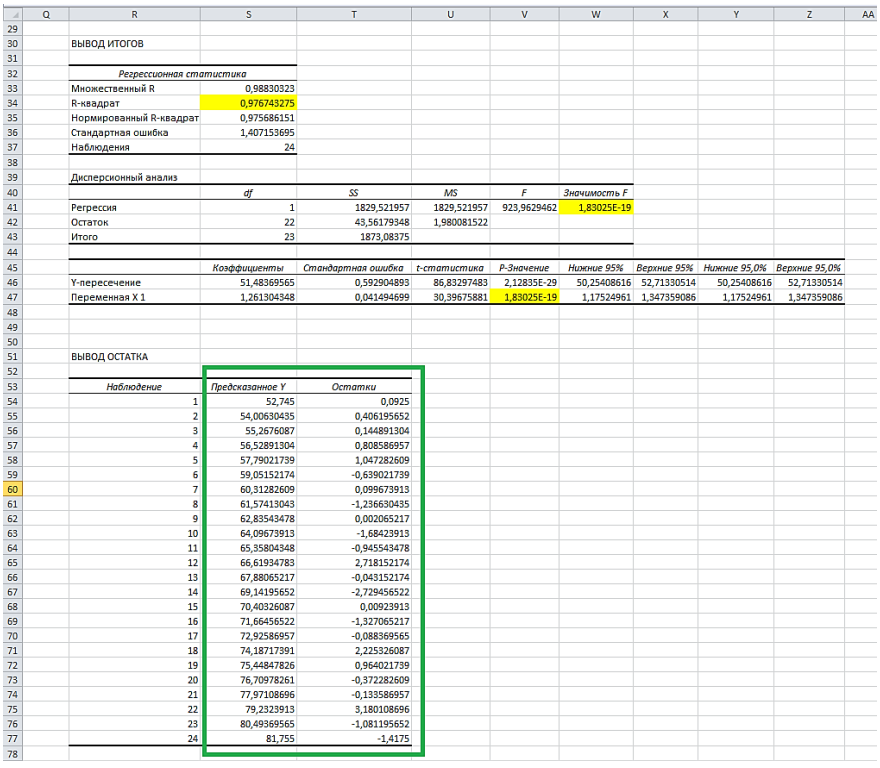


| № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E |
|---|---|---|---|---|---|
| 1 | 45 | | | -7,8375 | 52,8375 |
| 2 | 55 | | | 0,5875 | 54,4125 |
| 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 |
| 4 | 50 | 57 | -7 | -7,3375 | 57,3375 |
| 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 |
| 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 |
| 7 | 75 | 60 | 15 | 14,5875 | 60,4125 |
| 8 | 53 | 61 | -8 | -7,3375 | 60,3375 |
| 9 | 55 | 62 | -7 | -7,8375 | 62,8375 |
| 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 |
| 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 |
| 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 |
| 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 |
| 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 |
| 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 |
| 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 |
| 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 |
| 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 |
| 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 |
| 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 |
| 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 |
| 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 |
| 23 | 94 | | | 14,5875 | 79,4125 |
| 24 | 73 | | | -7,3375 | 80,3375 |



Рис. 7.23

В результате получаем.



ВЫВОД ИТОГОВ

| *Регрессионная статистика* | |
|---|---|
| Множественный R | 0,98830323 |
| R-квадрат | 0,976743275 |
| Нормированный R-квадрат | 0,975686151 |
| Стандартная ошибка | 1,407153695 |
| Наблюдения | 24 |

Дисперсионный анализ

| | *df* | *SS* | *MS* | *F* | *Значимость F* |
|---|---|---|---|---|---|
| Регрессия | 1 | 1829,521957 | 1829,521957 | 923,9629462 | 1,83025E-19 |
| Остаток | 22 | 43,56179348 | 1,980081522 | | |
| Итого | 23 | 1873,08375 | | | |

| | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
|---|---|---|---|---|---|---|---|---|
| Y-пересечение | 51,48369565 | 0,592904893 | 86,83297483 | 2,12835E-29 | 50,25408616 | 52,71330514 | 50,25408616 | 52,71330514 |
| Переменная X 1 | 1,261304348 | 0,041494699 | 30,39675881 | 1,83025E-19 | 1,17524961 | 1,347359086 | 1,17524961 | 1,347359086 |

ВЫВОД ОСТАТКА

| *Наблюдение* | *Предсказанное Y* | *Остатки* |
|---|---|---|
| 1 | 52,745 | 0,0925 |
| 2 | 54,00630435 | 0,406195652 |
| 3 | 55,2676087 | 0,144891304 |
| 4 | 56,52891304 | 0,808586957 |
| 5 | 57,79021739 | 1,047282609 |
| 6 | 59,05152174 | -0,639021739 |
| 7 | 60,31282609 | 0,099673913 |
| 8 | 61,57413043 | -1,236630435 |
| 9 | 62,83543478 | 0,002065217 |
| 10 | 64,09673913 | -1,68423913 |
| 11 | 65,35804348 | -0,945543478 |
| 12 | 66,61934783 | 2,718152174 |
| 13 | 67,88065217 | -0,043152174 |
| 14 | 69,14195652 | -2,729456522 |
| 15 | 70,40326087 | 0,00923913 |
| 16 | 71,66456522 | -1,327065217 |
| 17 | 72,92586957 | -0,088369565 |
| 18 | 74,18717391 | 2,225326087 |
| 19 | 75,44847826 | 0,964021739 |
| 20 | 76,70978261 | -0,372282609 |
| 21 | 77,97108696 | -0,133586957 |
| 22 | 79,2323913 | 3,180108696 |
| 23 | 80,49369565 | -1,081195652 |
| 24 | 81,755 | -1,4175 |

Рис. 7.24

➢Указание: Определить тренд можно также следующими способами.

1) Добавив на графике линию тренда и показав уравнение на диаграмме.

2) Воспользоваться функцией «*ТЕНДЕНЦИЯ*», которая возвращает значение в соответствии с линейной аппроксимацией по методу наименьших квадратов. Рекомендуется использовать, если применялся метод скользящего среднего.

3) Воспользоваться функцией «*РОСТ*», которая возвращает значение в соответствии с экспоненциальным трендом. Рекомендуется использовать, если применялось экспоненциальное сглаживание.

Проверяем полученные результаты на правильность вычислений. Значение «*Множественный R*» должно быть близким к 1. Значение «*Значимость F*» дисперсионного анализа должно быть меньше 0,05, что было определено выше и указывает на величину ошибки. «*P-Значение*» переменной «*Переменная X1*» – меньше 0,05. Таким образом, полученным результатам можно доверять.

Выделенные значения «*Предсказанное Y*» является трендом, а «*Остатки*» – случайной компонентой. Копируя данные значения, вставляем их в исходную таблицу.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,0063 | 0,406195652 | |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,26761 | 0,144891304 | |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,52891 | 0,808586957 | |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79022 | 1,047282609 | |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,05152 | -0,639021739 | |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,31283 | 0,099673913 | |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,57413 | -1,236630435 | |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,83543 | 0,002065217 | |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,09674 | -1,68423913 | |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,35804 | -0,945543478 | |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,61935 | 2,718152174 | |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,88065 | -0,043152174 | |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,14196 | -2,729456522 | |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,40326 | 0,00923913 | |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,66457 | -1,327065217 | |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,92587 | -0,088369565 | |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,18717 | 2,225326087 | |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,44848 | 0,964021739 | |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,70978 | -0,372282609 | |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,97109 | -0,133586957 | |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,23239 | 3,180108696 | |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,4937 | -1,081195652 | |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | |
| 26 | | | | | | | | | |

Рис. 7.25

Сделаем проверку вычислений. Сумма «**Сезонной компоненты**», «**Тренда**» и «**Случайной компоненты**» в каждом квартале должны дать значение «**Количества краж**».

