

**Федеральное агентство по образованию  
Государственное образовательное учреждение  
высшего профессионального образования  
Уфимский государственный авиационный технический университет**

**СТАТИСТИЧЕСКИЙ АНАЛИЗ  
СРЕДСТВАМИ ПАКЕТА MS EXCEL**

**Лабораторный практикум  
по дисциплине «Статистика»**

**Уфа 2008**

Федеральное агентство по образованию  
Государственное образовательное учреждение  
высшего профессионального образования  
Уфимский государственный авиационный технический университет

Кафедра автоматизированных систем управления

СТАТИСТИЧЕСКИЙ АНАЛИЗ  
СРЕДСТВАМИ ПАКЕТА MS EXCEL

Лабораторный практикум  
по дисциплине «Статистика»

Уфа 2008

Составители: В.Ю.Арьков, А.В.Жевак, А.И.Абдулнагимов

УДК 311: 004.45 (07)

ББК 60.6: 32.973-018.2 (Я7)

Статистический анализ средствами пакета MS Excel: Лабораторный практикум по дисциплине «Статистика» / Уфимск. гос. авиац. техн. ун-т; Сост.: В.Ю. Арьков, А.В. Жевак, А.И. Абдулнагимов. – Уфа, 2008. – 51 с.

Содержатся основные сведения по информационным технологиям статистических исследований с использованием пакета MS Excel. Рассматриваются особенности компьютерной обработки и визуализации данных. Приводится порядок проведения лабораторных работ.

Предназначены для студентов 2 курса специальности 080801 «Прикладная информатика в экономике».

Табл. 6. Ил.35. Библиогр.: 6 назв.

Рецензенты: канд. техн. наук, Л.М. Бакусов  
канд. техн. наук, Р.В. Насыров

© Уфимский государственный  
авиационный технический университет, 2008

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	5
ЛАБОРАТОРНАЯ РАБОТА №1	
СВОДКА И ГРУППИРОВКА ДАННЫХ.....	6
1. ЦЕЛЬ РАБОТЫ И ЗАДАЧИ.....	6
2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ.....	6
3. КРАТКОЕ ОПИСАНИЕ ПРОГРАММНОГО КОМПЛЕКСА.....	11
4. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ.....	12
4.1. Упражнение 1. Генерация исходных данных.....	12
4.2. Упражнение 2. Вычисление показателей вариации.....	16
4.3. Упражнение 3. Группировка с помощью статистической надстройки.....	17
4.4. Упражнение 4. Группировка с помощью формул.....	20
4.5. Упражнение 5. Построение графиков.....	22
4.6. Упражнение 6. Сравнение фактического распределения с теоретическим.....	24
Требования к содержанию и оформлению отчета.....	25
Порядок выполнения работы.....	26
Контрольные вопросы.....	27
Критерии результативности выполнения работы.....	27
ЛАБОРАТОРНАЯ РАБОТА №2	
КОРРЕЛЯЦИЯ И РЕГРЕССИЯ.....	27
1. ЦЕЛЬ РАБОТЫ И ЗАДАЧИ.....	27
2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ.....	28
3. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ.....	30
3.1. Упражнение 1. Генерация исходных данных.....	30
3.2. Упражнение 2. Корреляционный анализ.....	31
3.3. Упражнение 3. Регрессионный анализ.....	32
3.4. Упражнение 4. Условное среднее.....	35
3.5. Упражнение 5. Анализ качества модели связи.....	36
Требования к содержанию и оформлению отчета.....	37
Порядок выполнения работы.....	37
Контрольные вопросы.....	38
Критерии результативности выполнения работы.....	38

## ЛАБОРАТОРНАЯ РАБОТА №3

РЯДЫ ДИНАМИКИ .....	39
1. ЦЕЛЬ РАБОТЫ И ЗАДАЧИ .....	39
2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ.....	39
3. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ .....	40
3.1 Упражнение 1. Компоненты динамики .....	40
3.2. Упражнение 2. Загрузка исторических данных.....	41
3.3. Упражнение 3. Анализ тренда .....	43
3.4. Упражнение 4. Скользящая средняя простая.....	45
3.5 Упражнение 5. Скользящая средняя взвешенная .....	46
3.6. Упражнение 6. Экспоненциально взвешенная скользящая средняя.....	46
3.7. Упражнение 7. Биржевые графики .....	47
Требования к содержанию и оформлению отчета.....	49
Порядок выполнения работы .....	50
Контрольные вопросы .....	50
Критерии результативности выполнения работы .....	51
СПИСОК ЛИТЕРАТУРЫ .....	51

## ВВЕДЕНИЕ

Общая теория статистики включает два крупных раздела: описательную статистику и аналитическую статистику. Методы описательной статистики охватывают сбор информации, группировку данных, построение статистических таблиц и графиков, а также вычисление основных статистических показателей. К аналитической статистике обычно относят такие темы, как выборка, вариация, корреляция, регрессия, динамика, структура и индексы.

Предлагаемые лабораторные работы позволяют освоить информационные технологии описательной статистики на примере популярного офисного пакета **Microsoft Office Excel**. В системе помощи дается только *краткое* изложение используемых методов. Поэтому для грамотного использования программы требуется понимание соответствующих разделов общей теории статистики.

Лабораторные работы выполняются с помощью модуля **Анализ данных**, а также с использованием встроенных функций **MS Excel**.

В тексте методических указаний названия файлов, программ, комбинаций клавиш и пунктов меню даются шрифтом **Courier**, чтобы отличать их от основного текста, который набран гарнитурой **Times**. Например, запись [**Сервис** → **Надстройки**] означает, что необходимо подвести курсор к пункту меню **Сервис** и щелкнуть левой кнопкой мыши, затем подвести курсор к пункту **Надстройки** выпадающего меню и щелкнуть левой кнопкой мыши. Выражение [**Shift+F12**] означает, что необходимо нажать на клавишу **Shift** на клавиатуре и затем, не отпуская ее, нажать функциональную клавишу **F12**, после чего отпустить обе клавиши.

# ЛАБОРАТОРНАЯ РАБОТА №1

## СВОДКА И ГРУППИРОВКА ДАННЫХ

### 1. ЦЕЛЬ РАБОТЫ И ЗАДАЧИ

Целью работы является изучение принципов обработки статистических данных средствами **Microsoft Excel** и способов представления результатов в статистических таблицах и на графиках.

Задачами работы является:

- закрепление навыков формирования и заполнения таблиц с использованием встроенных формул;
- овладение методикой группировки данных и вычисления основных показателей вариации;
- приобретение навыков построения статистических графиков.

### 2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Для изучения формы эмпирического распределения проводят группировку данных. Результаты группировки представляют в виде таблиц и графиков.

Аналитическая группировка данных предназначена для анализа корреляционной взаимосвязи. Такая группировка заключается в разбиении диапазона возможных значений на интервалы и подсчете итогов по каждой группе.

Можно использовать следующие правила группирования:

- границы интервалов должны быть круглыми числами (например, 10-20, 70-80, 120-150 и т.п.);
- интервалы должны иметь одинаковую ширину (например: <100, 100-120, 120-140, 140-160, >160). При этом ширина первого и последнего интервалов принимается равной остальным;
- рекомендуемое количество интервалов должно быть на порядок меньше объема выборки;
- желательно избегать появления пустых и малочисленных интервалов.

Для каждого из интервалов необходимо вычислить следующие показатели:

- $n_i$  – частота (количество элементов выборки, попадающих в данный интервал);
- $n_i\%$  – относительная частота, частость (доля числа элементов в данном интервале от объема выборки);
- $K_i\%$  – накопленная частота (сумма частостей текущего интервала и всех предыдущих).

Исследование вариации в статистике имеет большое значение, помогает познать сущность изучаемого явления. Измерение вариации, выяснение ее причины, выявление влияния отдельных факторов дает важную информацию (например, о продолжительности жизни людей, доходах и расходах населения, финансовом положении предприятия и т.п.) для принятия научно обоснованных управленческих решений.

*Вариация* – это различие в значениях какого-либо признака у разных единиц данной совокупности в один и тот же период или момент времени. Вариация возникает в результате того, что индивидуальные значения признака складываются под совокупным влиянием разнообразных факторов (условий), которые по-разному сочетаются в каждом отдельном случае.

К основным показателям вариации относятся: размах вариации, объем выборки, медиана, мода, среднее, дисперсия и т.д. (см. табл.1.1).

Любая случайная величина имеет *функцию распределения* - зависимость плотности вероятности от значения случайной величины. Для *нормального распределения* (распределения Гаусса) функция распределения имеет следующий вид:

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

где  $\mu$  – математическое ожидание,  $\sigma$  – стандартное отклонение. Стандартным нормальным распределением называется нормальное распределение с математическим ожиданием 0 и стандартным отклонением 1.



Таблица 1.1

## Показатели вариации

Название	Обозначение	Название в сводной таблице	Метод вычисления	Формула Excel
1	2	3	4	5
Размах вариации	R	<b>Интервал</b>	Разница максимального и минимального значений	<b>МАКС (интервал) – МИН (интервал)</b>
Объём выборки	n	<b>Счёт</b>	Количество статистических единиц	<b>СЧЁТ (интервал)</b>
Медиана	Me	<b>Медиана</b>	Центральное значение отсортированной выборки	<b>МЕДИАНА (интервал)</b>
Мода	Mo	<b>Мода</b>	Наиболее часто встречающееся значение	<b>МОДА (интервал)</b>
Среднее	$\bar{x}$	<b>Среднее</b>	Среднее арифметическое	<b>СРЗНАЧ (интервал)</b>
Среднее линейное отклонение	d	-	Средний модуль отклонения от среднего значения	<b>СРОТКЛ (интервал)</b>
Дисперсия	D	<b>Дисперсия</b>	Средний квадрат отклонения от среднего значения	<b>ДИСП (интервал)</b>

Окончание таблицы 1.1

1	2	3	4	5
Среднее квадратичное отклонение	$\sigma$	-	Среднее квадратическое отклонение от среднего значения	<b>СТАНДОТКЛОНП</b> (интервал)
Среднее квадратичное отклонение (несмещённая оценка)	$\sigma$	<b>Стандартное отклонение</b>	Среднее квадратическое отклонение от среднего значения с поправкой на объём выборки	<b>СТАНДОТКЛОН</b> (интервал) – <b>несмещённая оценка</b>
Коэффициент осцилляции	$V_R$	-	$V_R = \frac{R}{\bar{x}} \cdot 100$	-
Линейный коэффициент вариации	$V_d$	-	$V_d = \frac{d}{\bar{x}} \cdot 100$	-
Коэффициент вариации	$V_\sigma$	-	$V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100$	-
Коэффициенты асимметрии	A	<b>Асимметрич- ность</b>	$A_s = \frac{\sum (x_i - \bar{x})^3}{n\sigma^3}$	-
Коэффициент эксцесса	E	<b>Эксцесс</b>	$E_x = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} - 3$	-

График функции плотности распределения вероятностей и интегральной функции распределения представлен на рис. 1.1. и 1.2.

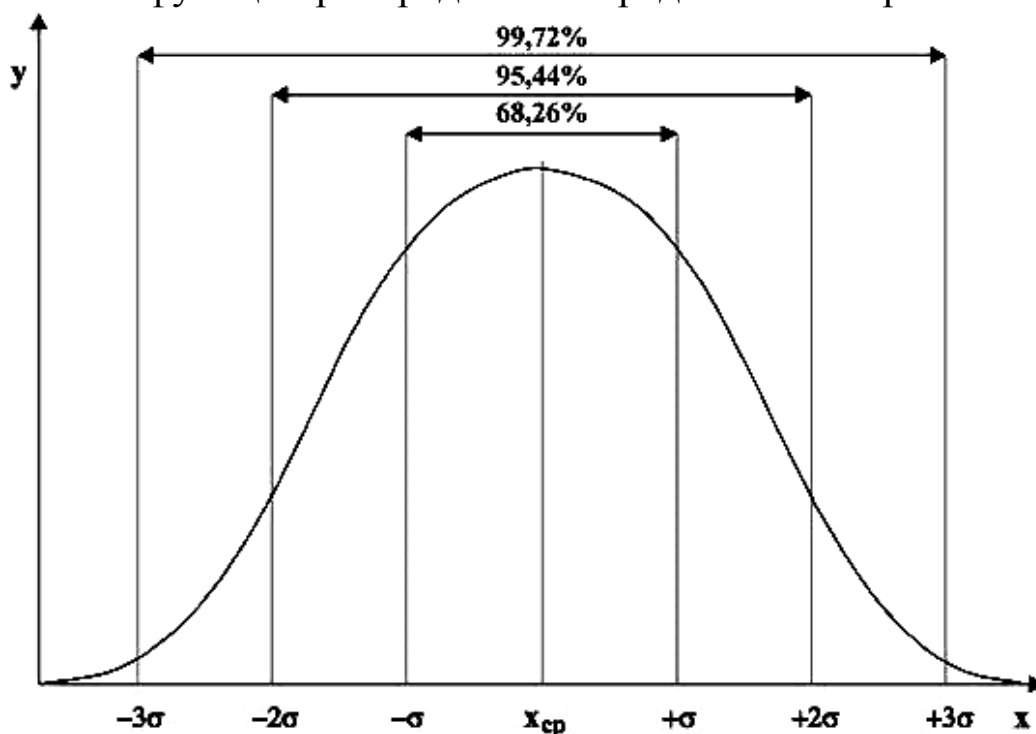


Рис. 1.1. График функции плотности распределения

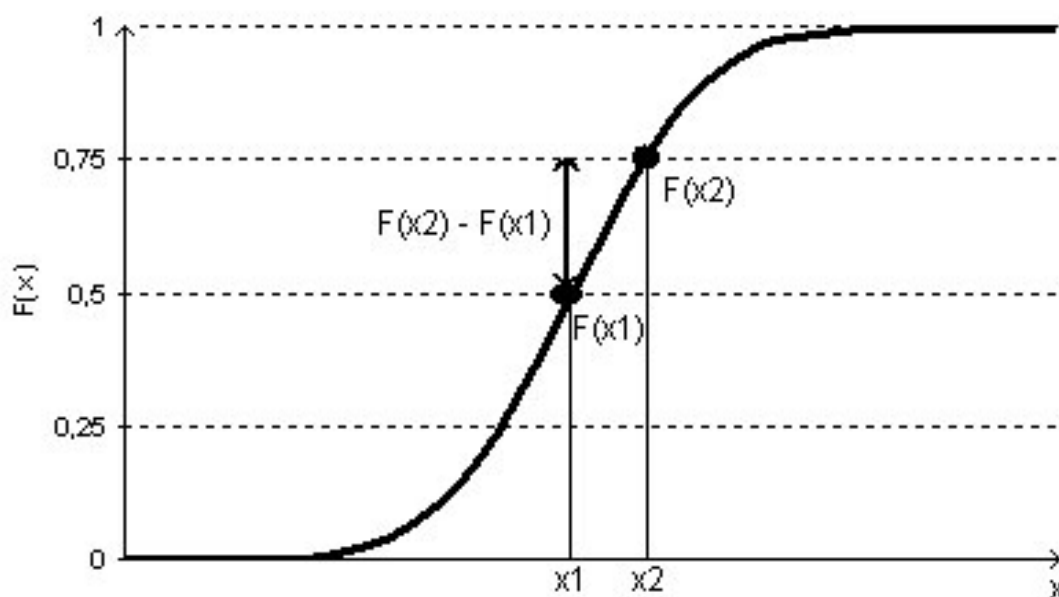


Рис. 1.2. Интегральная функция распределения

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

$$f(x) = \begin{cases} 0, & x < a \\ C, & a \leq x \leq b, \\ 0, & x > b \end{cases} \quad \text{где } C = \frac{1}{b-a}.$$

Постоянная величина  $C$  может быть определена из условия равенства единице площади, ограниченной кривой распределения. Плотность равномерного распределения представлена на рис. 1.3.

