Глава 5

Надстройка Пакет анализа

В состав Microsoft Excel входит надстройка Пакет анализа, которая содержит 19 статистических процедур и около 50 функций. Функции в основном относятся к категориям инженерных и финансовых и поэтому здесь не рассматриваются. Статистические процедуры, содержащиеся в надстройке Пакет анализа, предоставляют широкий спектр средств для статистического анализа начиная от простой описательной статистики или сглаживания данных и заканчивая анализом Фурье и проведением различных тестов. Полный список этих средств и их краткое описание представлены в табл. 5.1 (названия средств приводятся в соответствии со списком из диалогового окна Анализ данных).

Таблица 5.1. Статистические средства надстройки Пакет анализа

Средство	*Описание*
Однофакторный дисперсионный анализ	Используется для проверки гипотезы о равенстве математических ожиданий двух или более выборок
Двухфакторный дисперсионный анализ без повторений	Двухфакторный дисперсионный анализ на основе одной выборки
Двухфакторный дисперсионный анализ с повторениями	Двухфакторный дисперсионный анализ на основе нескольких выборок
Корреляция	Вычисляет корреляционную матрицу
Ковариация	Вычисляет матрицу ковариаций
Описательная статистика	Создает отчет, содержащий статистические характеристики представленной выборки
Экспоненциальное сглаживание	Реализует метод экспоненциального сглаживания данных
Двухвыборочный F-тест для дисперсий	Применяется для сравнения дисперсий двух генеральных совокупностей
Анализ Фурье	Реализует метод быстрого преобразования Фурье (БПФ) для анализа данных
Гистограмма	Используется для анализа распределения выборочных данных и построения гистограмм
Скользящее среднее	Используется для сглаживания данных
Генерация случайных чисел	Генерирует случайные числа, имеющие заданное распределение
Ранг и персентиль	Используется для вычисления рангов и квантилей

Окончание табл. 5.1

Средство	*Описание*
Регрессия	Используется для построения линейной регрессии
Выборка	Создает случайную выборку, рассматривая входной диапазон значений как генеральную совокупность
Парный двухвыборочный t-тест для средних	Используется для проверки гипотезы о равенстве математических ожиданий для двумерной выборки данных
Двухвыборочный t-тест с одинаковыми дисперсиями	Служит для проверки гипотезы о равенстве математических ожиданий для двух выборок. Предполагается равенство дисперсий генеральных совокупностей
Двухвыборочный t-тест с разными дисперсиями	Используется для проверки гипотезы о равенстве математических ожиданий для двух выборок. Не требует предположения о равенстве дисперсий генеральных совокупностей
Двухвыборочный z-тест для средних	Используется для проверки гипотезы о различии между математическими ожиданиям двух генеральных совокупностей

Отметим, что эти средства имеют определенные ограничения и иногда удобнее воспользоваться статистическими функциями или другими средствами Excel. Преимуществом функций перед данными средствами является то, что функции автоматически пересчитываются при любых изменениях, сделанных в выборке, тогда как эти средства необходимо выполнять заново, если выборка изменилась. В "оправдание" этих средств скажем, что они сохраняют установки, сделанные пользователем при последнем применении средства, но только в течение одного сеанса работы с Excel.

Средства, которые включены в надстройку Пакет анализа, доступны через команду Сервис⇨Анализ данных. (Если команды Анализ данных нет в меню Сервис, подключите эту надстройку. Для этого выполните команду Сервис⇨Надстройки и в открывшемся диалоговом окне Надстройки в списке Доступные надстройки установите флажок напротив опции Пакет анализа.) Команда Сервис⇨Анализ данных открывает одноименное диалоговое окно, в списке Инструменты анализа которого следует выбрать необходимое средство (рис. 5.1). После выбора какого-либо средства (и последующего щелчка на кнопке ОК) открывается диалоговое окно этого средства.