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E | |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | =G2+E2+H2 | |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,0063 | 0,406195652 | | |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,26761 | 0,144891304 | | |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,52891 | 0,808586957 | | |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79022 | 1,047282609 | | |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,05152 | -0,639021739 | | |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,31283 | 0,099673913 | | |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,57413 | -1,236630435 | | |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,83543 | 0,002065217 | | |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,09674 | -1,68423913 | | |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,35804 | -0,945543478 | | |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,61935 | 2,718152174 | | |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,88065 | -0,043152174 | | |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,14196 | -2,729456522 | | |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,40326 | 0,00923913 | | |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,66457 | -1,327065217 | | |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,92587 | -0,088369565 | | |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,18717 | 2,225326087 | | |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,44848 | 0,964021739 | | |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,70978 | -0,372282609 | | |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,97109 | -0,133586957 | | |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,23239 | 3,180108696 | | |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,4937 | -1,081195652 | | |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | | |
| 26 | | | | | | | | | | |

Рис. 7.26

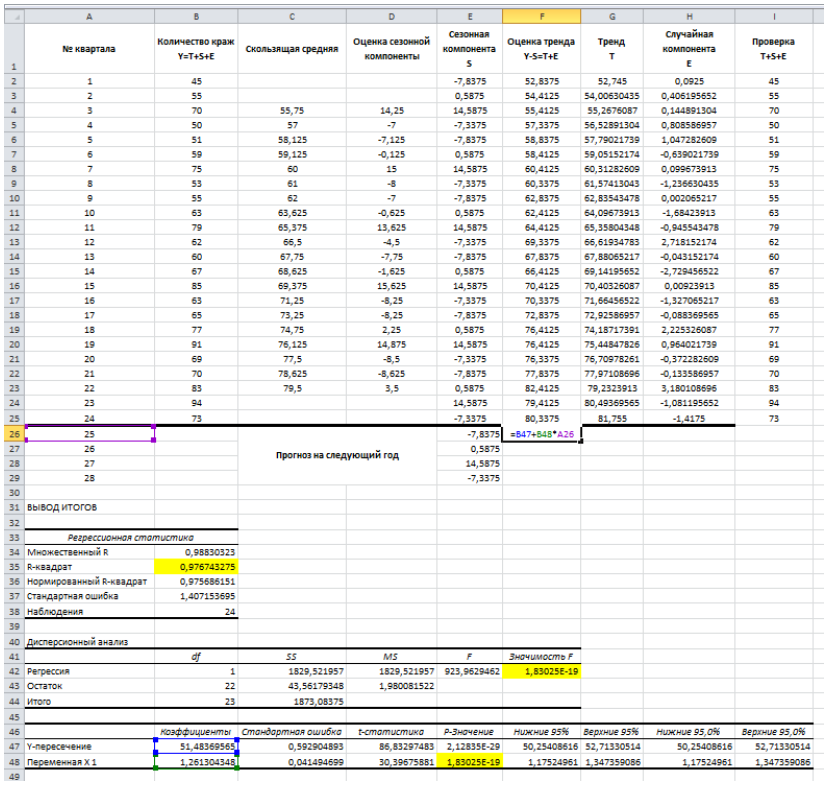
Получаем те же значения.

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E | |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | 45 | |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,0063 | 0,406195652 | 55 | |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,26761 | 0,144891304 | 70 | |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,52891 | 0,808586957 | 50 | |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79022 | 1,047282609 | 51 | |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,05152 | -0,639021739 | 59 | |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,31283 | 0,099673913 | 75 | |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,57413 | -1,236630435 | 53 | |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,83543 | 0,002065217 | 55 | |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,09674 | -1,68423913 | 63 | |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,35804 | -0,945543478 | 79 | |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,61935 | 2,718152174 | 62 | |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,88065 | -0,043152174 | 60 | |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,14196 | -2,729456522 | 67 | |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,40326 | 0,00923913 | 85 | |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,66457 | -1,327065217 | 63 | |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,92587 | -0,088369565 | 65 | |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,18717 | 2,225326087 | 77 | |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,44848 | 0,964021739 | 91 | |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,70978 | -0,372282609 | 69 | |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,97109 | -0,133586957 | 70 | |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,23239 | 3,180108696 | 83 | |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,4937 | -1,081195652 | 94 | |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | 73 | |
| 26 | | | | | | | | | | |

Рис. 7.27

Приступим к прогнозу на следующий за отчетным периодом год. Для этого скопируем значения сезонной компоненты, которая повторяется из года в год в 25, 26, 27 и 28 кварталы.

Рассчитаем тренд в исследуемом году, используя уравнение регрессии. Буквы номеров ячеек со значениями коэффициентов заключим знаком $ для неизменности при копировании формулы.



| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | 45 |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,00630435 | 0,406195652 | 55 |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,2676087 | 0,144891304 | 70 |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,52891304 | 0,808586957 | 50 |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79021739 | 1,047282609 | 51 |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,05152174 | -0,639021739 | 59 |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,31282609 | 0,099673913 | 75 |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,57413043 | -1,236630435 | 53 |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,83543478 | 0,002065217 | 55 |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,09673913 | -1,68423913 | 63 |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,35804348 | -0,945543478 | 79 |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,61934783 | 2,718152174 | 62 |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,88065217 | -0,043152174 | 60 |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,14195652 | -2,729456522 | 67 |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,40326087 | 0,00923913 | 85 |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,66456522 | -1,327065217 | 63 |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,92586957 | -0,088369565 | 65 |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,18717391 | 2,225326087 | 77 |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,44847826 | 0,964021739 | 91 |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,70978261 | -0,372282609 | 69 |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,97108696 | -0,133586957 | 70 |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,2323913 | 3,180108696 | 83 |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,49369565 | -1,081195652 | 94 |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | 73 |
| 26 | 25 | | | | -7,8375 | =B47+B48*A26 | | | |
| 27 | 26 | | Прогноз на следующий год | | 0,5875 | | | | |
| 28 | 27 | | | | 14,5875 | | | | |
| 29 | 28 | | | | -7,3375 | | | | |
| 30 | | | | | | | | | |
| 31 | ВЫВОД ИТОГОВ | | | | | | | | |
| 32 | | | | | | | | | |
| 33 | *Регрессионная статистика* | | | | | | | | |
| 34 | Множественный R | 0,9883023 | | | | | | | |
| 35 | R-квадрат | 0,976743275 | | | | | | | |
| 36 | Нормированный R-квадрат | 0,975686151 | | | | | | | |
| 37 | Стандартная ошибка | 1,407153695 | | | | | | | |
| 38 | Наблюдения | 24 | | | | | | | |
| 39 | | | | | | | | | |
| 40 | Дисперсионный анализ | | | | | | | | |
| 41 | | *df* | *SS* | *MS* | *F* | *Значимость F* | | | |
| 42 | Регрессия | 1 | 1829,521957 | 1829,521957 | 923,9629462 | 1,83025E-19 | | | |
| 43 | Остаток | 22 | 43,56179348 | 1,980081522 | | | | | |
| 44 | Итого | 23 | 1873,08375 | | | | | | |
| 45 | | | | | | | | | |
| 46 | | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
| 47 | Y-пересечение | 51,48369565 | 0,592904893 | 86,83297483 | 2,12835E-29 | 50,25408616 | 52,71330514 | 50,25408616 | 52,71330514 |
| 48 | Переменная X 1 | 1,261304348 | 0,041494699 | 30,39675881 | 1,83025E-19 | 1,17524961 | 1,347359086 | 1,17524961 | 1,347359086 |
| 49 | | | | | | | | | |

Рис. 7.28

В результате получим.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | 45 |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,006 | 0,406195652 | 55 |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,268 | 0,144891304 | 70 |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,529 | 0,808586957 | 50 |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79 | 1,047282609 | 51 |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,052 | -0,639021739 | 59 |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,313 | 0,099673913 | 75 |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,574 | -1,236630435 | 53 |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,835 | 0,002065217 | 55 |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,097 | -1,68423913 | 63 |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,358 | -0,945543478 | 79 |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,619 | 2,718152174 | 62 |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,881 | -0,043152174 | 60 |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,142 | -2,729456522 | 67 |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,403 | 0,00923913 | 85 |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,665 | -1,327065217 | 63 |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,926 | -0,088369565 | 65 |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,187 | 2,225326087 | 77 |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,448 | 0,964021739 | 91 |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,71 | -0,372282609 | 69 |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,971 | -0,133586957 | 70 |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,232 | 3,180108696 | 83 |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,494 | -1,081195652 | 94 |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | 73 |
| 26 | 25 | | | | -7,8375 | 83,0163043 | | | |
| 27 | 26 | | Прогноз на следующий год | | 0,5875 | 84,2776087 | | | |
| 28 | 27 | | | | 14,5875 | 85,538913 | | | |
| 29 | 28 | | | | -7,3375 | 86,8002174 | | | |
| 30 | | | | | | | | | |

Рис. 7.29

Делаем прогноз по количеству краж в исследуемом году. Данные значения определяем как сумму «Сезонной компоненты» и «Оценки тренда».