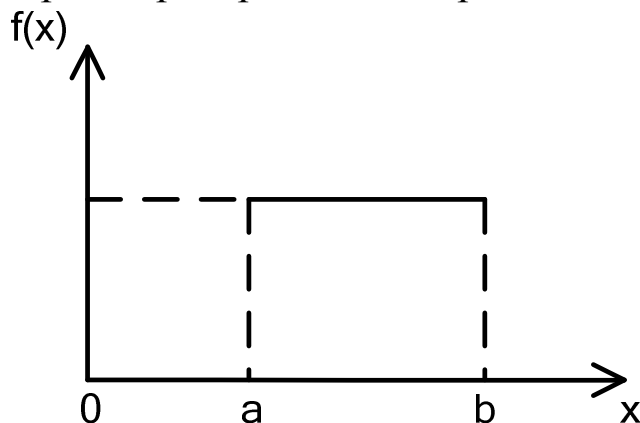


Рис. 1.3. Плотность равномерного распределения

Функция нормального распределения  $F(x)$  на отрезке  $[a, b]$  равна

$$F(x) = \int_{-\infty}^x f(x) dx = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}.$$

### 3. КРАТКОЕ ОПИСАНИЕ ПРОГРАММНОГО КОМПЛЕКСА

Лабораторные работы выполняются в пакете **Microsoft Excel**, который представляет собой электронную таблицу.

Документ **Excel** имеет расширение **\*.xls** и называется *рабочей книгой*. Рабочая книга состоит из листов. Переключаться между листами можно, используя закладки (ярлычки) в нижней части окна «Лист 1» и т.д.

Каждый лист представляет собой таблицу. Таблица состоит из столбцов и строк. Количество столбцов в листе – 256, строк – 65536. Столбцы обозначаются буквами латинского алфавита (в обычном режиме) от **A** до **Z**, затем идет **AA-AZ**, **BB-BZ** и т.д. Строки обозначаются обычными арабскими числами.

На пересечении столбца и строки находится ячейка. Каждая ячейка имеет свой уникальный (в пределах данного листа) адрес, который составляется из буквы столбца (в обычном режиме) и

номера строки. Адрес ячейки используется для работы с данными (ячейками) и формулами.

Статистические расчеты выполняются тремя способами: с помощью формул, функций и статистической надстройки.

Формулы вводятся в ячейку путем набора с клавиатуры. Формулы начинаются со знака равенства (=), например,

**=A1 \* \$B\$2**

**=СУММ(B2 : B151) / 150**

Статистические функции вставляются в формулы, выбрав в верхнем меню [**Вставка** → **Функция** → **Категория** → **Статистические**]. Справочные материалы по этим функциям можно получить в справочном руководстве Microsoft Excel ([**Справка** → **Справка Microsoft Excel**] или **F1**). Пример: вычисления среднего значения:

**=СРЗНАЧ(B2 : B151)**

## **4. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ**

### **4.1. Упражнение 1. Генерация исходных данных**

При выполнении работы также используется *статистическая надстройка* Microsoft Excel. Чтобы активировать надстройку, необходимо выбрать [**Сервис** → **Надстройки**] в меню и поставить галочку напротив пункта [**Пакет анализа**] (см. рис. 1.4). После этого станет доступным пункт меню [**Сервис** → **Анализ данных**].

Исходные данные представляют собой выборку  $\{x_1, x_2, \dots, x_n\}$ , сгенерированную по одному из законов распределения в зависимости от варианта задания (табл. 1.2). Для генерации исходных данных используется функция [**Генерация случайных чисел**] статистической надстройки Microsoft Excel (см. рис. 1.5). После вызова функции, статистическая надстройка предложит указать параметры генерации выборки, как показано на рис. 1.6, 1.7.

**Число переменных** устанавливается равным 1, объем выборки  $n$  (**Число случайных чисел**). Параметры распределения задаются согласно варианту задания (см. табл. 1.2). При генерировании равномерного распределения параметрами являются минимальное и максимальное значения диапазона (min и max), которые вводятся в окне [**Параметры** → **Между**]. Для

нормального распределения указывают среднее значение  $\mu$  и стандартное отклонение  $\sigma$ .

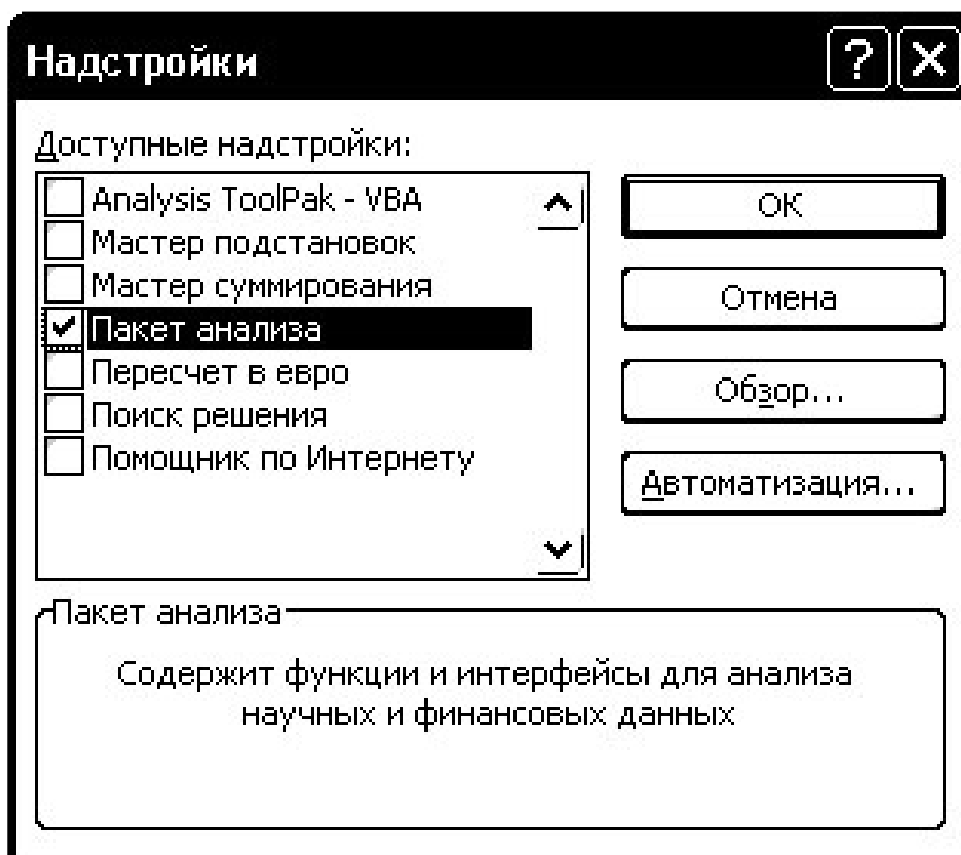


Рис. 1.4. Включение статистической надстройки MS Excel

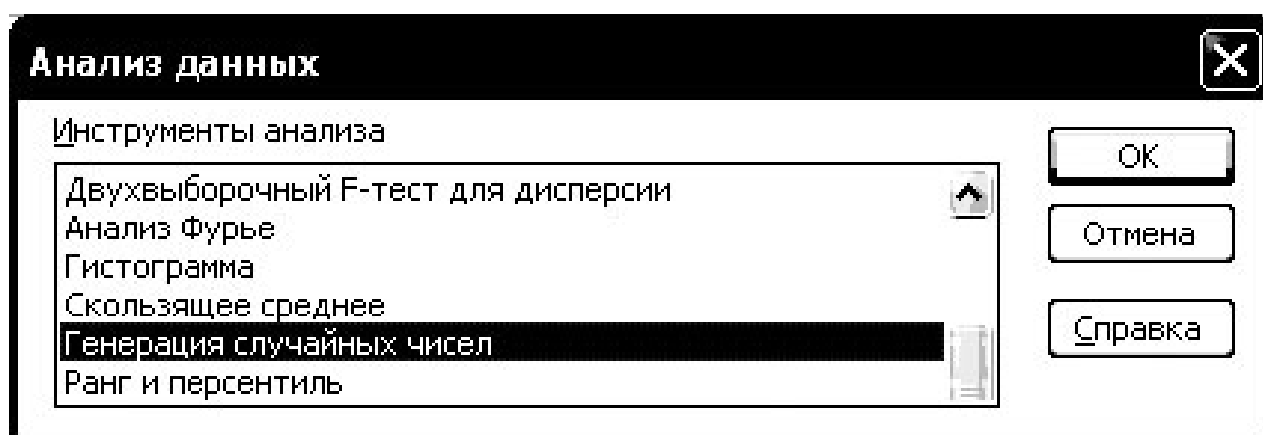


Рис. 1.5. Вызов функции «Генерация случайных чисел»

**Случайное рассеивание** – параметр генератора псевдослучайных чисел, определяющий начало последовательности. Задавая одно и то же значение параметра, можно каждый раз получать одну и ту же последовательность. В качестве параметра задаются четыре последние цифры номера зачетной книжки.

**Генерация случайных чисел**

Число переменных:

Число случайных чисел:

Распределение:

Параметры

Среднее =

Стандартное отклонение =

Случайное рассеивание:

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Рис. 1.6. Выбор параметров генератора нормального распределения

Сгенерированные случайные числа являются вещественными, их необходимо округлить до целых и поместить во второй столбец таблицы. Для этого можно воспользоваться математической функцией **ОКРУГЛ**, имеющей 2 аргумента: округляемое число и количество десятичных разрядов, до которого его нужно округлить. Число разрядов равно 0 в случае округления до целого.

Задаем функцию округления для одной из ячеек столбца «Округлённые числа». Затем ячейку с формулой копируем в буфер и вставляем во все остальные ячейки столбца (см. рис. 1.8). В дальнейшем, в работе используются только округлённые значения.



Рис. 1.7. Выбор параметров генератора равномерного распределения

B2		fx =ОКРУГЛ(A:A;0)			
	A	B	C	D	E
1	<b>Генератор</b>	<b>Рост, см</b>		<b>Рост, см</b>	
2	186,207007	186			
3	180,8470412	181	Среднее		171,3666667
4	174,0076907	174	Стандартная ошибка		1,395531913
5	186,4932707	186	Медиана		173,5
6	175,9891049	176	Мода		175
7	170,5842769	171	Стандартное отклонение		17,09170554
8	183,2322153	183	Дисперсия выборки		292,1263982
9	166,4345225	166	Экссесс		67,75392775
10	174,5641957	175	Асимметричность		-6,802877293
11	160,9356975	161	Интервал		189
12	185,1100192	185	Минимум		150
13	168,5142064	169	Максимум		189
14	155,6472976	156	Сумма		25705
15	189,3441572	189	Счет		150

Рис. 1.8. Генерация случайных чисел и описательная статистика



## Варианты заданий

№ варианта	Случайная величина	Параметры		<i>n</i>
		<i>min</i>	<i>max</i>	
	<i>Равномерное распределение</i>			
1	Рост, см	155	190	150
2	Вес, кг	40	90	100
3	Расходы на потребление	4000	7000	170
4	Месячная зарплата, руб.	10000	20000	180
5	ВВП, трлн. дол.	2	5	180
	<i>Нормальное распределение</i>	$\mu$	$\sigma$	
6	Цена автомобиля, тыс. руб.	400	20	140
7	Число студентов в группе, чел.	25	5	190
8	Расходы семьи на продукты питания в месяц, руб.	5000	500	170
9	Численность рабочих малого предприятия, чел.	20	4	150
10	Продолжительность жизни, лет	60	15	140

#### 4.2. Упражнение 2. Вычисление показателей вариации

Для вычисления показателей вариации применяется функция **Описательная статистика** статистической надстройки Microsoft Excel (см. рис. 1.9). В диалоговом окне нужно выбрать **Входной интервал**, **Метки в первой строке**, **Выходной интервал** и

**Итоговая статистика.** Вычислите показатели вариации, указанные в табл. 1.1.

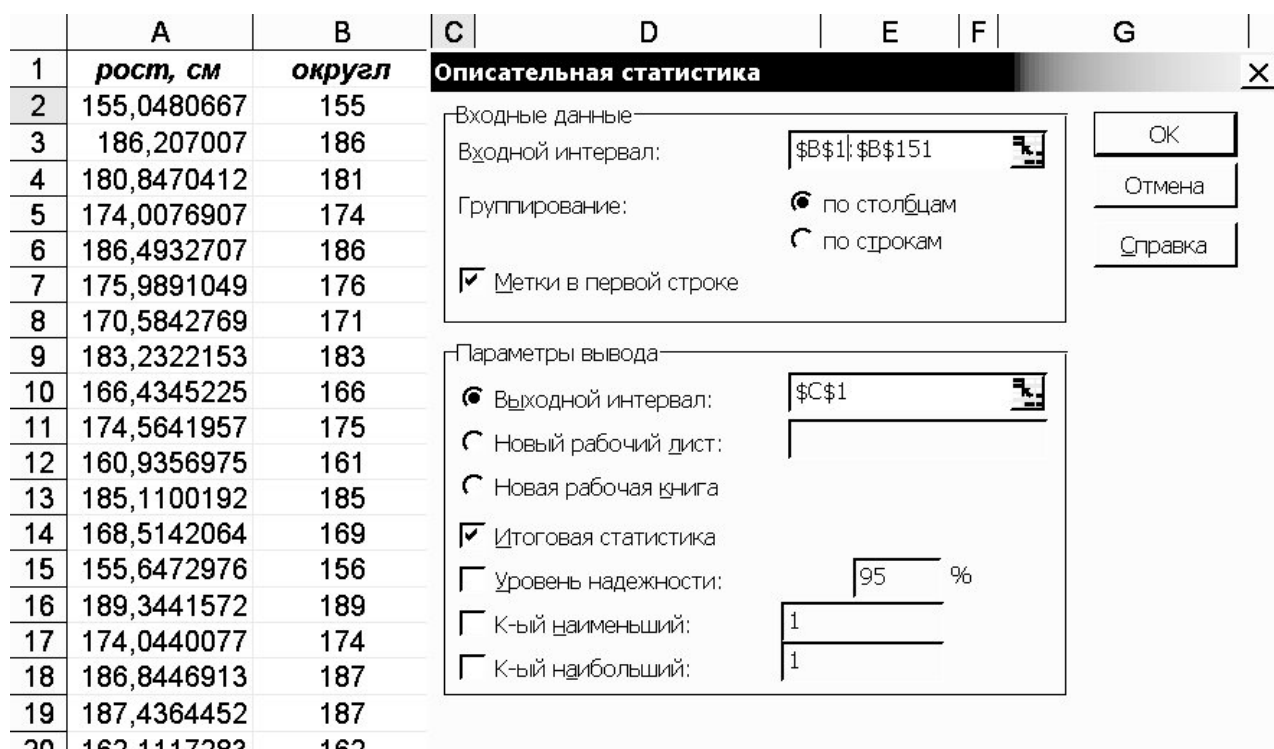


Рис. 1.9. Использование функции «Описательная статистика»

При анализе показателей вариации можно использовать следующие правила:

- выборка считается однородной, если коэффициент вариации  $V_{\sigma} \leq 30\%$ ;
- если коэффициенты асимметрии и эксцесса близки к нулю, то форму распределения можно считать близкой нормальному.