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | 45 |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,006 | 0,406195652 | 55 |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,268 | 0,144891304 | 70 |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,529 | 0,808586957 | 50 |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79 | 1,047282609 | 51 |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,052 | -0,639021739 | 59 |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,313 | 0,099673913 | 75 |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,574 | -1,236630435 | 53 |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,835 | 0,002065217 | 55 |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,097 | -1,68423913 | 63 |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,358 | -0,945543478 | 79 |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,619 | 2,718152174 | 62 |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,881 | -0,043152174 | 60 |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,142 | -2,729456522 | 67 |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,403 | 0,00923913 | 85 |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,665 | -1,327065217 | 63 |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,926 | -0,088369565 | 65 |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,187 | 2,225326087 | 77 |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,448 | 0,964021739 | 91 |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,71 | -0,372282609 | 69 |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,971 | -0,133586957 | 70 |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,232 | 3,180108696 | 83 |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,494 | -1,081195652 | 94 |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | 73 |
| 26 | 25 | =E26+F26 | | | -7,8375 | 83,0163043 | | | |
| 27 | 26 | | Прогноз на следующий год | | 0,5875 | 84,2776087 | | | |
| 28 | 27 | | | | 14,5875 | 85,538913 | | | |
| 29 | 28 | | | | -7,3375 | 86,8002174 | | | |
| 30 | | | | | | | | | |

Рис. 7.30

И для всего года имеем следующие значения.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | 45 |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,006 | 0,406195652 | 55 |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,268 | 0,144891304 | 70 |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,529 | 0,808586957 | 50 |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79 | 1,047282609 | 51 |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,052 | -0,639021739 | 59 |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,313 | 0,099673913 | 75 |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,574 | -1,236630435 | 53 |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,835 | 0,002065217 | 55 |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,097 | -1,68423913 | 63 |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,358 | -0,945543478 | 79 |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,619 | 2,718152174 | 62 |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,881 | -0,043152174 | 60 |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,142 | -2,729456522 | 67 |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,403 | 0,00923913 | 85 |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,665 | -1,327065217 | 63 |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,926 | -0,088369565 | 65 |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,187 | 2,225326087 | 77 |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,448 | 0,964021739 | 91 |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,71 | -0,372282609 | 69 |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,971 | -0,133586957 | 70 |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,232 | 3,180108696 | 83 |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,494 | -1,081195652 | 94 |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | 73 |
| 26 | 25 | 75,17880435 | | | -7,8375 | 83,0163043 | | | |
| 27 | 26 | 84,8651087 | Прогноз на следующий год | | 0,5875 | 84,2776087 | | | |
| 28 | 27 | 100,126413 | | | 14,5875 | 85,538913 | | | |
| 29 | 28 | 79,46271739 | | | -7,3375 | 86,8002174 | | | |
| 30 | | | | | | | | | |

Рис. 7.31

Округлим полученные значения до целого. Окончательно получаем.

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | № квартала | Количество краж Y=T+S+E | Скользящая средняя | Оценка сезонной компоненты | Сезонная компонента S | Оценка тренда Y-S=T+E | Тренд T | Случайная компонента E | Проверка T+S+E |
| 2 | 1 | 45 | | | -7,8375 | 52,8375 | 52,745 | 0,0925 | 45 |
| 3 | 2 | 55 | | | 0,5875 | 54,4125 | 54,006 | 0,406195652 | 55 |
| 4 | 3 | 70 | 55,75 | 14,25 | 14,5875 | 55,4125 | 55,268 | 0,144891304 | 70 |
| 5 | 4 | 50 | 57 | -7 | -7,3375 | 57,3375 | 56,529 | 0,808586957 | 50 |
| 6 | 5 | 51 | 58,125 | -7,125 | -7,8375 | 58,8375 | 57,79 | 1,047282609 | 51 |
| 7 | 6 | 59 | 59,125 | -0,125 | 0,5875 | 58,4125 | 59,052 | -0,639021739 | 59 |
| 8 | 7 | 75 | 60 | 15 | 14,5875 | 60,4125 | 60,313 | 0,099673913 | 75 |
| 9 | 8 | 53 | 61 | -8 | -7,3375 | 60,3375 | 61,574 | -1,236630435 | 53 |
| 10 | 9 | 55 | 62 | -7 | -7,8375 | 62,8375 | 62,835 | 0,002065217 | 55 |
| 11 | 10 | 63 | 63,625 | -0,625 | 0,5875 | 62,4125 | 64,097 | -1,68423913 | 63 |
| 12 | 11 | 79 | 65,375 | 13,625 | 14,5875 | 64,4125 | 65,358 | -0,945543478 | 79 |
| 13 | 12 | 62 | 66,5 | -4,5 | -7,3375 | 69,3375 | 66,619 | 2,718152174 | 62 |
| 14 | 13 | 60 | 67,75 | -7,75 | -7,8375 | 67,8375 | 67,881 | -0,043152174 | 60 |
| 15 | 14 | 67 | 68,625 | -1,625 | 0,5875 | 66,4125 | 69,142 | -2,729456522 | 67 |
| 16 | 15 | 85 | 69,375 | 15,625 | 14,5875 | 70,4125 | 70,403 | 0,00923913 | 85 |
| 17 | 16 | 63 | 71,25 | -8,25 | -7,3375 | 70,3375 | 71,665 | -1,327065217 | 63 |
| 18 | 17 | 65 | 73,25 | -8,25 | -7,8375 | 72,8375 | 72,926 | -0,088369565 | 65 |
| 19 | 18 | 77 | 74,75 | 2,25 | 0,5875 | 76,4125 | 74,187 | 2,225326087 | 77 |
| 20 | 19 | 91 | 76,125 | 14,875 | 14,5875 | 76,4125 | 75,448 | 0,964021739 | 91 |
| 21 | 20 | 69 | 77,5 | -8,5 | -7,3375 | 76,3375 | 76,71 | -0,372282609 | 69 |
| 22 | 21 | 70 | 78,625 | -8,625 | -7,8375 | 77,8375 | 77,971 | -0,133586957 | 70 |
| 23 | 22 | 83 | 79,5 | 3,5 | 0,5875 | 82,4125 | 79,232 | 3,180108696 | 83 |
| 24 | 23 | 94 | | | 14,5875 | 79,4125 | 80,494 | -1,081195652 | 94 |
| 25 | 24 | 73 | | | -7,3375 | 80,3375 | 81,755 | -1,4175 | 73 |
| 26 | 25 | 75 | Прогноз на следующий год | | -7,8375 | 83,0163043 | | | |
| 27 | 26 | 85 | | | 0,5875 | 84,2776087 | | | |
| 28 | 27 | 100 | | | 14,5875 | 85,538913 | | | |
| 29 | 28 | 79 | | | -7,3375 | 86,8002174 | | | |
| 30 | | | | | | | | | |

Рис. 7.32

Согласно прогнозу, в первом, втором, третьем и четвертом квартале следующего за отчетным периодом года будет предположительно совершено 75, 85, 100, 79 краж соответственно.

График распределения количества краж, сезонной компоненты, тренда, случайной компоненты, а также прогноз количества краж на будущий год имеет вид:

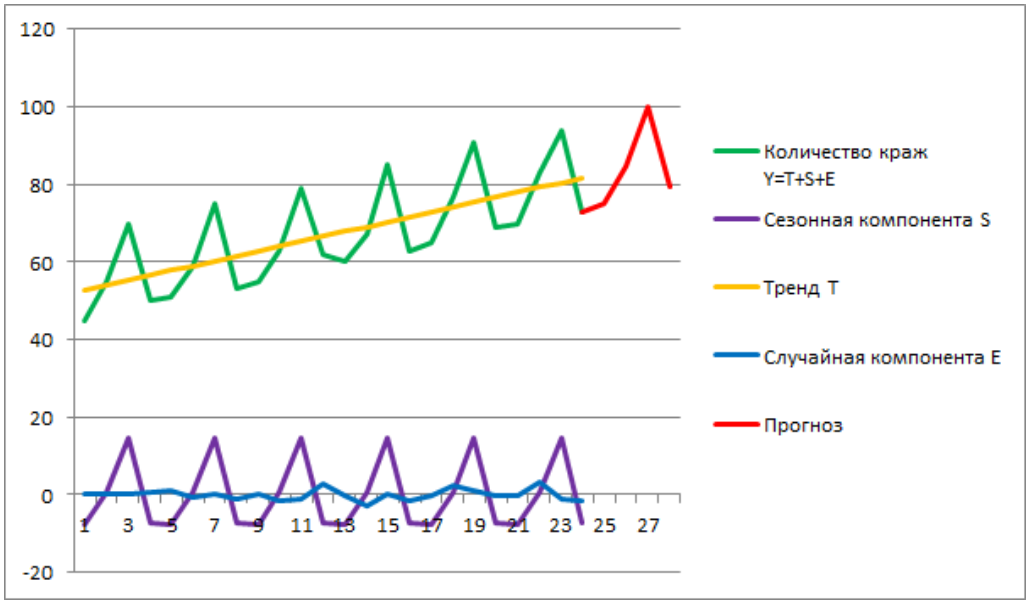



Рис. 7.33

## Задание для самостоятельного выполнения

На основании имеющихся данных сделать прогноз оценки количества преступлений на год, следующий за отчетным периодом, методом скользящей средней.

Вариант 1.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 60 |
| | 2 | 70 |
| | 3 | 85 |
| | 4 | 65 |
| 2 | 5 | 66 |
| | 6 | 74 |
| | 7 | 90 |
| | 8 | 68 |
| 3 | 9 | 70 |
| | 10 | 78 |
| | 11 | 94 |
| | 12 | 77 |
| 4 | 13 | 75 |
| | 14 | 82 |
| | 15 | 100 |
| | 16 | 78 |
| 5 | 17 | 80 |
| | 18 | 92 |
| | 19 | 106 |
| | 20 | 84 |
| 6 | 21 | 85 |
| | 22 | 98 |
| | 23 | 109 |
| | 24 | 88 |
| 7 | 25 | 90 |
| | 26 | 100 |
| | 27 | 115 |
| | 28 | 94 |

Вариант 2.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 84 |
| | 2 | 108 |
| | 3 | 144 |
| | 4 | 60 |
| 2 | 5 | 120 |
| | 6 | 156 |
| | 7 | 204 |
| | 8 | 84 |
| 3 | 9 | 156 |
| | 10 | 192 |
| | 11 | 252 |
| | 12 | 120 |
| 4 | 13 | 192 |
| | 14 | 228 |
| | 15 | 300 |
| | 16 | 144 |
| 5 | 17 | 240 |
| | 18 | 264 |
| | 19 | 348 |
| | 20 | 168 |
| 6 | 21 | 288 |
| | 22 | 324 |
| | 23 | 420 |
| | 24 | 192 |
| 7 | 25 | 317 |
| | 26 | 365 |
| | 27 | 478 |
| | 28 | 211 |

Вариант 3.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 336 |
| | 2 | 735 |
| | 3 | 567 |
| | 4 | 504 |
| 2 | 5 | 294 |
| | 6 | 609 |
| | 7 | 462 |
| | 8 | 420 |
| 3 | 9 | 252 |
| | 10 | 525 |
| | 11 | 399 |
| | 12 | 336 |
| 4 | 13 | 210 |
| | 14 | 441 |
| | 15 | 336 |
| | 16 | 273 |
| 5 | 17 | 147 |
| | 18 | 357 |
| | 19 | 273 |
| | 20 | 210 |
| 6 | 21 | 105 |
| | 22 | 252 |
| | 23 | 189 |
| | 24 | 147 |
| 7 | 25 | 75 |
| | 26 | 150 |
| | 27 | 98 |
| | 28 | 70 |

Вариант 4.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 67 |
| | 2 | 77 |
| | 3 | 92 |
| | 4 | 72 |
| 2 | 5 | 73 |
| | 6 | 81 |
| | 7 | 97 |
| | 8 | 75 |
| 3 | 9 | 77 |
| | 10 | 85 |
| | 11 | 101 |
| | 12 | 84 |
| 4 | 13 | 82 |
| | 14 | 89 |
| | 15 | 107 |
| | 16 | 85 |
| 5 | 17 | 87 |
| | 18 | 99 |
| | 19 | 113 |
| | 20 | 91 |
| 6 | 21 | 92 |
| | 22 | 105 |
| | 23 | 116 |
| | 24 | 95 |
| 7 | 25 | 97 |
| | 26 | 107 |
| | 27 | 122 |
| | 28 | 101 |

Вариант 5.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 85 |
| | 2 | 111 |
| | 3 | 58 |
| | 4 | 70 |
| 2 | 5 | 80 |
| | 6 | 104 |
| | 7 | 56 |
| | 8 | 69 |
| 3 | 9 | 67 |
| | 10 | 95 |
| | 11 | 53 |
| | 12 | 65 |
| 4 | 13 | 63 |
| | 14 | 85 |
| | 15 | 52 |
| | 16 | 59 |
| 5 | 17 | 58 |
| | 18 | 81 |
| | 19 | 51 |
| | 20 | 55 |
| 6 | 21 | 55 |
| | 22 | 77 |
| | 23 | 48 |
| | 24 | 49 |
| 7 | 25 | 51 |
| | 26 | 70 |
| | 27 | 40 |
| | 28 | 46 |

Вариант 6.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 75 |
| | 2 | 96 |
| | 3 | 81 |
| | 4 | 71 |
| 2 | 5 | 69 |
| | 6 | 90 |
| | 7 | 79 |
| | 8 | 66 |
| 3 | 9 | 65 |
| | 10 | 87 |
| | 11 | 73 |
| | 12 | 61 |
| 4 | 13 | 59 |
| | 14 | 81 |
| | 15 | 63 |
| | 16 | 56 |
| 5 | 17 | 58 |
| | 18 | 75 |
| | 19 | 59 |
| | 20 | 51 |
| 6 | 21 | 49 |
| | 22 | 71 |
| | 23 | 55 |
| | 24 | 47 |
| 7 | 25 | 46 |
| | 26 | 66 |
| | 27 | 51 |
| | 28 | 41 |

Вариант 7.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 34 |
| | 2 | 44 |
| | 3 | 59 |
| | 4 | 39 |
| 2 | 5 | 40 |
| | 6 | 48 |
| | 7 | 64 |
| | 8 | 42 |
| 3 | 9 | 44 |
| | 10 | 52 |
| | 11 | 68 |
| | 12 | 51 |
| 4 | 13 | 49 |
| | 14 | 56 |
| | 15 | 74 |
| | 16 | 52 |
| 5 | 17 | 54 |
| | 18 | 66 |
| | 19 | 80 |
| | 20 | 58 |
| 6 | 21 | 59 |
| | 22 | 72 |
| | 23 | 83 |
| | 24 | 62 |
| 7 | 25 | 64 |
| | 26 | 74 |
| | 27 | 89 |
| | 28 | 68 |

Вариант 8.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 83 |
| | 2 | 81 |
| | 3 | 130 |
| | 4 | 93 |
| 2 | 5 | 93 |
| | 6 | 86 |
| | 7 | 137 |
| | 8 | 98 |
| 3 | 9 | 100 |
| | 10 | 88 |
| | 11 | 144 |
| | 12 | 107 |
| 4 | 13 | 110 |
| | 14 | 90 |
| | 15 | 161 |
| | 16 | 114 |
| 5 | 17 | 117 |
| | 18 | 95 |
| | 19 | 176 |
| | 20 | 136 |
| 6 | 21 | 119 |
| | 22 | 98 |
| | 23 | 188 |
| | 24 | 144 |
| 7 | 25 | 132 |
| | 26 | 112 |
| | 27 | 194 |
| | 28 | 142 |

Вариант 9.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 136 |
| | 2 | 157 |
| | 3 | 142 |
| | 4 | 132 |
| 2 | 5 | 130 |
| | 6 | 151 |
| | 7 | 140 |
| | 8 | 127 |
| 3 | 9 | 126 |
| | 10 | 148 |
| | 11 | 134 |
| | 12 | 122 |
| 4 | 13 | 120 |
| | 14 | 142 |
| | 15 | 124 |
| | 16 | 117 |
| 5 | 17 | 119 |
| | 18 | 136 |
| | 19 | 120 |
| | 20 | 112 |
| 6 | 21 | 110 |
| | 22 | 132 |
| | 23 | 116 |
| | 24 | 108 |
| 7 | 25 | 107 |
| | 26 | 127 |
| | 27 | 112 |
| | 28 | 102 |

Вариант 10.

| Год | № квартала | Количество преступлений |
|---|---|---|
| 1 | 1 | 102 |
| | 2 | 100 |
| | 3 | 161 |
| | 4 | 115 |
| 2 | 5 | 115 |
| | 6 | 107 |
| | 7 | 170 |
| | 8 | 121 |
| 3 | 9 | 123 |
| | 10 | 109 |
| | 11 | 178 |
| | 12 | 132 |
| 4 | 13 | 136 |
| | 14 | 111 |
| | 15 | 199 |
| | 16 | 140 |
| 5 | 17 | 144 |
| | 18 | 117 |
| | 19 | 218 |
| | 20 | 168 |
| 6 | 21 | 147 |
| | 22 | 121 |
| | 23 | 233 |
| | 24 | 178 |
| 7 | 25 | 163 |
| | 26 | 138 |
| | 27 | 240 |
| | 28 | 175 |

## Контрольные вопросы

1. Что такое временной ряд?
2. Что такое интервальный временной ряд?
3. Что такое моментальный временной ряд?
4. Что называют уровнем ряда?
5. Что такое абсолютный прирост?
6. Что такое темп роста?
7. Что такое темп прироста?