Критические значения  $A$  и  $E$  вычисляются по оценкам дисперсий:

$$D(A) = \frac{6(n-1)}{(n+1)(n+3)}; \quad D(E) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

Критерий согласия:

$$|A| \leq 3\sqrt{D(A)}; \quad |E| \leq 5\sqrt{D(E)}$$

### 4.3. Упражнение 3. Группировка с помощью статистической надстройки

Группировка данных проводится двумя способами: с помощью стандартных функций Excel и статистической надстройки.

Вначале следует создать таблицу нижних границ интервалов группирования (рис. 1.10).

	A	B	C	D	E	F	G	H	I
1	Сгенерированные числа	Округлённые числа	Карманы						
2	70,6124072	71	80						
3	111,5359853	112	90						
4	96,78733502	97	100						
5	106,6241455	107	110						
6	110,1162868	110	120						
7	117,4143679	117							
8	106,1346782	106							
9	100,9696123	101							
10	109,879227	110							
11	126,6270945	127							
12	89,89264895	90							
13	110,9991106	111							
14	97,80183089	98							
15	100,8367238	101							
16	77,73811629	78							
17	88,80405201	89							
18	96,55121883	97							
19	102,8880777	103							
20	96,7026497	97							
21	82,20373527	82							
22	96,99767841	97							
23	95,75112497	96							

**Гистограмма**

Входные данные

Входной интервал:

Интервал карманов:

Метки

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Парето (отсортированная гистограмма)

Интегральный процент

Вывод графика

Рис. 1.10. Настройка функции **Гистограмма**

Для группировки данных с помощью статистической надстройки выбираем меню [Сервис → Анализ данных → Гистограмма]. Указываем следующие параметры:

**Входные данные:**

- **Входной интервал** – выборка исходных данных;
- **Интервал карманов** – нижние границы интервалов группирования.

**Параметры вывода:**

- **Выходной интервал** – расположение результатов группировки на листе;
- **Вывод графика** – построение гистограммы.
- **Интегральный процент** – вычисление накопленных частот.

Результат работы функции **Гистограмма** представлен на рис. 1.11. Сгенерированную таблицу необходимо дополнить недостающими столбцами. График необходимо настроить для

корректного отображения. На рис. 1.12 показан пример рекомендуемого стиля оформления таблицы и графика.

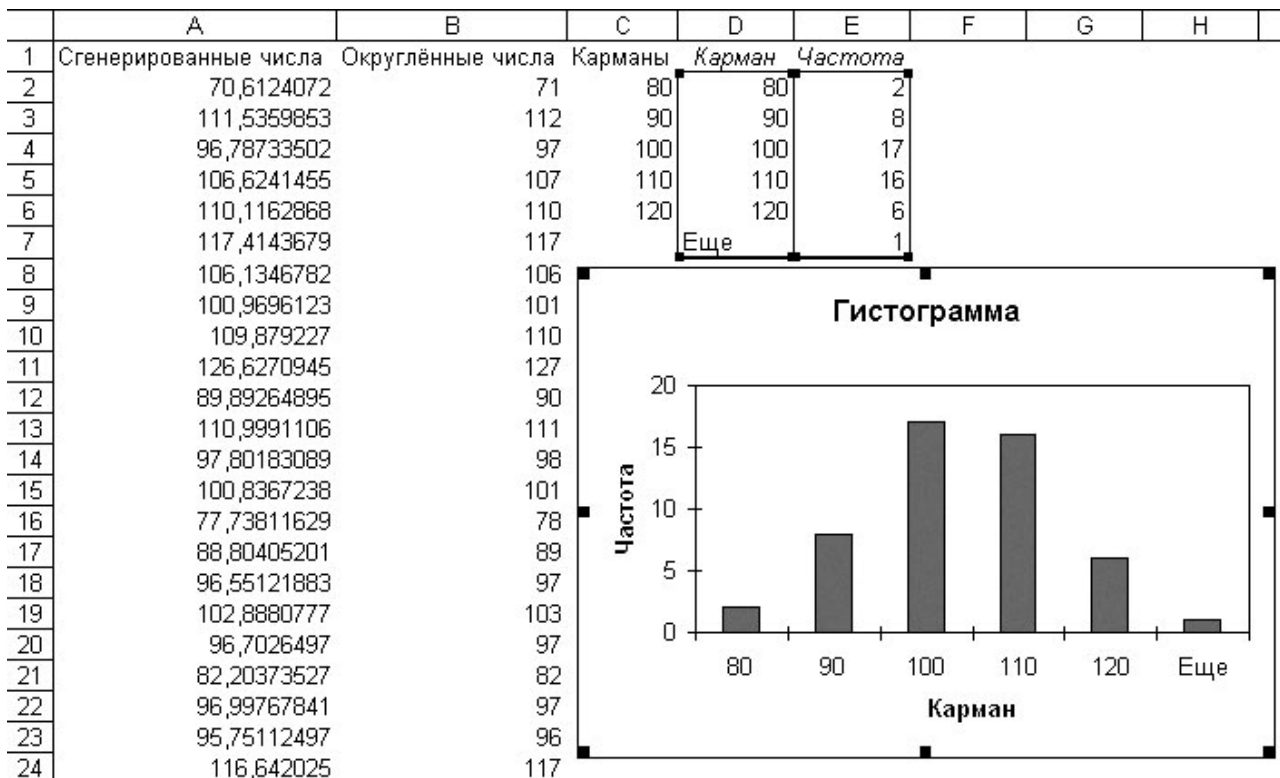


Рис. 1.11. Результат вызова функции «Гистограмма»

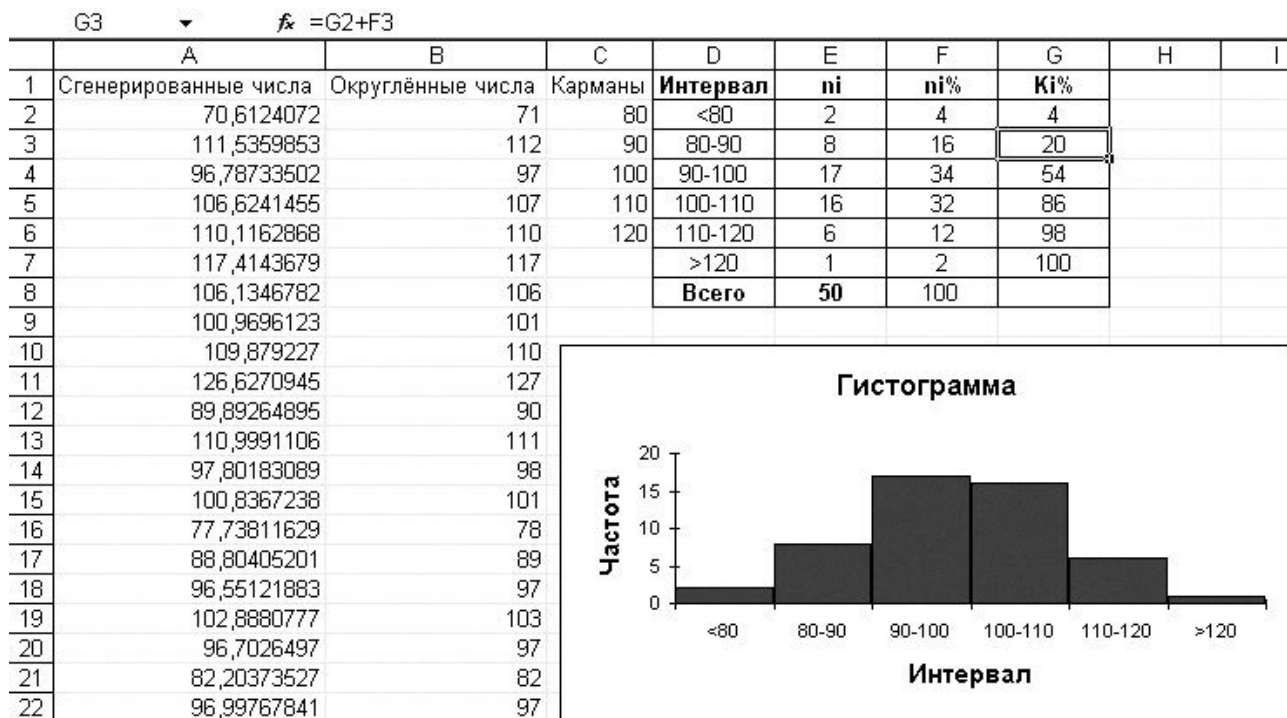


Рис. 1.12. Настройка результатов группировки

Таблицу необходимо дополнить следующими столбцами:

- *Интервал* – подписи для столбцов гистограммы в виде текстовых меток, описывающих границы интервалов группировки;
- *Частота*  $n_i$  (%);
- *Кумулята*  $K_i$  (%).

В графе *Всего* выполняют подсчет суммы частот.

Для вычисления недостающих характеристик интервалов следует использовать формулы. Например, для вычисления накопленной частоты для интервала 90..100 (ячейка **G3**) используется формула **=G2+F3**. Вычисление относительной частоты для интервала 80..90 (ячейка **F3**) выполняется с помощью формулы: **=100\*E3/\$E\$8**.

После вычислений следует убедиться в отсутствии грубых ошибок. Например, накопленная частота должна быть равна 100%.

Расположение столбиков гистограммы должно соответствовать границам интервалов группирования данных. Для настройки графика щелкните по столбику гистограммы курсором и нажмите правую кнопку мыши. Выберите [**Формат рядов данных → Параметры**] и установите нулевое значение параметров **Перекрытие** и **Ширина зазора**.

#### 4.4. Упражнение 4. Группировка с помощью формул

Пример группировки с помощью формул приводится на рис. 1.13.

Подсчет частоты попадания в интервал значений определяется как разность количества значений меньше верхней границы и меньше нижней границы интервала. Например, частота для первого интервала (ячейка **G6**) рассчитана с помощью функции **СЧЁТЕСЛИ**:

**СЧЁТЕСЛИ (В : В ; "<=" &E6) – СЧЁТЕСЛИ (В : В ; "<=" &D6) .**

Первая часть приведённой формулы вычисляет количество ячеек столбца **В**, значения которых меньше или равно содержимому ячейки **E6** (верхняя граница диапазона). Вторая часть формулы вычисляет количество ячеек столбца **В**, значения которых меньше или равно содержимому ячейки **D6** (нижняя граница диапазона). Таким образом, функция в целом даёт количество ячеек столбца **В**,

значения которых попадают в интервал между значениями ячеек D6 и E6.

G11			=СЧЁТЕСЛИ(В:В; "<="&E11)-СЧЁТЕСЛИ(В:В; "<="&D11)						
	A	B	C	D	E	F	G	H	I
1	70,61241	71							
2	111,536	112	80						
3	96,78734	97	90						
4	106,6241	107	100						
5	110,1163	110	110						
6	117,4144	117	120						
7	106,1347	106							
8	100,9696	101							
9	109,8792	110							
10	126,6271	127							
11	89,89265	90							
12	110,9991	111							
13	97,80183	98							

Сводная таблица					
Интервал			ni	ni, %	ki %
Нижняя граница	Верхняя граница	Середина			
0	70	35	0	0	0
70	80	75	3	3	3
80	90	85	13	13	16
90	100	95	37	37	53
100	110	105	32	32	85
110	120	115	13	13	98
120	130	125	2	2	100
Всего			100	100	-

Рис. 1.13. Группировка данных с использованием формул

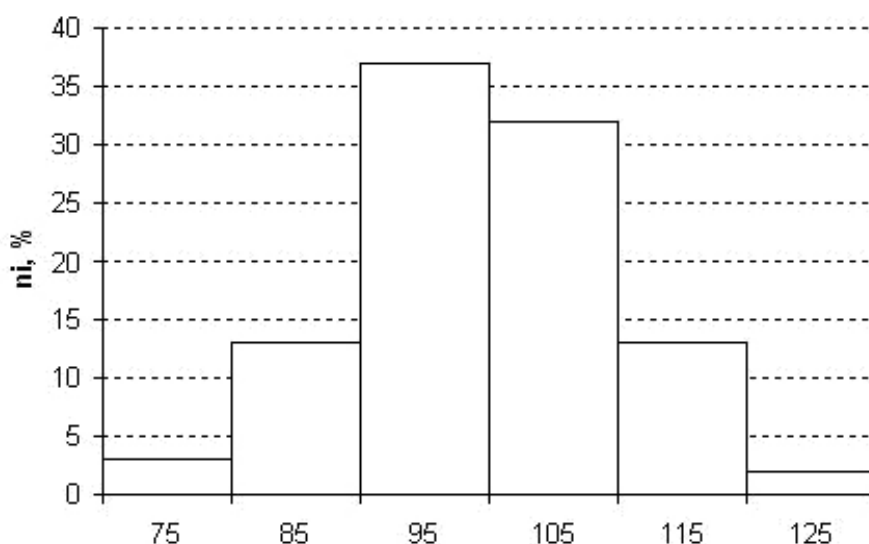


Рис. 1.14. Гистограмма по результатам группировки

По рассматриваемой таблице строится гистограмма (рис. 1.14). для этого выбираем в меню [Вставка → Диаграмма → Гистограмма → Обычная гистограмма]. Переходим на закладку Ряд и нажимаем кнопку **Добавить**. Нажимаем кнопку **Значения** и указываем диапазон значений частот. Нажимаем кнопку **Подписи оси X** и указываем диапазон меток для оси X. Нажимаем кнопки **Далее** → **Готово**. В качестве меток можно указать середины интервалов группировки.