# 4. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

## Лабораторная работа № 8
## Проверка статистических гипотез

**Цель лабораторной работы:** изучить основные понятия, связанные с проверкой статистических гипотез, рассмотреть решение основных примеров, в том числе в табличных процессорах, и самостоятельно выполнить задания согласно своему варианту.

### Краткие теоретические сведения и рассмотрение примеров

### Общая схема проверки статистических гипотез

Проверка статистических гипотез тесно связана с теорией статистического оценивания, так как часто результаты наблюдений используются для проверки предположений (гипотез) относительно либо самого вида распределения генеральной совокупности, либо значения параметров уже известного распределения – статистических гипотез.

*Определение.* **Статистическая гипотеза** это любое предположение о виде распределения случайной величины или параметрах неизвестного закона распределения.

Пусть известно распределение случайной величины *X*, и по выборке необходимо проверить гипотезу о значении некоторого параметра ($\bar{x}_г$, $D_г$ или $\sigma_г$) этого распределения.

*Определение.* **Выдвигаемая (проверяемая) гипотеза –** это нулевая (основная) гипотеза, которая обозначается $H_0$. Она формулируется так: «нет различий», «=».

*Определение.* **Альтернативная (конкурирующая) гипотеза –** это гипотеза, которая противоречит нулевой, т. е. является логическим отрицанием $H_0$. Она обозначается $H_1$ или $H_a$. Формулируется так: «различия», «≠».

Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез.

Выдвинутая гипотеза $H_0$ может соответствовать истине или нет.

Для проверки гипотез используется статистический критерий – правило, по которому гипотеза $H_0$ отвергается или принимается. Все возможные значения статистики критерия (критической статистики) $\tilde{\Theta}_n$ разделяются на две непересекающиеся области:

*Определение.* **Критическая область (область отклонения гипотезы)** $w_{кр}$ – это область таких значений критерия $k$, попадание в которую критерием маловероятно ( $p = \alpha$ ) при справедливости $H_0$.

*Определение.* **Область принятия гипотезы** $\overline{W}$ – это область, попадание в которую критерием $k$ вероятно при справедливости $H_0$.

Когда наблюдаемое значение критерия $k$ не попало в критическую область $w_{кр}$ (попало в область принятия гипотезы) – гипотеза $H_0$ верна и она не отвергается, т. е. $k \notin w_{кр}$.

В случае попадания наблюдаемого значения критерия $k$ в критическую область $w_{кр}$ – гипотеза $H_0$ неверна и она отвергается, т. е.

$$p(k \in w_{кр} | H_0) = \alpha .$$

Для проверки статистических гипотез используется выборка.

Суть проверки статистических гипотез заключается в следующем.

Используется специально составленная выборочная характеристика (статистика) $\tilde{\Theta}_n(x_1,\ x_2,\ \text{К},\ x_n)$, полученная по выборке $X_1$, $X_2$, ... , $X_n$, точное или приближенное распределение которой известно. Далее по этому выборочному распределению определяется критическое значение $\Theta_{кр}$ – такое, что если гипотеза $H_0$ верна, то вероятность $P(\tilde{\Theta}_n > \Theta_{кр}) = \alpha$ мала. В этом случае событие $\tilde{\Theta}_n > \Theta_{кр}$ можно (с некоторым риском) считать практически невозможным и гипотеза $H_0$ – отвергается. В случае $\tilde{\Theta}_n \le \Theta_{кр}$ событие считается совместимым с гипотезой $H_0$. Нет оснований отвергать гипотезу $H_0$ – она принимается.

Если фактически наблюдаемое значение статистики критерия $\tilde{\Theta}_n$ попадает в критическую область $w_{кр}$, то гипотезу $H_0$ отвергают. При этом возможны четыре случая, представленные в следующей таблице.

| Гипотеза $H_0$ | Принимается $H_0$ | Отвергается $H_0$ |
|---|---|---|
| Верна | Правильное решение | Ошибка *I*-го рода ($\alpha$) |
| Неверна | Ошибка *II*-го рода ($\beta$) | Правильное решение |

При проверке гипотезы $H_0$ по результатам выборки могут быть допущены ошибки двух родов:

- **ошибка *I*-го рода** – отвергнута правильная гипотеза

$$p(I\ \ рода) = p(H_1 | H_0\ \ верна) = \alpha$$ – уровень значимости.

- **ошибка *II*-го рода** – принята неправильная гипотеза

$$p(II\ \ рода) = p(H_0 | H_1\ \ верна) = \beta .$$

Чем больше $\alpha$, тем меньше $\beta$, и наоборот.

Последствия этих ошибок неравнозначны, и роль каждой оценивается до конца по условиям конкретной задачи.

Вероятность $\alpha$ допустить ошибку *I*-го рода (отвергнуть гипотезу $H_0$, когда она верна) называется уровнем значимости (размером, критерия). В расчетах величина $\alpha$ принимает значения 0,05; 0,01; 0,005.

Вероятность допустить ошибку *II*-го рода (принять гипотезу $H_0$, когда она неверна) обозначают $\beta$.

Вероятность $(\mu = 1 - \beta)$ не допустить ошибку *II*-го рода, т.е. отвергнуть гипотезу $H_0$, когда она неверна, называется мощностью критерия.

Смысл величин $\alpha$ и $\beta$ для различных предметных областей можно представить следующим образом:

– в юридических терминах:

$\alpha$ – вероятность вынесения судом обвинительного приговора, когда на самом деле обвиняемый невиновен,

$\beta$ – вероятность вынесения судом оправдательного приговора, когда на самом деле обвиняемый виновен в совершении преступления;

– в технических терминах:

$\alpha$ – вероятность того, что предназначавшийся наблюдателю сигнал не будет им принят,

$\beta$ – вероятность того, что наблюдатель примет ложный сигнал.

Критическую область $w_{кр}$ следует выбирать таким образом, чтобы вероятность попадания в нее статистики критерия $\tilde{\Theta}_n$ была минимальной и равной $\alpha$, если верна нулевая гипотеза $H_0$, и максимальной в противоположном случае.

Другими словами, критическая область должна быть такой, чтобы при заданном уровне значимости $\alpha$ мощность критерия $\mu$ была максимальной.

*Пример 8.1.* По двум независимым выборкам, объемы которых $n_1 = 11$ и $n_2 = 14$, извлеченным из нормальных генеральных совокупностей $X$ и $Y$, найдены исправленные выборочные дисперсии $s_X^2 = 0{,}76$ и $s_Y^2 = 0{,}38$. При уровне значимости $\alpha = 0{,}05$, проверить нулевую гипотезу $H_0 : D(X) = D(Y)$ о равенстве генеральных дисперсий, при конкурирующей гипотезе $H_1 : D(X) > D(Y)$.

*Решение.* Найдем отношение большей исправленной дисперсии к меньшей

$$F_{набл} = 0{,}76 / 0{,}38 = 2.$$

По условию конкурирующая гипотеза $H_1$ имеет вид $D(X) > D(Y)$, следовательно критическая область – правосторонняя.

Для заданного уровня значимости $\alpha = 0{,}05$ и числам степеней свободы

$$k_1 = 11 - 1 = 10,$$
$$k_2 = 14 - 1 = 13$$

с помощью таблицы определим критическую точку

$$F_{кр}(0{,}05;\ 10;\ 13) = 2{,}67.$$

Так как $F_{набл} < F_{кр}$ – нет оснований отвергать гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные дисперсии различаются незначительно.

*Ответ.* $F_{набл} < F_{кр}$ – принятая нулевая гипотеза $H_0$ не отвергается.

*Пример 8.2.* По двум независимым выборкам, объемы которых $n_1 = 14$ и $n_2 = 10$, извлеченным из нормальных генеральных совокупностей $X$ и $Y$, найдены исправленные выборочные дисперсии $s_X^2 = 0{,}84$ и $s_Y^2 = 2{,}52$. При уровне значимости $\alpha = 0{,}1$, проверить нулевую гипотезу $H_0 : D(X) = D(Y)$ о равенстве генеральных дисперсий, при конкурирующей гипотезе $H_1 : D(X) \neq D(Y)$.

*Решение.* Найдем отношение большей исправленной дисперсии к меньшей

$$F_{набл} = 2{,}52 / 0{,}84 = 3.$$

По условию конкурирующая гипотеза $H_1$ имеет вид $D(X) \neq D(Y)$, следовательно, критическая область – двухсторонняя. В соответствии с правилом 2, при отыскании критической точки следует брать уровень значимости, вдвое меньший заданного.

По уровню значимости $\alpha/2 = 0{,}1/2 = 0{,}05$ и числам степеней свободы, определяемым по формулам

$$k_1 = 10 - 1 = 9,$$
$$k_2 = 14 - 1 = 13.$$

С помощью таблицы, определим критическую точку

$$F_{кр}(0{,}05;\ 9;\ 13) = 2{,}72.$$

Так как $F_{набл} > F_{кр}$ – нулевую гипотезу о равенстве генеральных дисперсий отвергаем.

*Ответ.* $F_{набл} > F_{кр}$ – принятая нулевая гипотеза $H_0$ о равенстве генеральных дисперсий отвергается.

*Пример 8.3.* Двумя методами проведены измерения одной и той же физической величины. Получены следующие результаты:

а) в первом случае $x_1 = 9{,}6$, $x_2 = 10{,}0$, $x_3 = 9{,}8$, $x_4 = 10{,}2$, $x_5 = 10{,}6$;

б) во втором случае $y_1 = 10{,}4$, $y_2 = 9{,}7$, $y_3 = 10{,}0$, $y_4 = 10{,}3$.

Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости $\alpha = 0{,}1$?

Предполагается, что результаты измерений распределены нормально и выборки независимы.

*Решение.* Будем судить о точности методов по величинам дисперсий. Таким образом, нулевая гипотеза имеет вид $H_0 : D(X) = D(Y)$. В качестве конкурирующей примем гипотезу $H_1 : D(X) \neq D(Y)$.

Найдем выборочные дисперсии. Для упрощения вычислений перейдем к условным вариантам

$$u_i = 10x_i - 100,$$
$$v_i = 10y_i - 100.$$

В итоге получим условные варианты

| $u_i$ | – 4 | 0 | – 2 | 2 | 6 |
|---|---|---|---|---|---|
| $v_i$ | 4 | – 3 | 0 | 3 | |

Найдем исправленные выборочные дисперсии по формулам

$$s_u^2 = \frac{\sum u_i^2 - [\sum u_i]^2 / n_1}{n_1 - 1},$$
$$s_v^2 = \frac{\sum v_i^2 - [\sum v_i]^2 / n_2}{n_1 - 1}.$$

Получим

$$s_u^2 = \frac{\left|(-4)^2 + (-2)^2 + 2^2 + 6^2\right| - 2^2/5}{5-1} = 14{,}8,$$
$$s_v^2 = \frac{\left|4^2 + (-3)^2 + 3^2\right| - 4^2/4}{4-1} = 10{,}0.$$

Сравним дисперсии. По формуле (4.3) найдем отношение большей исправленной дисперсии к меньшей (каждая из дисперсий увеличилась в $10^2$ раз, но их отношение не изменилось)

$$F_{набл} = 14{,}8/10 = 1{,}48.$$

По условию конкурирующая гипотеза $H_1$ имеет вид $D(X) \neq D(Y)$, поэтому критическая область двусторонняя и, в соответствии с правилом 2, при отыскании критической точки следует брать уровень значимости вдвое меньший заданного.

По таблице, по уровню значимости $\alpha/2 = 0{,}1/2 = 0{,}05$ и числам степеней свободы, определяемым

$$k_1 = 5 - 1 = 4,$$
$$k_2 = 4 - 1 = 3.$$

находим критическую точку $F_{кр}(0{,}05;\ 4;\ 3) = 9{,}12$.

Так как $F_{набл} < F_{кр}$ – нет оснований отвергать гипотезу о равенстве генеральных дисперсий. Другими словами, выборочные исправленные

дисперсии различаются незначительно и, следовательно, оба метода обеспечивают одинаковую точность измерений.

*Ответ.* $F_{набл} < F_{кр}$ – принятая нулевая гипотеза $H_0$ не отвергается, оба метода обеспечивают одинаковую точность измерений.

### Задания для самостоятельного выполнения

**Задание 1.** Имеются две независимые выборки объемов $n_1$ и $n_2$, которые извлечены из нормальных генеральных совокупностей $X$ и $Y$. Для них определены исправленные выборочные дисперсии $s_X^2$ и $s_Y^2$. Для уровня значимости $\alpha$ проверить нулевую гипотезу $H_0 : D(X) = D(Y)$ при конкурирующей гипотезе $H_1$.

1)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 12 | 15 | 0,78 | 0,4 | 0,1 | $D(X) > D(Y)$ |

2)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 15 | 12 | 0,86 | 2,58 | 0,1 | $D(X) \neq D(Y)$ |

3)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 21 | 35 | 0,46 | 0,21 | 0,05 | $D(X) > D(Y)$ |

4)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 18 | 31 | 3,02 | 0,97 | 0,05 | $D(X) \neq D(Y)$ |

5)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 18 | 14 | 0,25 | 0,32 | 0,1 | $D(X) > D(Y)$ |

6)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 24 | 17 | 1,12 | 2,64 | 0,2 | $D(X) \neq D(Y)$ |

7)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 22 | 28 | 0,31 | 0,49 | 0,05 | $D(X) > D(Y)$ |

8)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 13 | 24 | 1,52 | 2,48 | 0,1 | $D(X) \neq D(Y)$ |

9)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 30 | 25 | 0,39 | 0,28 | 0,05 | $D(X) > D(Y)$ |

10)

| $n_1$ | $n_2$ | $s_X^2$ | $s_Y^2$ | $\alpha$ | $H_1$ |
|---|---|---|---|---|---|
| 15 | 19 | 1,35 | 2,94 | 0,1 | $D(X) \neq D(Y)$ |

**Задание 2.** На практическом занятии по физике были проведены измерения силы тока в электрической цепи двумя электроизмерительными приборами. Полученные результаты представлены в таблице.

Можно ли считать, что оба прибора обеспечивают одинаковую точность измерений силы тока в электрической цепи? Предполагается, что результаты измерений распределены нормально и выборки независимы.

1)

| | | | | | | |
|---|---|---|---|---|---|---|
| $X$ | 8,3 | 8,7 | 8,5 | 8,9 | 9,3 | 9,1 |
| $Y$ | 9,1 | 8,4 | 8,7 | 9,0 | 8,3 | 8,5 |

Уровень значимости принять равным $\alpha = 0{,}1$.

2)

| | | | | | | |
|---|---|---|---|---|---|---|
| $Y$ | 6,1 | 6,5 | 6,3 | 6,7 | 7,1 | 6,9 |
| $Z$ | 6,7 | 6,0 | 6,3 | 6,6 | 5,9 | 6,1 |

Уровень значимости принять равным $\alpha = 0{,}2$.

3)

| | | | | | |
|---|---|---|---|---|---|
| $X$ | 4,7 | 5,1 | 4,9 | 5,3 | 5,7 |
| $Y$ | 4,8 | 4,1 | 4,4 | 4,7 | 4,0 |

Уровень значимости принять равным $\alpha = 0{,}1$.

4)

| | | | | | | |
|---|---|---|---|---|---|---|
| $Z$ | 7,6 | 8,0 | 7,8 | 8,2 | 8,6 | |
| $Y$ | 8,4 | 7,7 | 8,0 | 8,5 | 7,6 | 7,8 |

Уровень значимости принять равным $\alpha = 0{,}05$.

5)

| | | | | | |
|---|---|---|---|---|---|
| $X$ | 6,7 | 7,1 | 6,9 | 7,3 | 7,7 |
| $Z$ | 7,3 | 6,6 | 6,7 | 7,1 | |

Уровень значимости принять равным $\alpha = 0{,}25$.

6)

| | | | | | | |
|---|---|---|---|---|---|---|
| $X$ | 5,4 | 5,8 | 5,6 | 6,0 | 6,4 | 6,2 |
| $Y$ | 6,0 | 5,3 | 5,8 | 5,6 | 5,2 | 5,4 |

Уровень значимости принять равным $\alpha = 0{,}1$.

7)

| | | | | | | |
|---|---|---|---|---|---|---|
| $Y$ | 10,0 | 10,4 | 10,2 | 10,6 | 11,0 | 10,8 |
| $Z$ | 11,5 | 10,8 | 11,1 | 11,4 | 10,7 | 10,6 |

Уровень значимости принять равным $\alpha = 0{,}2$.