## 4.5. Упражнение 5. Построение графиков

Чтобы добавить полигон на график гистограммы, необходимо добавить к новый ряд данных. Для этого нужно щелкнуть правой кнопкой мыши по области графика и в контекстном меню выбрать [Исходные данные → Ряд] (рис. 1.15).

Сводная таблица					
Интервал			ni	ni, %	ki %
Нижняя граница	Верхняя граница	Середина			
0	70	35	0	0	0
70	80	75	3	3	3
80	90	85	13	13	16
90	100	95	37	37	53
100	110	105	32	32	85
110	120	115	13	13	98
120	130	125	2	2	100
Всего			100	100	-

Исходные данные

Диапазон данных Ряд

ni, %

75 85 95 105 115 125

Ряд

Гистограмма  
Полигон

Имя: Полигон

Значения: =Лист1!\$N\$7:\$N\$12

Добавить Удалить

Подписи оси X: =Лист1!\$F\$7:\$F\$12

OK Отмена

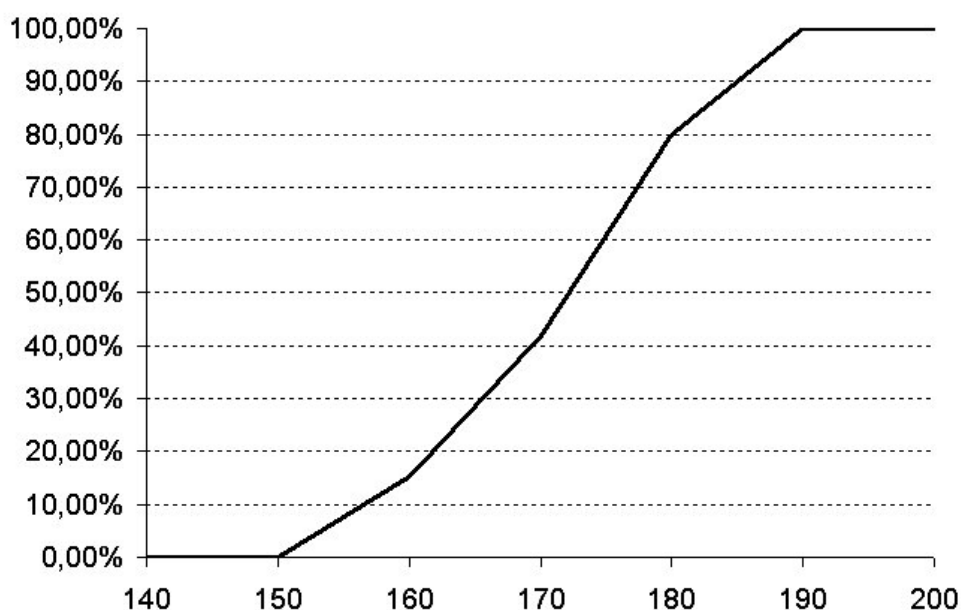
Рис. 1.15. Добавление данных для построения полигона

Нажимаем кнопки **Добавить** → **Значения** и указываем тот же диапазон ячеек, что и для гистограммы. При работе с рядами данных в окне **Имя** можно ввести комментарий, поясняющий содержание набора данных для графика.

Кумулята строится на отдельном графике. Выбираем в меню **[Вставка → Диаграмма → Точечная диаграмма → Точечная диаграмма, на которой значения соединены отрезками]**. Добавляем ряд значений, указав диапазоны по оси X и по оси Y. Кумулята – это линейный график накопленных частот, представляющий собой «интеграл от гистограммы». Пример кумуляты показан на рис. 1.16. На рис. 1.17 можно видеть окончательные варианты графика гистограммы и полигона.

Все графики должны быть оформлены надлежащим образом. Например, минимальное и максимальное значения по оси Y кумуляты должны быть 0 и 100, соответственно (правый клик по оси Y, **[Формат оси → Шкала]**); линия кумуляты должна начинаться от  $x = 0$  (правый клик по оси X, **[Формат оси → Шкала]**, убрать галочку **[Пересечение с осью Y (значений) в максимальном значении]**). Подписи по осям должны быть информативными.

Положение оси Y устанавливается с помощью пункта **[Ось Y (значений) пересекает в значении...]**.



*Рис. 1.16. Пример кумуляты*



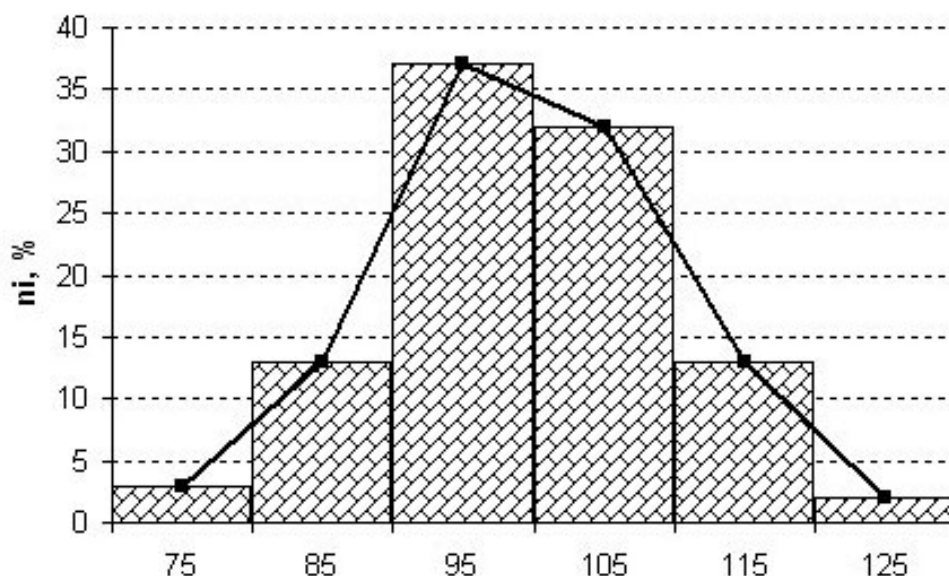


Рис. 1.17. Пример гистограммы и полигона

#### 4.6. Упражнение 6. Сравнение фактического распределения с теоретическим

Исходная выборка генерируется по стандартному закону распределения с параметрами в соответствии с вариантом задания (см. разд. 4.1). Чтобы сравнить фактическое и теоретическое распределения, необходимо построить их графики. Для работы с теоретическими распределениями используются готовые статистические функции, например, **НОРМРАСП** и **НОРМОБР**. Описание статистических функций необходимо самостоятельно изучить, вызвав **Справку**.

На рис. 1.18 показан пример графика теоретического и фактического распределений.

Чтобы вычислить значение теоретической вероятности попадания случайной величины в интервал  $[x_1, x_2)$ , необходимо найти разность вероятности попадания в интервалы  $[0, x_2)$  и  $[0, x_1)$  (см. рис. 1.2). Например, вероятность попадания случайной величины в интервал  $[100, 110)$  для нормального закона распределения со средним 100 и стандартным отклонением 20, равна:

$$= \text{НОРМРАСП}(110; 100; 20; \text{ИСТИНА}) - \text{НОРМРАСП}(100; 100; 20; \text{ИСТИНА}) .$$

В случае равномерного распределения с нижней границей 50 и верхней 150, может быть использована следующая формула:

$$= (110 - 50) / (150 - 50) - (100 - 50) / (150 - 50) \\ = (110 - 100) / (150 - 50) * 100\% .$$

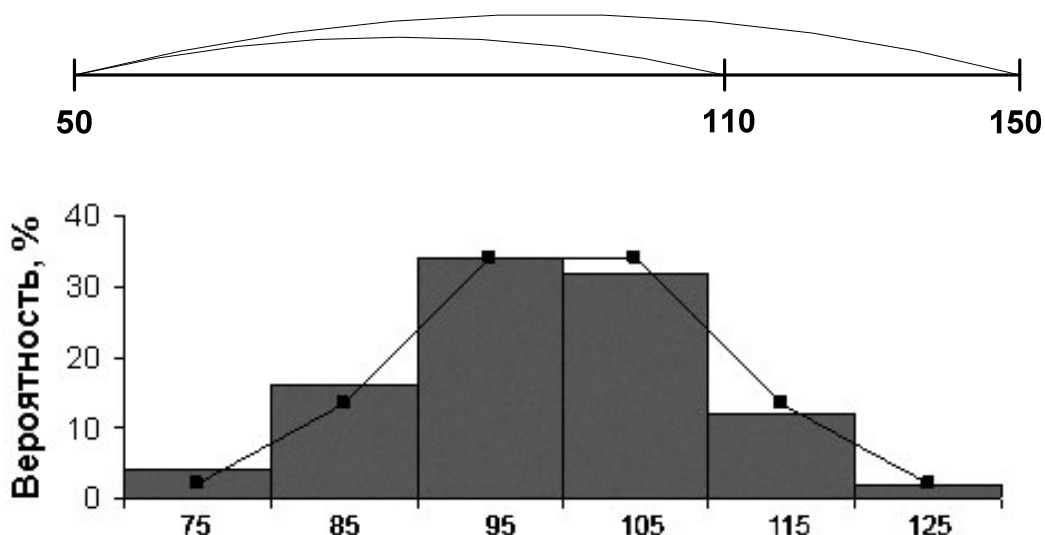


Рис. 1.18. Эмпирическое (столбики) и теоретическое распределение (линия)

### Требования к содержанию и оформлению отчета

Отчет должен быть продемонстрирован как на бумажном носителе, содержащем показатели вариации, результаты группировки, формы распределения, статистические графики, так и в электронной форме в виде файла с заполненной таблицей и графиками.

Статистические таблицы должны содержать все элементы, необходимые для анализа материала: заголовки, названия, единицы измерения, диапазоны измерения. Основной заголовок должен отвечать на три вопроса: «Что? Где? Когда?», т.е. содержать информацию об объекте исследования с привязкой к времени и месту.

Рисунки также должны иметь заголовок (подрисуючную подпись). Обозначения по осям должны содержать названия статистических показателей и единицы измерения. Если на одном графике изображено несколько кривых, необходимо указать, что изображает каждая из них. Такая информация дается в виде легенды или в составе подрисуючной подписи.

Отчёт заканчивается выводами относительно изучаемой выборки. В выводах не пересказывают этапы проведённых работ, а кратко излагают результаты. Например:

- какие данные проанализированы;
- есть ли явные ошибки и несоответствия в данных;

- какие закономерности в данных обнаружены;
- какими графиками и показателями это подтверждается.

Титульный лист отчета должен содержать всю информацию, необходимую для однозначной идентификации авторов и работы. Для этого на титульном листе указывают название дисциплины, тему и номер работы, вариант задания, номер группы, фамилии и инициалы студентов, должность, фамилию и инициалы преподавателя и т.п. (в соответствии со стандартом УГАТУ на оформление текстовых документов).

### **Порядок выполнения работы**

1. Ознакомьтесь с описанием следующих функций:  
**СТАНДОКЛ, СТАНДОТКЛП, НОРМРАСП.**
2. Сгенерируйте исходные данные.
3. Вычислите основные статистические показатели тремя способами:
  - с помощью формул;
  - с помощью статистических функций;
  - с помощью статистической надстройки.
4. Сравните результаты расчетов.
5. Сделайте вывод об однородности выборки.
6. Сделайте вывод о близости к нормальному распределению.
7. Проведите группировку данных двумя способами:
  - с помощью стандартных функций Excel;
  - с помощью статистической надстройки.
8. Постройте сводные таблицы и сравните результаты группировки, полученные двумя способами.
9. Постройте гистограмму, полигон и кумуляту двумя способами:
  - с помощью статистической надстройки
  - «вручную».
10. По виду гистограммы определите модальный интервал.
11. По графику кумуляты найдите 80%-й квантиль.
12. Постройте гистограммы теоретического и эмпирического распределений на одном графике.
13. Сделайте вывод о близости эмпирического распределения к теоретическому.
14. Оформите отчёт в соответствии с требованиями.

### **Контрольные вопросы**

1. Как вычисляются основные статистические показатели?
2. Как проводится группировка?
3. Какие существуют статистические графики, как они строятся и в чём их назначение?
4. В чём разница между теоретическим и эмпирическим распределением?
5. Какие требования предъявляются к оформлению таблиц и графиков?
6. Что должны содержать выводы?

### **Критерии результативности выполнения работы**

Лабораторная работа считается выполненной в том случае, если студент

1. выполнил все указанные задания, следуя порядку выполнения работы, и представил отчет, содержащий показатели вариации, результаты группировки и статистические графики;
2. освоил методику выполнения типовых заданий и способен продемонстрировать работу программы;
3. ответил на все контрольные и дополнительные вопросы.

## **ЛАБОРАТОРНАЯ РАБОТА №2**

### **КОРРЕЛЯЦИЯ И РЕГРЕССИЯ**

#### **1. ЦЕЛЬ РАБОТЫ И ЗАДАЧИ**

Освоить методы статистического исследования корреляционной взаимосвязи.

Задачами работы является:

- приобретение навыков построения корреляционного поля и уравнения регрессии;
- закрепление навыков нахождения коэффициентов уравнения регрессии и условного среднего.

## 2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Анализ статистических данных позволяет выявить взаимосвязь исследуемых явлений. При этом наблюдаются два вида связи: причинно-следственные отношения (изменения в одном явлении являются причиной изменений в другом) и корреляция (изменения в обоих явлениях происходят одновременно и вызваны общей причиной). Корреляционная зависимость выглядит как разброс точек относительно линии на диаграмме рассеяния. Модель взаимосвязи отражает количественные отношения и строится методами корреляционного и регрессионного анализа. Корреляционный анализ позволяет исследовать тесноту связи, т.е. степень разброса точек от линии. Регрессионный анализ позволяет построить уравнение связи. Для правильной интерпретации модели необходим этап качественного анализа изучаемого явления, его природы и внутренних механизмов.

Теснота линейной связи оценивается с помощью коэффициента линейной корреляции

$$r_{yx} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}.$$

Коэффициент корреляции принимает значения от  $-1$  до  $+1$ , включительно; его знак указывает на обратную или прямую связь показателей. Величина коэффициента характеризует тесноту линейной связи (см. табл. 2.1).

Таблица 2.1

Оценка тесноты линейной связи

Величина коэффициента корреляции	Характер связи
$ r  < 0,3$	Практически отсутствует
$0,3 \leq  r  < 0,5$	Слабая
$0,5 \leq  r  < 0,7$	Умеренная
$0,7 \leq  r  < 1,0$	Сильная
$ r  = 1,0$	Функциональная

Низкое значение коэффициента говорит об отсутствии *линейной* связи. Фактической причиной могут быть полное отсутствие связи,

высокий уровень случайных отклонений, либо наличие существенно нелинейной связи. Теснота *нелинейной* связи может оцениваться с помощью коэффициентов ранговой корреляции.