8)

| | | | | | |
|---|---|---|---|---|---|
| $X$ | 12,1 | 12,5 | 12,3 | 12,7 | 13,1 |
| $Y$ | 12,6 | 11,9 | 12,2 | 12,5 | 11,8 |

Уровень значимости принять равным $\alpha = 0{,}1$.

9)

| | | | | | | |
|---|---|---|---|---|---|---|
| $Z$ | 12,6 | 13,0 | 12,8 | 13,2 | 13,6 | |
| $Y$ | 13,1 | 12,3 | 12,4 | 12,9 | 12,2 | 12,5 |

Уровень значимости принять равным $\alpha = 0{,}05$.

10)

| | | | | | |
|---|---|---|---|---|---|
| $X$ | 13,4 | 13,8 | 13,6 | 14,0 | 14,3 |
| $Z$ | 15,6 | 14,9 | 15,2 | 15,5 | |

Уровень значимости принять равным $\alpha = 0{,}25$.

## Контрольные вопросы

1. Что такое статистическая гипотеза?
2. Что такое выдвигаемая (проверяемая) гипотеза?
3. Что такое альтернативная (конкурирующая) гипотеза?
4. Что такое критическая область (область отклонения гипотезы)?
5. Область принятия гипотезы?
6. Что такое ошибка I-го рода?
7. Что такое ошибка II-го рода?

# РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

## Основная литература

1. Атласов И. В. Типовой расчет. Теория вероятностей и математическая статистика : учебно-методическое пособие / И. В. Атласов, В. Н. Думачев. – Воронеж : Воронежский институт МВД России, 2017. – URL : https://library.vimvd.ru/MegaPro/Download/MObject/3583.

2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие : рек. М-вом общ. и проф. образования РФ / В. Е. Гмурман. – 4-е изд., стереотип. – Москва : Высшая школа, 1997. – 399 с.

3. Данилова О. Ю. Правовая статистика : учебное пособие / О. Ю. Данилова, В. В. Меньших, С. В. Синегубов. – Воронеж : Воронежский институт МВД России, 2018. – 302 с.

4. Копылов А. Н. Теория вероятностей : учебное пособие / А. Н. Копылов. – Воронеж : Воронежский институт МВД России, 2020. – 109 с.

## Дополнительная литература

1. Вентцель Е. С. Теория вероятностей и ее инженерные приложения : учебное пособие : рек. М-вом образ. РФ / Е. С. Вентцель. – 3-е изд., перераб. и доп. – Москва : Академия, 2003. – 458 с.

2. Высшая математика в упражнениях и задачах: [в 2 ч.] / П. Е. Данко [и др.]. – 6-е изд. – Москва : Оникс : Мир и образование, 2007. – 304 с.

3. Гмурман В. Е. Теория вероятностей и математическая статистика : учебное пособие : рек. М-вом общ. и проф. образования РФ / В. Е. Гмурман. – 6-е изд., стереотип. – Москва : Высшая школа, 1998. – 479 с.

4. Королев В. Ю. Теория вероятностей и математическая статистика : учебник : доп. М-вом образ. РФ / В. Ю. Королев. – Москва : Проспект, 2008. – 160 с.

5. Тихонов В. И. Статистический анализ и синтез радиотехнических устройств и систем : учебное пособие : рек. УМО по ун-тскому политехн. образованию / В. И. Тихонов, В. Н. Харисов. – Москва : Радио и связь, 2004. – 608 с.

Приложение 1

Таблица значений вероятностей распределения Пуассона $P_n(k) = (pn)^k \frac{e^{-pn}}{k!}$

| k \ pn | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,9018 | 0,8187 | 0,7408 | 0,6703 | 0,6065 | 0,5458 | 0,4966 | 0,4493 | 0,4066 |
| 1 | 0,0905 | 0,1638 | 0,2222 | 0,2681 | 0,3033 | 0,3293 | 0,3476 | 0,3595 | 0,3659 |
| 2 | 0,0045 | 0,0164 | 0,0333 | 0,0536 | 0,0758 | 0,0988 | 0,1217 | 0,1438 | 0,1647 |
| 3 | 0,0002 | 0,0019 | 0,0033 | 0,0072 | 0,0126 | 0,0198 | 0,0284 | 0,0383 | 0,0494 |
| 4 | 0 | 0,0001 | 0,0002 | 0,0007 | 0,0016 | 0,0030 | 0,0050 | 0,0077 | 0,0111 |
| 5 | 0 | 0 | 0 | 0,0001 | 0,0002 | 0,0004 | 0,0007 | 0,0012 | 0,0020 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0001 | 0,0002 | 0,0003 |

| k \ pn | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,3679 | 0,1353 | 0,0498 | 0,0183 | 0,0067 | 0,0025 | 0,0009 | 0,0003 | 0,0001 | 0,0000 |
| 1 | 0,3679 | 0,2707 | 0,1494 | 0,0733 | 0,0337 | 0,0149 | 0,0064 | 0,0027 | 0,0011 | 0,0005 |
| 2 | 0,1839 | 0,2707 | 0,2240 | 0,1465 | 0,0842 | 0,0446 | 0,0223 | 0,0107 | 0,0050 | 0,0023 |
| 3 | 0,0613 | 0,1804 | 0,2240 | 0,1954 | 0,1404 | 0,0892 | 0,0521 | 0,0286 | 0,0150 | 0,0076 |
| 4 | 0,0153 | 0,0902 | 0,1680 | 0,1954 | 0,1755 | 0,1339 | 0,0912 | 0,0572 | 0,0337 | 0,0189 |
| 5 | 0,0031 | 0,0361 | 0,1008 | 0,1563 | 0,1755 | 0,1606 | 0,1277 | 0,0916 | 0,0607 | 0,0378 |
| 6 | 0,0005 | 0,0120 | 0,0504 | 0,1042 | 0,1462 | 0,1606 | 0,1490 | 0,1221 | 0,0911 | 0,0631 |
| 7 | 0,0001 | 0,0037 | 0,0216 | 0,0595 | 0,1044 | 0,1377 | 0,1490 | 0,1396 | 0,1171 | 0,0901 |
| 8 | 0 | 0,0009 | 0,0081 | 0,0298 | 0,0653 | 0,1033 | 0,1304 | 0,1396 | 0,1318 | 0,1126 |
| 9 | 0 | 0,0002 | 0,0027 | 0,0132 | 0,0363 | 0,0688 | 0,1014 | 0,1241 | 0,1318 | 0,1251 |
| 10 | 0 | 0 | 0,0008 | 0,0053 | 0,0181 | 0,0413 | 0,0710 | 0,0993 | 0,1186 | 0,1251 |
| 11 | 0 | 0 | 0,0002 | 0,0019 | 0,0082 | 0,0225 | 0,0452 | 0,0722 | 0,0970 | 0,1137 |
| 12 | 0 | 0 | 0,0001 | 0,0006 | 0,0034 | 0,0126 | 0,0263 | 0,0481 | 0,0728 | 0,0948 |
| 13 | 0 | 0 | 0 | 0,0002 | 0,0013 | 0,0052 | 0,0142 | 0,0296 | 0,0504 | 0,0729 |
| 14 | 0 | 0 | 0 | 0,0001 | 0,0005 | 0,0022 | 0,0071 | 0,0169 | 0,0324 | 0,0521 |
| 15 | 0 | 0 | 0 | 0 | 0,0002 | 0,0009 | 0,0033 | 0,0090 | 0,0194 | 0,0347 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0,0003 | 0,0014 | 0,0045 | 0,0109 | 0,0217 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0,0001 | 0,0006 | 0,0021 | 0,0058 | 0,0128 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0002 | 0,0009 | 0,0029 | 0,0071 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0001 | 0,0004 | 0,0014 | 0,0037 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0002 | 0,0006 | 0,0019 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0001 | 0,0003 | 0,0009 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0001 | 0,0004 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0002 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,0001 |