Модель связи обычно строится в форме уравнения регрессии. Парная регрессия (связь 2 показателей) может описываться уравнениями:

$$\text{Прямой: } \bar{y}_x = a_1x + a_0;$$

$$\text{Параболы: } \bar{y}_x = a_2x^2 + a_1x + a_0;$$

$$\text{Кубическая парабола: } \bar{y}_x = a_3x^3 + a_2x^2 + a_1x + a_0.$$

Неизвестные коэффициенты  $a_0, a_1, \dots, a_k$  могут быть найдены методом наименьших квадратов (МНК), путём минимизации суммы квадратов невязок:

$$\sum (\bar{y}_x - y_x)^2 \rightarrow \min .$$

Системы уравнений для вычисления коэффициентов регрессии для полиномов различных степеней выглядят следующим образом:

$$\begin{cases} \sum y = a_0n + a_1 \sum x \\ \sum yx = a_0 \sum x + a_1 \sum x^2 \end{cases} \quad \text{для прямой;}$$

$$\begin{cases} \sum y = a_0n + a_1 \sum x + a_2 \sum x^2 \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 \\ \sum yx^2 = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 \end{cases} \quad \text{для параболы;}$$

$$\begin{cases} \sum y = a_0n + a_1 \sum x + a_2 \sum x^2 + a_3 \sum x^3 \\ \sum yx = a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 + a_3 \sum x^4 \\ \sum yx^2 = a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 + a_3 \sum x^5 \\ \sum yx^3 = a_0 \sum x^3 + a_1 \sum x^4 + a_2 \sum x^5 + a_3 \sum x^6 \end{cases} \quad \begin{array}{l} \text{для кубической} \\ \text{параболы.} \end{array}$$

Общий вид системы уравнений в матричной записи:  $Y = Z \cdot A$ .

Для дальнейшей работы целесообразно вычислить промежуточные значения, такие как  $\sum x^2$ ,  $\sum x^3$  и т.д. Используя полученные суммы, составляем матрицы для системы нормальных уравнений. Например, для построения линейного уравнения регрессии потребуются следующие матрицы

$$Y = \begin{pmatrix} \sum y \\ \sum yx \end{pmatrix}; \quad Z = \begin{pmatrix} n & \sum x \\ \sum x & \sum x^2 \end{pmatrix}.$$

Находим оценки коэффициентов уравнения регрессии:

$$Y = Z \cdot A \Rightarrow A = Z^{-1} \cdot Y$$

Таким образом, для нахождения значений матрицы коэффициентов регрессии  $A$  надо найти матрицу, обратную  $Z$  (т.е.  $Z^{-1}$ ), и умножить её слева на матрицу  $Y$ .

### 3. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ

#### 3.1. Упражнение 1. Генерация исходных данных

Для начала работы требуется сгенерировать значения двух переменных  $x$  и  $y$ , в соответствии с табл. 2.2. Объем выборки – 100 элементов.

Таблица 2.2

#### Варианты заданий

№	Фактор (x)	Результат (y)
1	Зарплата: 3000..10000 руб.	Потребление, руб.: $y = 1500 + 0,5 \cdot x + 500 \cdot e$
2	Доходы: 3500..11000 руб.	Сбережения, руб.: $y = -1000 + 0,5 \cdot x + 300 \cdot e$
3	Число студентов: 2000..10000	Число преподавателей ВУЗа: $y = 220 + 0,09 \cdot x + 50 \cdot e$
4	Цена товара: 15..50 руб/кг	Спрос, кг: $y = 200000 \cdot x^{-0,85} + 500 \cdot e$
5	Валовой национальный продукт: 1..8 трлн. руб.	Личные доходы, трлн. дол.: $y = -0,4 + 0,95 \cdot x + 0,4 \cdot e$
6	Расходы на рекламу: 0..10 млн. руб.	Прибыль, млн. руб.: $y = 10 + 6 \cdot x - 0,3 \cdot x^2 + 2 \cdot e$

7	Денежная масса: 100..350 млрд. руб.	Индекс цен, %: $y = 38 + 0,3 \cdot x + 7 \cdot e$
8	Индекс трудозатрат: 100..160 %	Индекс объёма продукции, %: $y = 7 \cdot x^{0,6} + 2 \cdot e$

Упомянутая в таблице случайная составляющая  $e$  имеет нормальное распределение с единичной дисперсией и нулевым математическим ожиданием. Значения  $e$  следует сгенерировать отдельно, при помощи функции «Генерация случайных чисел» статистической надстройки. Этот же способ можно использовать для генерации значений  $x$  (тип распределения – «равномерное», левая и правая границы – согласно варианту задания). Полученные значения  $x$  и  $y$  целесообразно округлить до того или иного знака после запятой (или до целого), в зависимости от порядка полученных величин (зависит от варианта). Для округления используется функция **ОКРУГЛ (число; число разрядов)**. Пример результата генерации данных и округления можно видеть на рис. 2.1. В дальнейшей работе используются только округлённые значения  $x$  и  $y$ .

E3		fx =ОКРУГЛ(C3:C102;0)				
	A	B	C	D	E	F
1	<b>Сгенерированные значения</b>				<b>Округленные значения</b>	
2	<b>e</b>	<b>x2</b>	<b>x</b>	<b>y</b>	<b>Зарплата</b>	<b>Потребление</b>
3	-2,99478	9096256	3016,45	1510,83266	3016	1511
4	1,23524	17656804	4202,31	4218,77212	4202	4219
5	0,63869	18757561	4331,34	3985,01378	4331	3985
6	0,10819	84750436	9206,37	6157,27755	9206	6157
7	1,28046	11108889	3332,83	3806,646	3333	3807
8	0,25254	73719396	8586,41	5919,47741	8586	5919
9	-0,13763	33246756	5765,86	4314,11408	5766	4314
10	0,86556	88228449	9393,08	6629,31998	9393	6629
11	-0,44904	30902481	5559,5	4055,22664	5559	4055

Рис. 2.1. Пример результата генерации и округления данных

### 3.2. Упражнение 2. Корреляционный анализ

Для вычисления коэффициентов корреляции можно использовать как функцию «Корреляция» статистической надстройки, так и функцию **КОРРЕЛ (диапазон\_x;**



диапазон  $y$ ). Полученное значение можно округлить с учетом числа значащих разрядов в исходных данных. Результаты расчёта приведены на рис. 2.2.

$f_x$  =ОКРУГЛ(КОРРЕЛ(D3:D102;E3:E102);3)

C	D	E	F	G	H	I
<b>исходные значения</b>	<b>Округленные значения</b>					
<b>y</b>	<b>Зарплата</b>	<b>Потребление</b>			<b>Зарплата</b>	<b>Потребление</b>
1510,832661	3016	1511		Зарплата	1	
4218,772125	4202	4219		Потребление	0,915899763	1
3985,013782	4331	3985				
6157,277552	9206	6157				
3806,645996	3333	3807				
5919,477414	8586	5919				
4314,114081	5766	4314				
6629,319981	9393	6629				

Показатель	Значение
Регрессия	0,916
a0	
a1	

Рис. 2.2. Пример вычисления коэффициентов корреляции

### 3.3. Упражнение 3. Регрессионный анализ

На рис. 2.3 показан пример вычисления промежуточных значений, таких как  $\sum x^2$ ,  $\sum x^3$  и т.д. Используется функция **СУММПРОИЗВ**, которая позволяет вычислить сумму попарных произведений нескольких столбцов.

$f_x$  =СУММПРОИЗВ(D4:D103;D4:D103;E4:E103)

C	D	E	F	G	H
				<b>Вычисление промежуточных сумм</b>	
1	2957,700105	3871	2958	n	100
1	6598,804167	9116	6599	Сумма x	628956
2	3353,568187	3997	3354	Сумма y	462708
3	2195,145292	3476	2195	Сумма x <sup>2</sup>	4426918638
7	6678,484727	8276	6678	Сумма x <sup>3</sup>	3,3934E+13
1	5978,923053	8847	5979	Сумма x <sup>4</sup>	2,76164E+17
3	3999,007407	3658	3999	Сумма xy	3152595054
5	5649,97021	6848	5650	Сумма yx <sup>2</sup>	2,3598E+13
2	5251,494479	8333	5251		

Рис. 2.3. Пример вычисления промежуточных сумм

Для работы с матрицами в пакете Excel используются функции, работающие с массивами. Матричные функции вводят в диапазон ячеек, как описано ниже. После ввода матричных функций, они автоматически отображаются в фигурных скобках.

На рис. 2.4 приведён пример матриц. Для нахождения обратной матрицы используется функция **МОБР (матрица\_Z)**, для умножения матриц – функция **МУМНОЖ (матрица\_Z<sup>-1</sup>; матрица\_Y)**.

	Z		A		
	0,118880678	-1,67213E-05	=МУМНОЖ(H18:I19;G15:G16)		
	-1,67213E-05	2,56796E-09			

15	прямая				
16	Y	Z	Z-1	A	
17	718089	100	718089	0,127976954	-1,64293E-05
18	5593596543	718089	5593596543	-1,64293E-05	2,28792E-09

20	парабола				
21	Y	Z	Z-1	A	
22	718089	100	718089	5593596543	1,350055314
23	5593596543	718089	5593596543	4,65092E+13	-0,000383858
24	4,65092E+13	5,594E+09	4,65092E+13	4,06818E+17	2,53215E-08

Рис. 2.4. Пример работы с матрицами

Данные функции возвращают в качестве результата не одно значение, а массивы чисел (диапазон ячеек). Для того чтобы получить результат, выполните следующие действия:

- выберите диапазон ячеек, в котором будет располагаться матрица, являющаяся результатом вычислений матричной функции;
- введите формулу в ячейку, являющуюся левым верхним углом выбранного диапазона, нажать **Enter**;
- выделите область ячеек (выбранный диапазон), см. рис. 2.4;
- нажмите **F2**;
- нажмите **Ctrl+Shift+Enter**.

Коэффициенты регрессии можно также найти с помощью функции **ЛИНЕЙН** или статистической надстройки (функция «Регрессия»). Однако, в этом случае необходимо создать дополнительные столбцы данных, содержащие значения регрессоров, например,  $x^2$  и  $x^3$ , т.к. данные функции вычисляют коэффициенты лишь для линейной множественной регрессии. Столбцы регрессоров должны располагаться в соседних колонках электронной таблицы, при этом они указываются в виде диапазона как второй аргумент функции **ЛИНЕЙН (известные\_значения\_y; известные\_значения\_x;)**. Вводя дополнительные переменные,

равные  $x^2$  и  $x^3$ , уравнения параболы и кубической параболы можно преобразовать к линейному виду:

$$\bar{y}_x = a_2x^2 + a_1x + a_0 \Rightarrow \bar{y}_x = a_2z + a_1x + a_0, \text{ где } z = x^2;$$

$$\bar{y}_x = a_3x^3 + a_2x^2 + a_1x + a_0 \Rightarrow \bar{y}_x = a_3t + a_2z + a_1x + a_0,$$

где  $z = x^2, t = x^3$ .

В случае множественной регрессии, функции **ЛИНЕЙН** в качестве параметра **известные\_значения\_x** надо указывать диапазон, содержащий сразу все переменные (например, оба столбца:  $x$  и  $z$  в случае параболы).

Аналогично. При вызове функции «Регрессия» статистической надстройки, надо указывать аналогичные диапазоны в качестве входного интервала  $x$ .

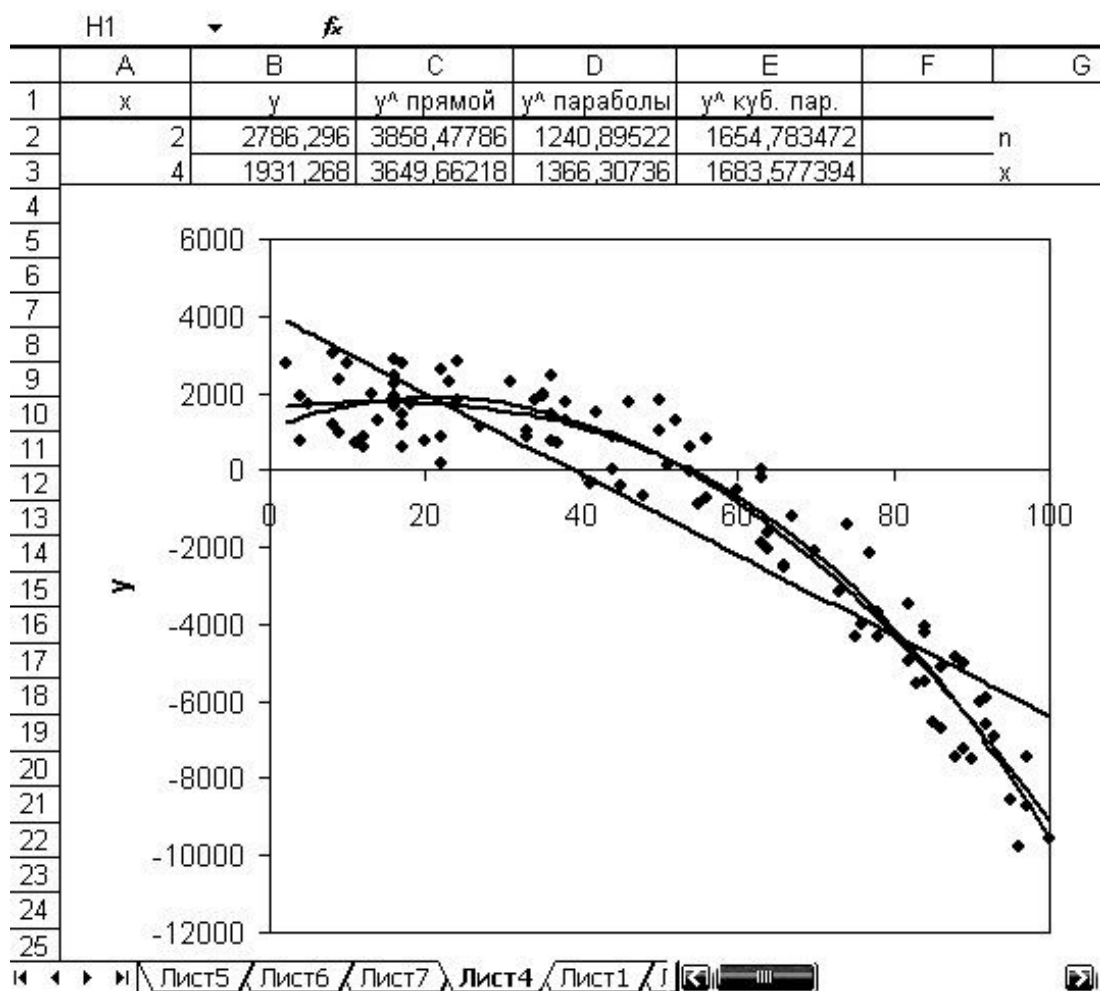


Рис. 2.5. Пример корреляционного поля с линиями регрессии

При изучении взаимосвязей, необходимо построить диаграмму разброса (корреляционное поле): меню **[Вставка → Диаграмма]**. На этой диаграмме исходные данные ( $x, y$ ) показаны

точками. Сюда же наносится линия регрессии. Для этого необходимо сформировать вспомогательные столбцы  $x$  и  $\bar{y}_x$  для каждого вида регрессии. Столбец вспомогательных значений факторного признака  $x$  должен содержать несколько значений с постоянным шагом от минимального до максимального. Для этого в первую ячейку вводим начальное значение, выбираем диапазон значений и вызываем [Правка → Заполнить → Прогрессия]. При этом нужно выбрать вид заполнения **По столбцам**, вид прогрессии **Арифметическая**, **Шаг**, **Предельное значение**. Количество вспомогательных промежуточных значений фактора выбирают таким образом, чтобы получить на графике гладкую кривую линию.

Тип диаграммы для линий регрессии – **Точечная диаграмма со значениями, соединёнными сглаживающими линиями без маркеров**, см. рис. 2.5.

### 3.4. Упражнение 4. Условное среднее

Условное среднее  $\bar{y}_x$  – это среднее арифметическое значений результативного признака  $y$  при условии, что соответствующие значения факторного признака  $x$  попадают в заданный интервал. Добавьте интервалы по  $x$ , которые выбираются по общим правилам группировки данных (см. лабораторная работа №1).

Для нахождения условного среднего можно использовать функцию **СУММЕСЛИ**, которая позволяет вычислить сумму при выполнении заданного условия. Формат функции следующий:

**СУММЕСЛИ (диапазон ; критерий ; диапазон\_суммирования) .**

**Диапазон** – ячейки, значения которых проверяются с помощью условия;

**Критерий** – условия суммирования, например, "**<=**" &W2;

**Диапазон\_суммирования** – ячейки, значения которых складывают при выполнении условия.

Полученная сумма делится на количество элементов, попадающих в диапазон. Для этого используется функция **СЧЕТЕСЛИ**.

Нижняя граница X	Верхняя граница X	Среднее X	Условное среднее Y
3000	4000	3500	3229
4000	5000	4500	3676
5000	6000	5500	4166
6000	7000	6500	4929
7000	8000	7500	5385
8000	9000	8500	5778
9000	10000	9500	6246

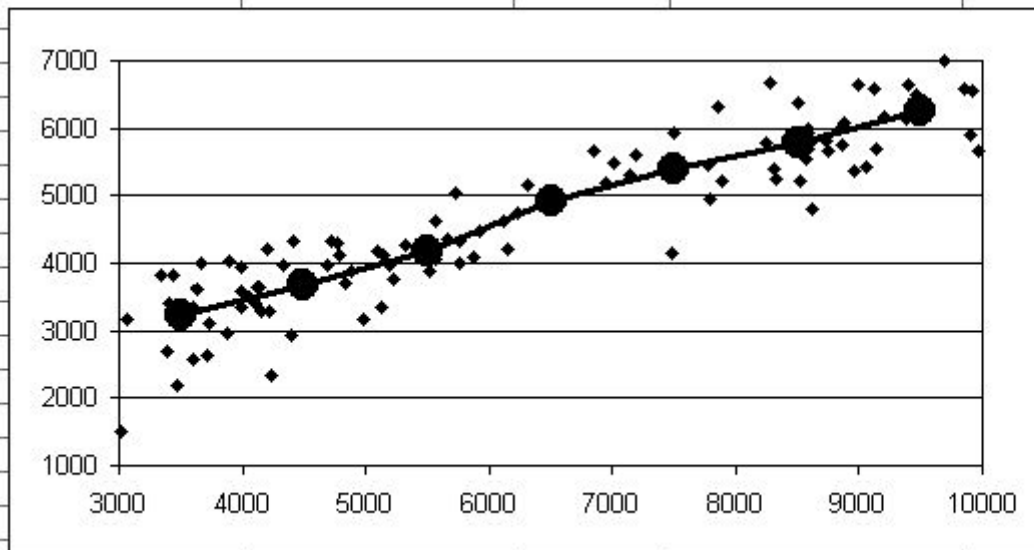


Рис. 2.6. Пример корреляционного поля с линией условного среднего

Формула для расчета условного среднего может быть построена следующим образом:

$$= \frac{(\text{СУММЕСЛИ}(\$E\$3:\$E\$102;"<="&W2;\$F\$3:\$F\$102) - \text{СУММЕСЛИ}(\$E\$3:\$E\$102;"<="&V2;\$F\$3:\$F\$102))}{(\text{СЧЁТЕСЛИ}(\$E\$3:\$E\$102;"<="&W2) - \text{СЧЁТЕСЛИ}(\$E\$3:\$E\$102;"<="&V2))}$$

Линия условного среднего (эмпирическая регрессия) наносится на корреляционное поле, см. рис. 2.6. В качестве значений  $x$  берутся середины интервалов, точки соединяются прямыми линиями.

### 3.5. Упражнение 5. Анализ качества модели связи

Для анализа полученной модели связи используют показатель остаточной дисперсии

$$\sigma_{\text{ост}}^2 = \frac{\sum (y_i - \bar{y}(x_i))^2}{n - k - 1},$$

где  $n$  – объём выборки,  $k$  – число коэффициентов уравнения регрессии. Остатки – это разность между фактическим значением

(точками на графике) и теоретическим прогнозом (линией регрессии). Учет числа коэффициентов  $k$  компенсирует постепенное приближение линии регрессии к исходным точкам на корреляционном поле за счет повышения порядка модели. Рекомендуется выбирать уравнение регрессии, дающее наименьшую остаточную дисперсию.

### **Требования к содержанию и оформлению отчета**

Отчет должен быть продемонстрирован как на бумажном носителе, содержащем статистические показатели, коэффициенты уравнений регрессии, представленные в виде таблицы, корреляционное поле с линиями регрессии, так и в электронной форме в виде файла с заполненной таблицей и графиками.

Выводы по результатам анализа взаимосвязи социально-экономических явлений могут содержать следующие положения:

- какая информация исследовалась и какими методами;
- есть ли линейная связь между показателями;
- совпадают ли результаты, полученные разными способами;
- какие прогнозы и рекомендации можно сделать.

Титульный лист отчета должен содержать всю информацию, необходимую для однозначной идентификации авторов и работы. Для этого на титульном листе указывают название дисциплины, тему и номер работы, вариант задания, номер группы, фамилии и инициалы студентов, должность, фамилию и инициалы преподавателя и т.п. (в соответствии со стандартом УГАТУ на оформление текстовых документов).

### **Порядок выполнения работы**

1. Ознакомьтесь с описанием следующих функций Excel:
2. **КОРРЕЛ**, **ОКРУГЛ**, **СУММПРОИЗВ**, **МОБР**, **МУМНОЖ**, **ЛИНЕЙН**, **СУММЕСЛИ**.
3. Сгенерируйте исходные данные.
4. Вычислите коэффициент корреляции с помощью надстройки и функции **КОРРЕЛ**.
5. Сделайте вывод о тесноте связи факторного и результативного признака.

6. Вычислите коэффициенты уравнений регрессии первого, второго и третьего порядков с помощью матричных функций, функции **ЛИНЕЙН** и надстройки.
7. Постройте корреляционное поле.
8. Нанесите на корреляционное поле линии регрессии.
9. Нанесите на корреляционное поле линию эмпирической регрессии.
10. Рассчитайте значение остаточной дисперсии для каждого уравнения регрессии.
11. Оформите отчёт в соответствии с требованиями.

### **Контрольные вопросы**

1. Что такое корреляционная зависимость?
2. Что изучает корреляционный анализ?
3. Как коэффициент корреляции характеризует взаимосвязь параметров?
4. Что изучает регрессионный анализ?
5. Что такое регрессия?
6. Как определить параметры уравнения регрессии?
7. Как определить оптимальный вид уравнения регрессии?
8. Что такое эмпирическая регрессия?
9. Что такое условное среднее?
10. Что такое корреляционное поле?

### **Критерии результативности выполнения работы**

Лабораторная работа считается выполненной в том случае, если студент

1. выполнил все указанные задания, следуя порядку выполнения работы;
2. освоил методику выполнения типовых заданий и способен продемонстрировать работу программы;
3. представил отчет, содержащий статистические показатели и коэффициенты уравнений регрессии и корреляционное поле с линиями регрессии;
4. ответил на все контрольные и дополнительные вопросы.

# ЛАБОРАТОРНАЯ РАБОТА №3

## РЯДЫ ДИНАМИКИ

### 1. ЦЕЛЬ РАБОТЫ И ЗАДАЧИ

Освоить методы статистического исследования динамики социально-экономических явлений.

Задачами работы является:

- закрепление навыков генерации рядов динамики по аддитивной и мультипликативной моделям;
- овладение методикой построения уравнения и линии тренда, и сглаживания ряда динамики;
- приобретение навыков построения биржевых графиков.

### 2. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

*Динамика* (в статистике) – это изменение явления во времени.  
*Ряд динамики* (временной ряд) – последовательность значений показателя, привязанных к моментам или интервалам времени.  
Отдельное значение показателя – *уровень ряда*.

При анализе динамики рассматриваются три составные части (компонента) ряда динамики:

$T$  – тренд (основная тенденция);

$S$  – колебания;

$E$  – случайная составляющая.

Чаще всего используются две модели рядов динамики (модели сезонности):

аддитивная (сумма):  $y_t = T + S + E$ ;

мультипликативная (произведение):  $y_t = T \cdot S \cdot E$ .

Аддитивная модель строится по следующим формулам:

$$T = a + b \cdot t;$$

$$S = c \cdot \sin(0,5236 \cdot t);$$

$$E = d \cdot N(0;1);$$

$$y_t = T + S + E.$$

Мультипликативная модель строится по следующим формулам:

$$T = a + b \cdot t;$$

$$S = 1 + f \cdot \sin(0,5236 \cdot t);$$

$$E = 1 + g \cdot N(0;1);$$



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

Сглаживание временных рядов – это метод локального усреднения, позволяющий изучать общий характер динамики.

При построении скользящей средней находят среднее значение на интервале фиксированной длины, причём начало интервала смещается по времени. Полученное сглаженное значение «привязывают» к середине интервала по времени.

Интервал сглаживания выбирают так, чтобы удалить случайные отклонения, но сохранить тренд и колебания. Сглаживание с периодом 12 месяцев удаляет не только случайную составляющую, но и сезонные колебания.

### 3. МЕТОДИКА ВЫПОЛНЕНИЯ РАБОТЫ

#### 3.1 Упражнение 1. Компоненты динамики

Для знакомства с моделями динамики сгенерируйте ряды динамики и постройте их графики согласно варианту задания, см. табл. 3.1. Используйте разные листы Excel для работы с аддитивной и мультипликативной моделями.

Создайте следующие столбцы:

$t$  – время, мес.,  $t = 1 \dots n$ ;

$T$  – тренд;

$S$  – сезонные колебания;

$E$  – случайная составляющая;

$y(t)$  – ряд динамики.

Таблица 3.1

Варианты параметров моделей динамики

№	$n$	$a$	$b$	$c$	$d$	$f$	$g$
1	60	2	0,3	3	0,5	0,3	0,05
2	70	3	0,2	2	0,5	0,5	0,07
3	80	4	0,1	2	0,5	0,4	0,06
4	90	5	0,2	4	0,6	0,3	0,06
5	100	6	0,3	3	0,4	0,3	0,06
6	110	7	0,5	4	0,8	0,2	0,05

Для создания столбца времени  $t$  введите начальное значение  $t = 1$ . Затем выделите столбец и заполните его, выбрав **Правка** → **Заполнить** → **Прогрессия** → **Арифметическая**. Введите  $n$  в окне **Предельное значение**.

Для создания сезонной составляющей используйте функцию **SIN**.

Случайные числа с нормальным распределением  $N(0;1)$  сгенерируйте с помощью статистической надстройки анализа данных: **Сервис** → **Анализ данных** → **Генерация случайных чисел**. Укажите нулевое среднее значение и единичную дисперсию.

Постройте график ряда динамики. Настройте свойства диаграммы, чтобы изображение занимало все доступное поле графика, а масштаб выражался круглыми числами, см. рис. 3.1.

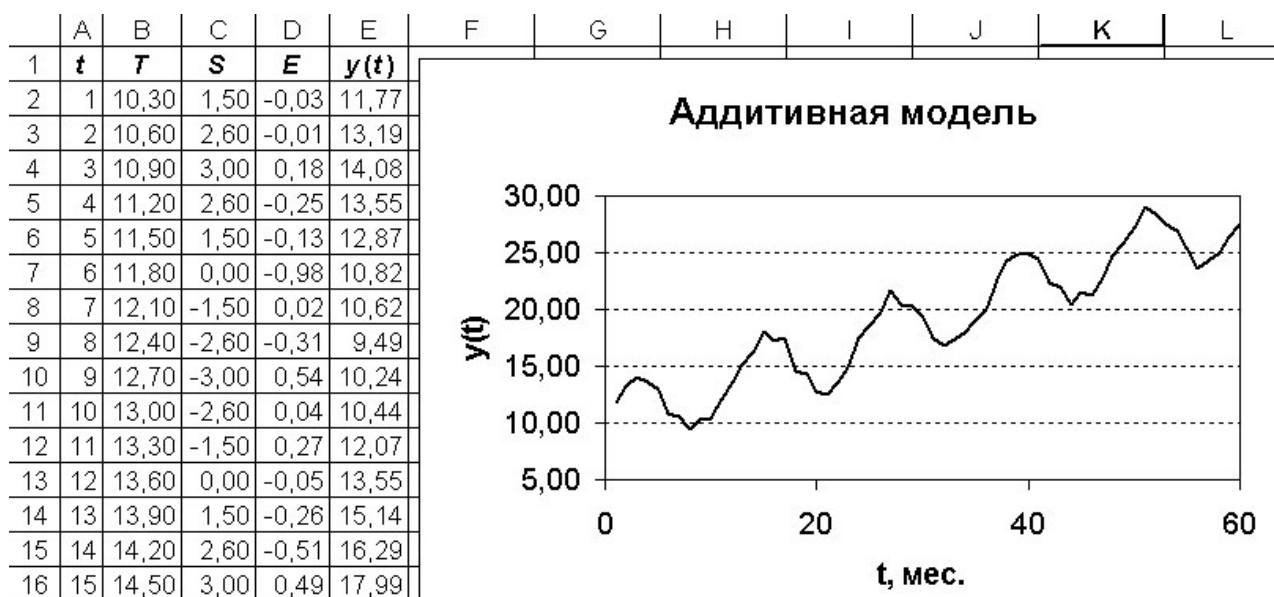


Рис. 3.1. Исходные данные и диаграмма динамики

Сравните полученные графики и сделайте вывод об изменении амплитуды сезонных колебаний в разных моделях сезонности.