Приложение 2

**Таблица значений интегральной функции Лапласа**

| t | Φ(t) | t | Φ(t) | t | Φ(t) | t | Φ(t) | t | Φ(t) | t | Φ(t) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0,00** | 0,0000 | **0,49** | 0,1879 | **0,98** | 0,3365 | **1,47** | 0,4292 | **1,96** | 0,4750 | **2,90** | 0,4981 |
| **0,01** | 0,0040 | **0,50** | 0,1915 | **0,99** | 0,3389 | **1,48** | 0,4306 | **1,97** | 0,4756 | **2,92** | 0,4982 |
| **0,02** | 0,0080 | **0,51** | 0,1950 | **1,00** | 0,3413 | **1,49** | 0,4319 | **1,98** | 0,4761 | **2,94** | 0,4984 |
| **0,03** | 0,0120 | **0,52** | 0,1985 | **1,01** | 0,3438 | **1,50** | 0,4332 | **1,99** | 0,4767 | **2,96** | 0,4985 |
| **0,04** | 0,0160 | **0,53** | 0,2019 | **1,02** | 0,3461 | **1,51** | 0,4345 | **2,00** | 0,4772 | **2,98** | 0,4986 |
| **0,05** | 0,0199 | **0,54** | 0,2054 | **1,03** | 0,3485 | **1,52** | 0,4357 | **2,02** | 0,4783 | **3,00** | 0,49865 |
| **0,06** | 0,0239 | **0,55** | 0,2088 | **1,04** | 0,3508 | **1,53** | 0,4370 | **2,04** | 0,4793 | **3,20** | 0,49931 |
| **0,07** | 0,0279 | **0,56** | 0,2123 | **1,05** | 0,3531 | **1,54** | 0,4382 | **2,06** | 0,4803 | **3,40** | 0,49966 |
| **0,08** | 0,0319 | **0,57** | 0,2157 | **1,06** | 0,3554 | **1,55** | 0,4394 | **2,08** | 0,4812 | **3,60** | 0,499841 |
| **0,09** | 0,0359 | **0,58** | 0,2190 | **1,07** | 0,3577 | **1,56** | 0,4406 | **2,10** | 0,4821 | **3,80** | 0,499928 |
| **0,10** | 0,0398 | **0,59** | 0,2224 | **1,08** | 0,3599 | **1,57** | 0,4418 | **2,12** | 0,4830 | **4,00** | 0,499968 |
| **0,11** | 0,0438 | **0,60** | 0,2257 | **1,09** | 0,3621 | **1,58** | 0,4429 | **2,14** | 0,4838 | **4,50** | 0,499997 |
| **0,12** | 0,0478 | **0,61** | 0,2291 | **1,10** | 0,3643 | **1,59** | 0,4441 | **2,16** | 0,4846 | **5,00** | 0,499997 |
| **0,13** | 0,0517 | **0,62** | 0,2324 | **1,11** | 0,3665 | **1,60** | 0,4452 | **2,18** | 0,4854 | | |
| **0,14** | 0,0557 | **0,63** | 0,2357 | **1,12** | 0,3686 | **1,61** | 0,4463 | **2,20** | 0,4861 | | |
| **0,15** | 0,0596 | **0,64** | 0,2389 | **1,13** | 0,3708 | **1,62** | 0,4474 | **2,22** | 0,4868 | | |
| **0,16** | 0,0636 | **0,65** | 0,2422 | **1,14** | 0,3729 | **1,63** | 0,4484 | **2,24** | 0,4875 | | |
| **0,17** | 0,0675 | **0,66** | 0,2454 | **1,15** | 0,3749 | **1,64** | 0,4495 | **2,26** | 0,4881 | | |
| **0,18** | 0,0714 | **0,67** | 0,2486 | **1,16** | 0,3770 | **1,65** | 0,4505 | **2,28** | 0,4887 | | |
| **0,19** | 0,0753 | **0,68** | 0,2517 | **1,17** | 0,3790 | **1,66** | 0,4515 | **2,30** | 0,4893 | | |
| **0,20** | 0,0793 | **0,69** | 0,2549 | **1,18** | 0,3810 | **1,67** | 0,4525 | **2,32** | 0,4898 | | |
| **0,21** | 0,0832 | **0,70** | 0,2580 | **1,19** | 0,3830 | **1,68** | 0,4535 | **2,34** | 0,4904 | | |
| **0,22** | 0,0871 | **0,71** | 0,2611 | **1,20** | 0,3849 | **1,69** | 0,4545 | **2,36** | 0,4909 | | |
| **0,23** | 0,0910 | **0,72** | 0,2642 | **1,21** | 0,3869 | **1,70** | 0,4554 | **2,38** | 0,4913 | | |
| **0,24** | 0,0948 | **0,73** | 0,2673 | **1,22** | 0,3883 | **1,71** | 0,4564 | **2,40** | 0,4918 | | |
| **0,25** | 0,0987 | **0,74** | 0,2703 | **1,23** | 0,3907 | **1,72** | 0,4573 | **2,42** | 0,4922 | | |
| **0,26** | 0,1026 | **0,75** | 0,2734 | **1,24** | 0,3925 | **1,73** | 0,4582 | **2,44** | 0,4927 | | |
| **0,27** | 0,1064 | **0,76** | 0,2764 | **1,25** | 0,3944 | **1,74** | 0,4591 | **2,46** | 0,4931 | | |
| **0,28** | 0,1103 | **0,77** | 0,2794 | **1,26** | 0,3962 | **1,75** | 0,4599 | **2,48** | 0,4934 | | |
| **0,29** | 0,1141 | **0,78** | 0,2823 | **1,27** | 0,3980 | **1,76** | 0,4608 | **2,50** | 0,4938 | | |
| **0,30** | 0,1179 | **0,79** | 0,2852 | **1,28** | 0,3997 | **1,77** | 0,4616 | **2,52** | 0,4941 | | |
| **0,31** | 0,1217 | **0,80** | 0,2881 | **1,29** | 0,4015 | **1,78** | 0,4625 | **2,54** | 0,4945 | | |
| **0,32** | 0,1255 | **0,81** | 0,2910 | **1,30** | 0,4032 | **1,79** | 0,4633 | **2,56** | 0,4948 | | |
| **0,33** | 0,1293 | **0,82** | 0,2939 | **1,31** | 0,4049 | **1,80** | 0,4641 | **2,58** | 0,4951 | | |
| **0,34** | 0,1331 | **0,83** | 0,2967 | **1,32** | 0,4066 | **1,81** | 0,4649 | **2,60** | 0,4953 | | |
| **0,35** | 0,1368 | **0,84** | 0,2995 | **1,33** | 0,4082 | **1,82** | 0,4656 | **2,62** | 0,4956 | | |
| **0,36** | 0,1406 | **0,85** | 0,3023 | **1,34** | 0,4099 | **1,83** | 0,4664 | **2,64** | 0,4959 | | |
| **0,37** | 0,1443 | **0,86** | 0,3051 | **1,35** | 0,4115 | **1,84** | 0,4671 | **2,66** | 0,4961 | | |
| **0,38** | 0,1480 | **0,87** | 0,3078 | **1,36** | 0,4131 | **1,85** | 0,4678 | **2,68** | 0,4963 | | |
| **0,39** | 0,1517 | **0,88** | 0,3106 | **1,37** | 0,4147 | **1,86** | 0,4686 | **2,70** | 0,4965 | | |
| **0,40** | 0,1554 | **0,89** | 0,3133 | **1,38** | 0,4162 | **1,87** | 0,4693 | **2,72** | 0,4967 | | |
| **0,41** | 0,1591 | **0,90** | 0,3159 | **1,39** | 0,4177 | **1,88** | 0,4699 | **2,74** | 0,4969 | | |
| **0,42** | 0,1628 | **0,91** | 0,3186 | **1,40** | 0,4192 | **1,89** | 0,4706 | **2,76** | 0,4971 | | |
| **0,43** | 0,1664 | **0,92** | 0,3212 | **1,41** | 0,4207 | **1,90** | 0,4713 | **2,78** | 0,4973 | | |
| **0,44** | 0,1700 | **0,93** | 0,3238 | **1,42** | 0,4222 | **1,91** | 0,4719 | **2,80** | 0,4974 | | |
| **0,45** | 0,1736 | **0,94** | 0,3264 | **1,43** | 0,4236 | **1,92** | 0,4726 | **2,82** | 0,4976 | | |
| **0,46** | 0,1772 | **0,95** | 0,3289 | **1,44** | 0,4251 | **1,93** | 0,4732 | **2,84** | 0,4977 | | |
| **0,47** | 0,1808 | **0,96** | 0,3315 | **1,45** | 0,4265 | **1,94** | 0,4738 | **2,86** | 0,4979 | | |
| **0,48** | 0,1844 | **0,97** | 0,3340 | **1,46** | 0,4279 | **1,95** | 0,4744 | **2,88** | 0,4980 | | |

Учебное издание

Анастасия Валерьевна Меньших,
*кандидат технических наук, доцент*;

Маргарита Александровна Панкова,
*кандидат технических наук*

# ОСНОВЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

*Практикум*

В авторской редакции
Компьютерная верстка А. В. Меньших
Объем 7,8 Мб

Воронежский институт МВД России
394065, Воронеж, просп. Патриотов, 53