### 3.2. Упражнение 2. Загрузка исторических данных

В качестве исходных данных используются дневные и недельные ряды котировок акций различных эмитентов на фондовой бирже ММВБ.

## Варианты заданий

№	Имя файла	Описание
1	GAZP_DAY.TXT	Газпром (дневные)
2	GAZP_WEEK.TXT	Газпром (недельные)
3	GMKN_DAY.TXT	ГМК Норильский никель (дневные)
4	GMKN_WEEK.TXT	ГМК Норильский никель (недельные)
5	ROSN_DAY.TXT	Роснефть (дневные)
6	ROSN_WEEK.TXT	Роснефть (недельные)
7	SBERP_DAY.TXT	Сбербанк (дневные)
8	SBERP_WEEK.TXT	Сбербанк (недельные)
9	VTBR_DAY.TXT	ВТБ (дневные)
10	VTBR_WEEK.TXT	ВТБ (недельные)

Загрузите файл в соответствии с вариантом задания, см. табл. 3.2. При попытке открыть текстовый файл **\*.TXT** через меню **Файл** → **Открыть** автоматически запускается **Мастер импорта** текстовых файлов MS Excel. В процессе настройки мастера в нижней части окна можно видеть **Образец разбора данных** и контролировать правильность импорта.

Используйте следующие настройки:

**Шаг 1 из 3: Формат исходных данных:** → с разделителями.

**Шаг 2 из 3: Символом-разделителем является:** → запятая.

**Шаг 3 из 3: Подробнее** → **Разделитель целой и дробной части** → **Точка**, см. рис. 3.2.

Убедитесь, что при импорте данных не была искажена информация. Для этого визуально сравните исходный текстовый файл и таблицу Excel.

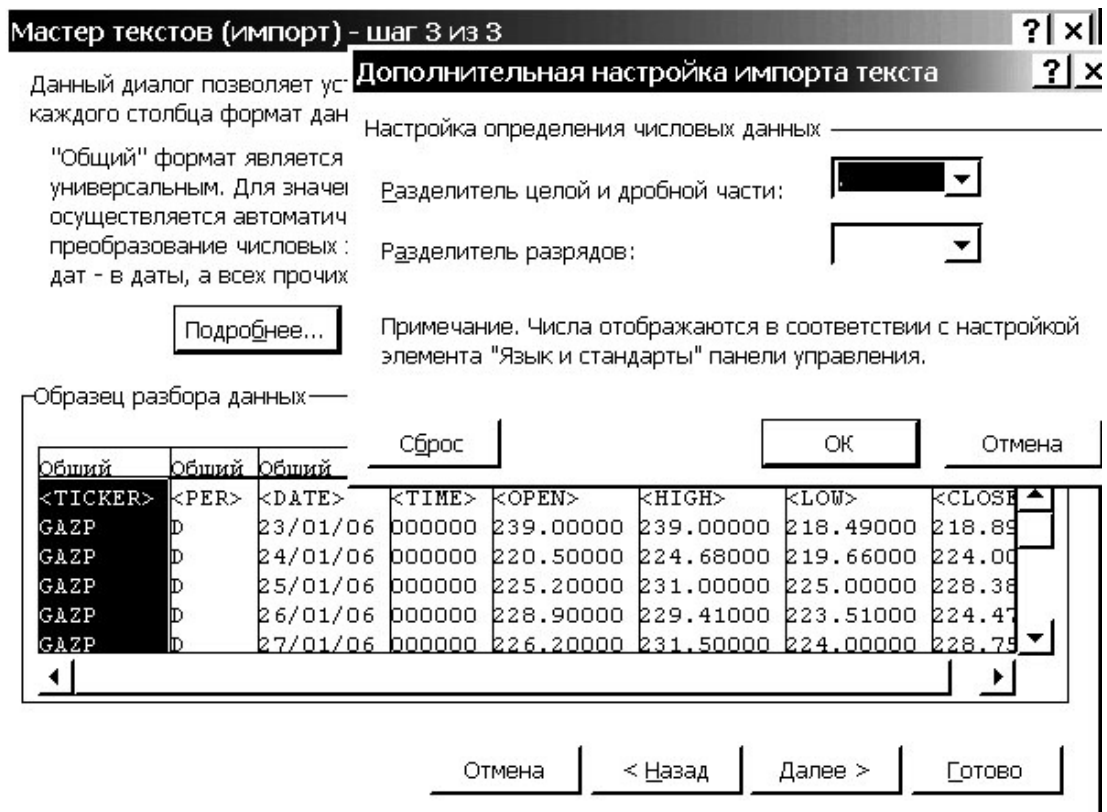


Рис. 3.2. Настройка мастера импорта текстовых файлов

После завершения импорта удалите малоинформативные колонки **TICKER**, **PER**, **TIME** и сохраните данные в формате **\*.XLS**.

### 3.3. Упражнение 3. Анализ тренда

Постройте график по ценам закрытия **CLOSE** в зависимости от времени: **Вставка** → **Диаграмма** → **Тип** → **Точечная** → **Без маркеров**.

Настройте диаграмму так, чтобы оптимально использовать всё поле графика, см. рис. 3.3. При анализе данных за несколько лет можно использовать сокращенное представление даты в формате «месяц–год».

Для построения линии тренда щелкните правой кнопкой мыши по линии графика. В контекстном меню выберите пункт **Добавить линию тренда**. В диалоговом окне выберите вид аппроксимирующей функции, см. рис. 3.4.

Уравнение тренда можно построить методами регрессионного анализа. При работе с историческими данными можно создать вспомогательный столбец времени  $t$ . Постройте уравнение тренда с помощью функции **ЛИНЕЙН** и с помощью статистической надстройки **Анализ данных**, см. описание лабораторной работы №2. На основе

уравнения тренда сформируйте столбец линейного прогноза  $T(t)$ .  
 Затем добавьте этот ряд в исходные данные диаграммы с привязкой к фактическим датам (столбец **DATE**).

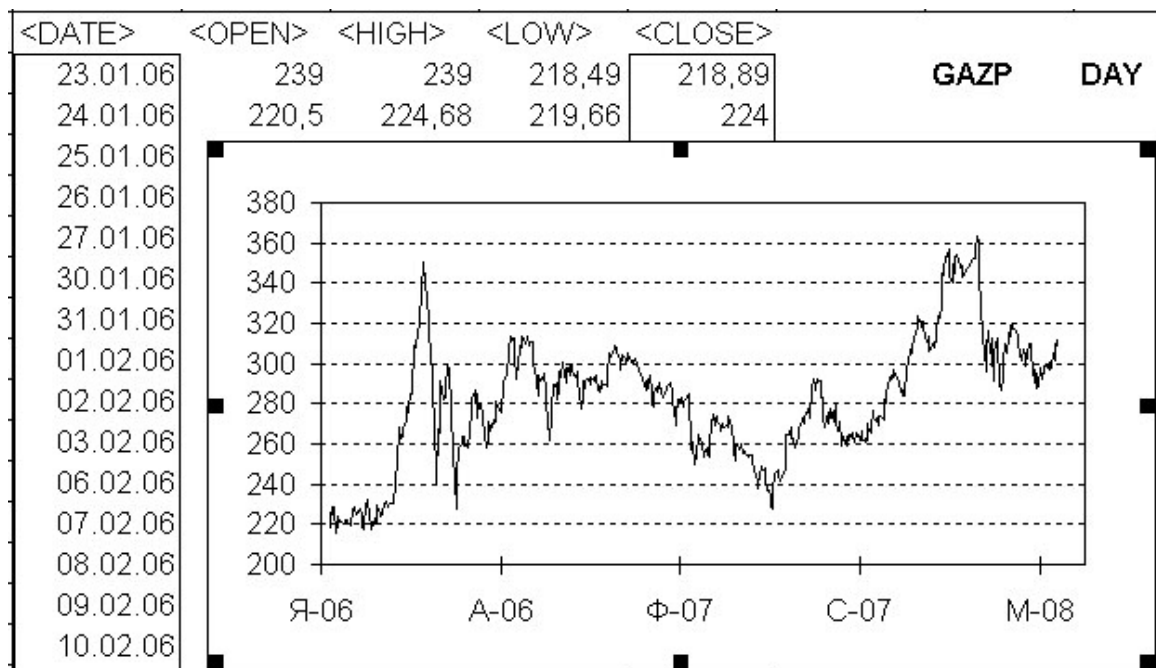


Рис. 3.3. Настройка диаграммы динамики

**Линия тренда**

Тип | Параметры

Построение линии тренда (аппроксимация и сглаживание)

Линейная

Логарифмическая

Полиномиальная

Степенная

Экспоненциальная

Линейная фильтрация

Степень: 2

Точки: 2

Построен на ряде: **CLOSE**

Рис. 3.4. Выбор характера тренда

### 3.4. Упражнение 4. Скользящая средняя простая

Для построения простой скользящей средней (СС) имеются несколько инструментов:

- функция **СРЗНАЧ (интервал\_сглаживания\_y)**;
- надстройка **Анализ данных** → **Скользящее среднее**;
- контекстное меню **Добавить линию тренда** → **Линейная фильтрация**.

При построении СС теряются несколько первых и последних уровней ряда.

Постройте СС с помощью функции **СРЗНАЧ** и отнесите результат сглаживания к середине интервала. Линия тренда будет соответствовать линии регрессии в традиционном понимании.

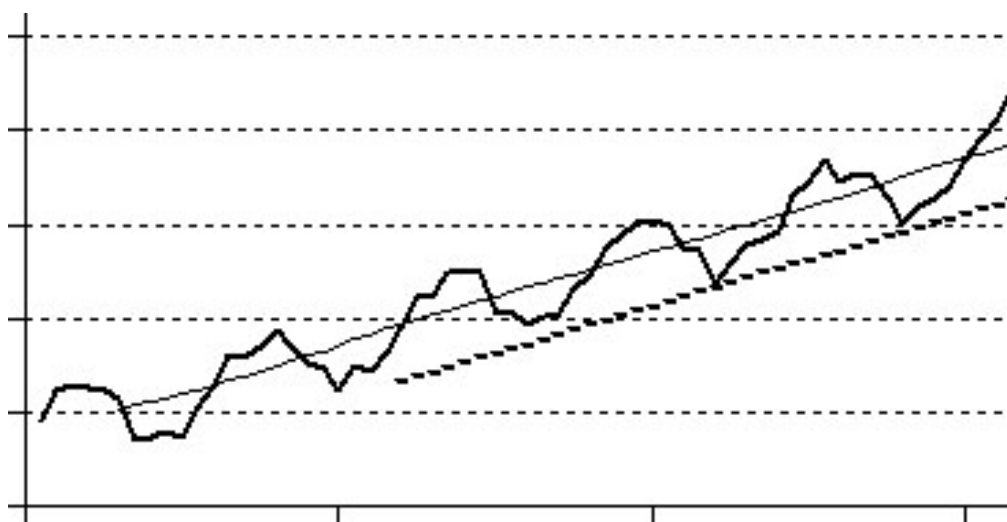


Рис. 3.5. Линия регрессии и линия поддержки

Постройте СС с помощью надстройки **Анализ данных**. В качестве выходного интервала укажите первую ячейку выходного столбца. Обратите внимание на автоматическую привязку к времени и диапазон неопределенных сглаженных значений. Нанесите линию на график исходного ряда. Подберите интервал сглаживания так, чтобы линия «повышательного» тренда выступила в роли линии поддержки, см. рис. 3.5. В этом случае локальные минимумы ряда динамики касаются линии поддержки.

При анализе тренда вместо среднего арифметического может использоваться медиана, см. функцию **МЕДИАНА (диапазон)**.

### 3.5 Упражнение 5. Скользящая средняя взвешенная

При построении взвешенной средней каждое усредняемое значение учитывается со своим весовым коэффициентом (весом). Сумма взвешенных значений делится на сумму весов, например:

$$\bar{y}_t = \frac{y_{t-1} \cdot w_{t-1} + y_t \cdot w_t + y_{t+1} \cdot w_{t+1}}{w_{t-1} + w_t + w_{t+1}}.$$

Часто используют биномиальные весовые коэффициенты, см. рис. 3.6. Например, усреднение по 5 точкам выглядит следующим образом:

$$\bar{y}_i = \frac{y_{i-2} + 4y_{i-1} + 6y_i + 4y_{i+1} + y_{i+2}}{16}.$$

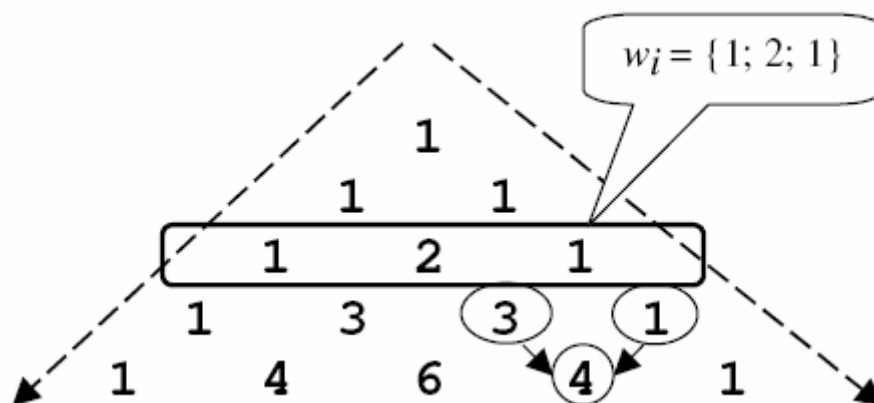


Рис. 3.6. Биномиальные весовые коэффициенты

### 3.6. Упражнение 6. Экспоненциально взвешенная скользящая средняя

Сглаженное значение  $\bar{y}_t$  вычисляется по текущему уровню ряда  $y_t$  и предыдущему сглаженному уровню  $\bar{y}_{t-1}$ :

$$\bar{y}_t = \alpha \cdot y_t + (1 - \alpha) \bar{y}_{t-1},$$

где  $\alpha$  – параметр сглаживания:

$$\alpha < 1;$$

$$\alpha = \frac{2}{n+1};$$

$$n = \frac{2 - \alpha}{\alpha} \text{ – эквивалентный период сглаживания ССП.}$$

Если преобразовать приведенные выше соотношения к стандартному виду скользящей средней взвешенной, то весовые коэффициенты будут убывать по экспоненте.

Сглаженные значения привязывают к концу интервала, см. рис.3.7.

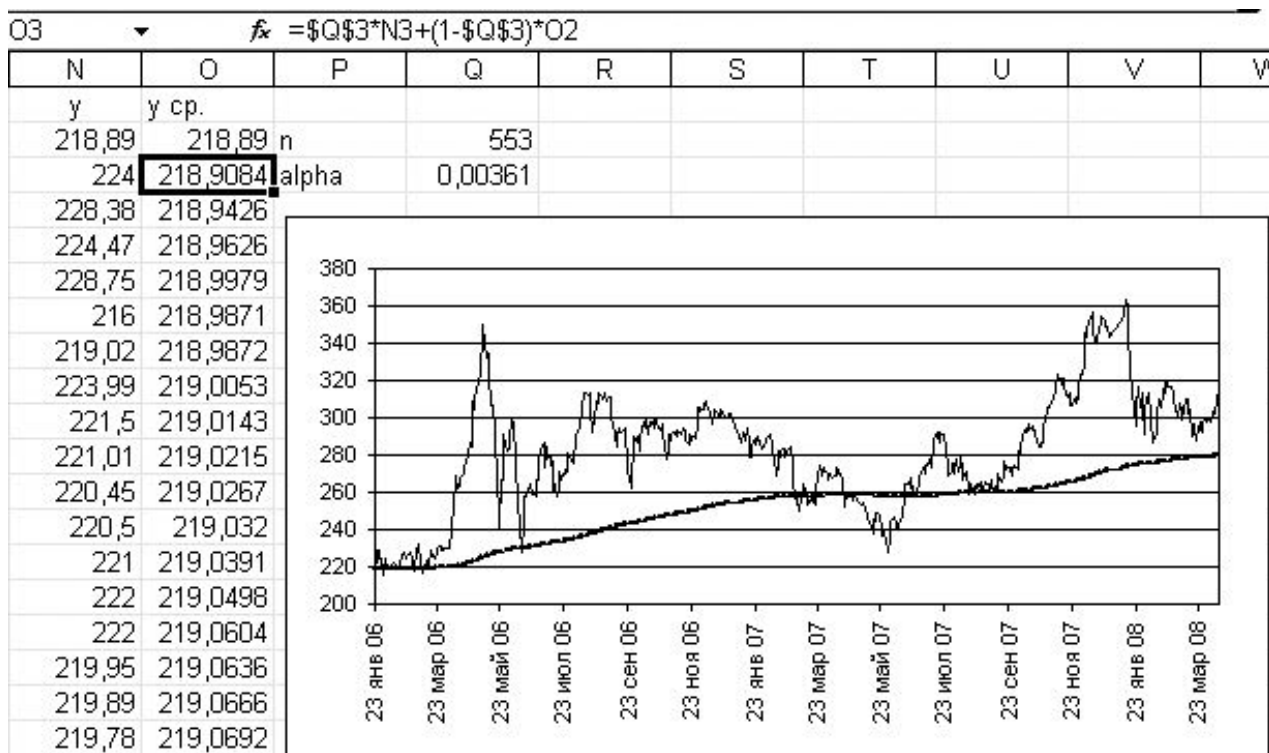


Рис. 3.7. Экспоненциальные весовые коэффициенты

### 3.7. Упражнение 7. Биржевые графики

По историческим данным о котировках акций в MS Excel можно построить два вида биржевых графиков: свечи и бары (штриховой график). Для этого вызывается меню **Вставка** → **Диаграмма** → **Биржевая**, см. рис. 3.8.

На диаграмме японских свечей отображаются четыре значения цены за интервал времени:

- OPEN – цена открытия (на начало интервала времени);
- HIGH – максимальная цена за интервал времени;
- LOW – минимальная цена за интервал времени;
- CLOSE – цена закрытия (на конец интервала времени).

Если цена закрытия выше, чем цена открытия (рынок растет), то тело свечи белое. При падении цены тело черное, см. рис. 3.9.





Рис. 3.8. Выбор вида биржевой диаграммы

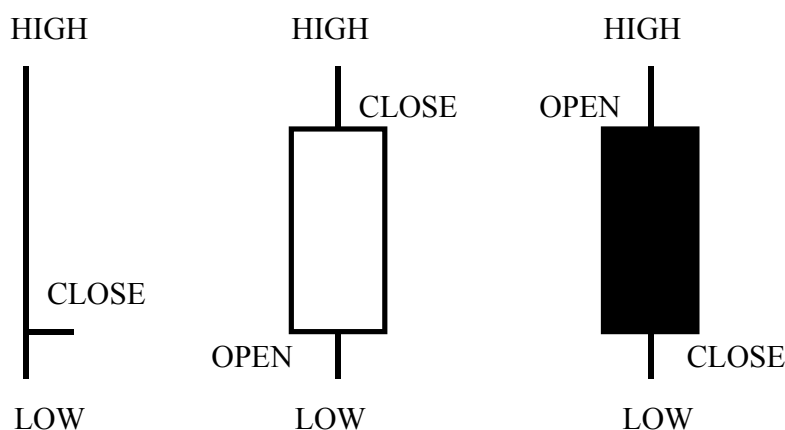


Рис. 3.9. Биржевые диаграммы: бар, растущая свеча, падающая свеча

На диаграмме баров MS Excel отображаются три значения цены за интервал времени: HIGH, LOW, CLOSE.

Таким образом, биржевые диаграммы отображают не только движение цены, но и размах значений за интервал времени.

После построения диаграммы можно настроить толщину линий. Выберите различные диапазоны по времени, чтобы проанализировать изменение направления динамики цен.

Торговля на фондовой бирже не проводится по выходным и праздничным дням, поэтому на графике появляются «пробелы» во времени – от двух до десяти дней. Поэтому при анализе можно удалить столбец времени из исходных данных диаграммы и использовать только порядковый номер свечи, см. рис.3.11.

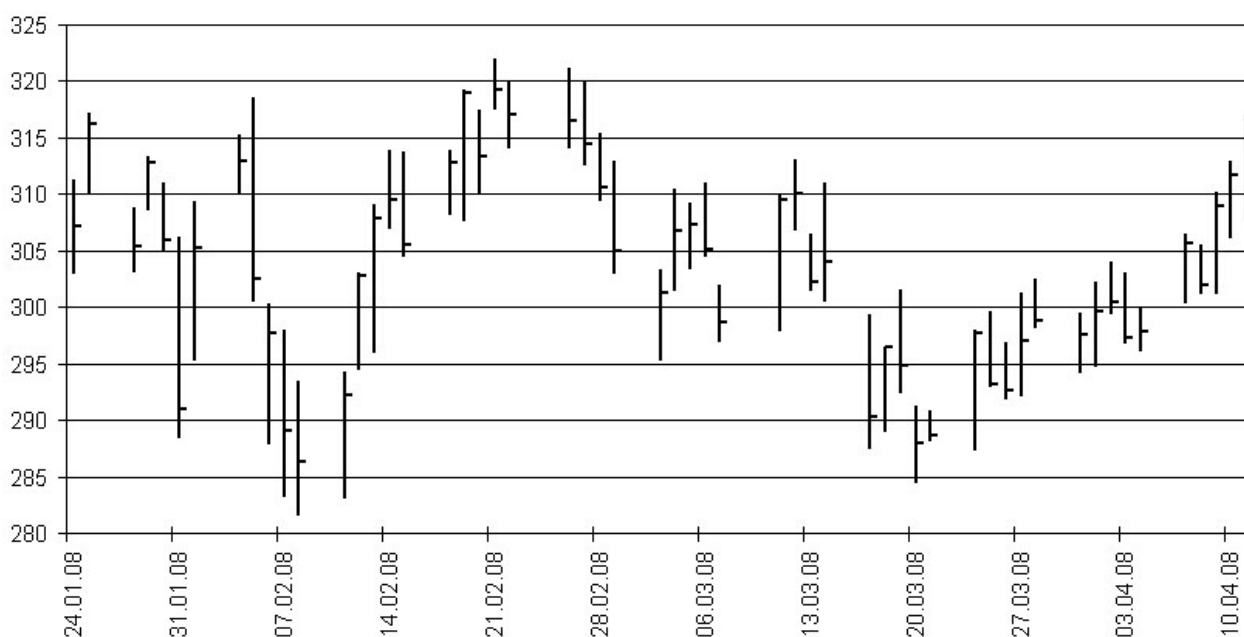


Рис. 3.10. Диаграмма баров с привязкой к календарным датам

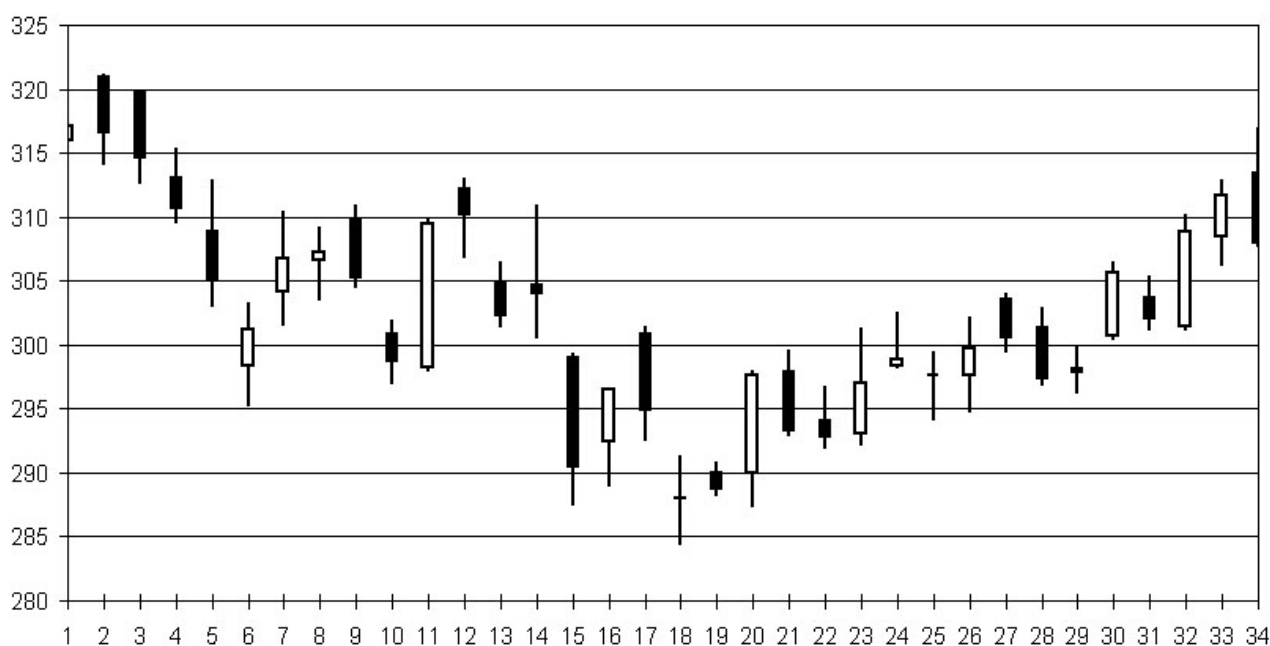


Рис. 3.11. Диаграмма японских свечей без указания времени

### Требования к содержанию и оформлению отчета

Отчет должен быть продемонстрирован как на бумажном носителе, содержащем графики исходного ряда, сглаженного ряда, тренда и уравнения моделей, так и в электронной форме в виде файла с заполненной таблицей и биржевыми графиками.

Выводы по результатам анализа динамики социально-экономических явлений могут содержать следующие положения:

- какая информация исследовалась и какими методами;
- какова основная тенденция (тренд) в данных;
- совпадают ли результаты, полученные разными способами;
- какие прогнозы и рекомендации можно сделать.

Титульный лист отчета должен содержать всю информацию, необходимую для однозначной идентификации авторов и работы. Для этого на титульном листе указывают название дисциплины, тему и номер работы, вариант задания, номер группы, фамилии и инициалы студентов, должность, фамилию и инициалы преподавателя и т.п. (в соответствии со стандартом УГАТУ на оформление текстовых документов).

### **Порядок выполнения работы**

1. Сгенерируйте ряды динамики по аддитивной и мультипликативной моделям согласно варианту и постройте графики.
2. Загрузите исторические данные согласно варианту задания и постройте график по ценам закрытия.
3. С помощью контекстного меню добавьте линию тренда на все три графика. Подберите наиболее подходящий тип тренда.
4. Постройте уравнение тренда с помощью функции **ЛИНЕЙН** и с помощью статистической надстройки и нанесите линию тренда на график исходных данных.
5. Проведите сглаживание ряда динамики с аддитивной моделью сезонности с помощью простой скользящей средней (тремя способами) и с помощью скользящей медианы. Подберите периоды усреднения для удаления случайной составляющей и для удаления сезонности.
6. Постройте биржевые графики баров и свечей по разным диапазонам времени (месяц, год).

### **Контрольные вопросы**

1. Что такое динамика?
2. Какие компоненты изучают при анализе рядов динамики?
3. Чем отличаются графики аддитивной и мультипликативной моделей?
4. Что такое тренд?
5. Как построить уравнение тренда?

6. Что такое скользящая средняя и скользящая медиана?
7. Для чего сглаженный ряд относят к середине или к концу интервала усреднения?
8. Что изображается на биржевых графиках баров и свечей?

### **Критерии результативности выполнения работы**

Лабораторная работа считается выполненной в том случае, если студент

1. выполнил все указанные задания, следуя порядку выполнения работы, и представил отчет, содержащий графики исходного ряда, сглаженного ряда, тренда, биржевые графики и уравнения моделей;
2. освоил методику выполнения типовых заданий и способен продемонстрировать работу программы;
3. ответил на все контрольные и дополнительные вопросы.

### **СПИСОК ЛИТЕРАТУРЫ**

1. Статистика / Под ред. Назарова М.Г.– М.:Кнорус, 2006.– 480 с.
2. Теория статистики / Под ред. Р.А.Шмойловой.– М.: Финансы и статистика, 2007.– 656 с.
3. Практикум по теории статистики / Под ред. Р.А.Шмойловой.– М.: Финансы и статистика, 2007.– 415 с.
4. *Вадзинский Р. Н.* Статистические вычисления в среде Excel.– СПб.: Питер, 2008.– 608 с.
5. *Веденеева Е.Е.* Функции и формулы Excel 2007.– СПб.: Питер, 2008.– 384 с.
6. *Козлов А.Ю.* Пакет анализа MS Excel в экономико-статистических расчетах: учебное пособие для студентов вузов.– М.: ЮНИТИ-ДАНА, 2003 .– 139 с.

Составители АРЬКОВ Валентин Юльевич,  
ЖЕВАК Алексей Владимирович,  
АБДУЛНАГИМОВ Ансаф Ирекович

## СТАТИСТИЧЕСКИЙ АНАЛИЗ СРЕДСТВАМИ ПАКЕТА MS EXCEL

Лабораторный практикум  
по дисциплине «Статистика»

Подписано в печать                      Формат 60×84 1/16

Бумага офсетная. Печать плоская. Гарнитура Times New Roman.

Усл. печ. л.      Усл. кр.-отг.      Уч.-изд.л.

Тираж 100 экз. Заказ № \_\_\_\_\_

ГОУ ВПО Уфимский государственный авиационный технический университет

Центр оперативной полиграфии УГАТУ

450000, Уфа-центр, ул. К.Маркса, 12