В большинстве таких диалоговых окон (на рис. 5.2 для примера показано диалоговое окно средства Описательная статистика) выделены области Входные данные и Параметры вывода. В области Входные данные указывается диапазон ячеек, в котором содержатся данные (поле Входной интервал), указывается, сгруппированы ли данные, и если сгруппированы, то по столбцам или по строкам (переключатели по столбцам и по строкам). Если задается входной диапазон данных вместе с заголовками, то устанавливается флажок опции Метки в первой строке (столбце). (Если заголовки не задаются, то данным автоматически присваиваются заголовки Столбец1, Столбец2 и т.д. или Строка1, Строка2 и т.д. в зависимости от того, расположены данные в столбцах или в строках.) В некоторых диалоговых окнах в области Входные данные необходимо указать несколько входных диапазонов (например, в окне Регрессия) либо дополнительные параметры для проведения выбранной статистической процедуры, например доверительный уровень для проведения тестов.

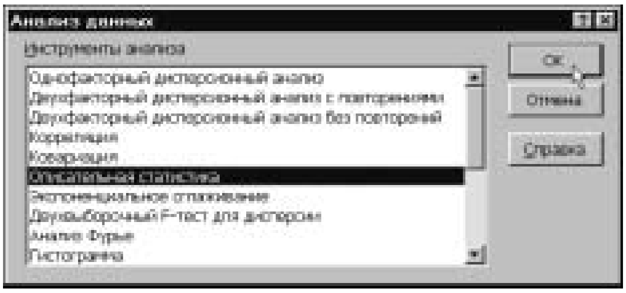


Рис. 5.1. Диалоговое окно Анализ данных *со списком инструментов статистического анализа*

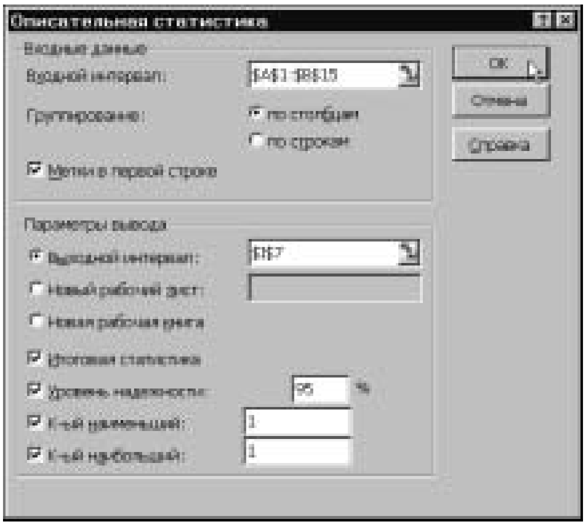


Рис. 5.2. Диалоговое окно средства Описательная статистика

В области Параметры вывода, как правило, надо указать, куда будут выводиться результаты расчетов. Предусмотрено три возможности: на текущий рабочий лист (переключатель Выходной интервал), при этом необходимо указать выходной интервал (достаточно указать адрес одной ячейки, которая определяет верхний левый угол выходного диапазона); на новый рабочий лист текущей рабочей книги начиная с ячейки A1 (переключатель Новый рабочий лист), при этом можно сразу задать имя этому листу; в новую рабочую книгу (переключатель Новая рабочая книга), в этом случае автоматически открывается новая рабочая книга. Также в этой области часто имеются опции, которые указывают, что именно необходимо вывести из возможного набора выходных результатов (например, графики либо дополнительные статистические характеристики).

В некоторых диалоговых окнах имеются другие области, в которых содержатся опции, необходимые для работы данного средства. Эти опции будут приведены при описании конкретных средств. Опции областей Входные данные и Параметры вывода будем упоминать только тогда, когда они будут отличаться от описанных выше.

Перейдем к описанию конкретных средств статистического анализа, при этом будем называть их так, как они названы в списке диалогового окна Анализ данных. Опишем их в порядке “от простого к сложному” (другими словами, в том порядке, который больше нравится автору).

5.1. Описательная статистика

Это средство (вместе со средством Гистограмма, которое будет описано в следующем разделе) является, по-видимому, наиболее часто используемым из всего пакета анализа, поскольку быстро и просто вычисляет основные статистические характеристики одномерных выборок. На рис. 5.3 показан рабочий лист, содержащий три ряда данных (три независимые выборки, имеющие разные распределения) и диалоговое окно Описательная статистика.

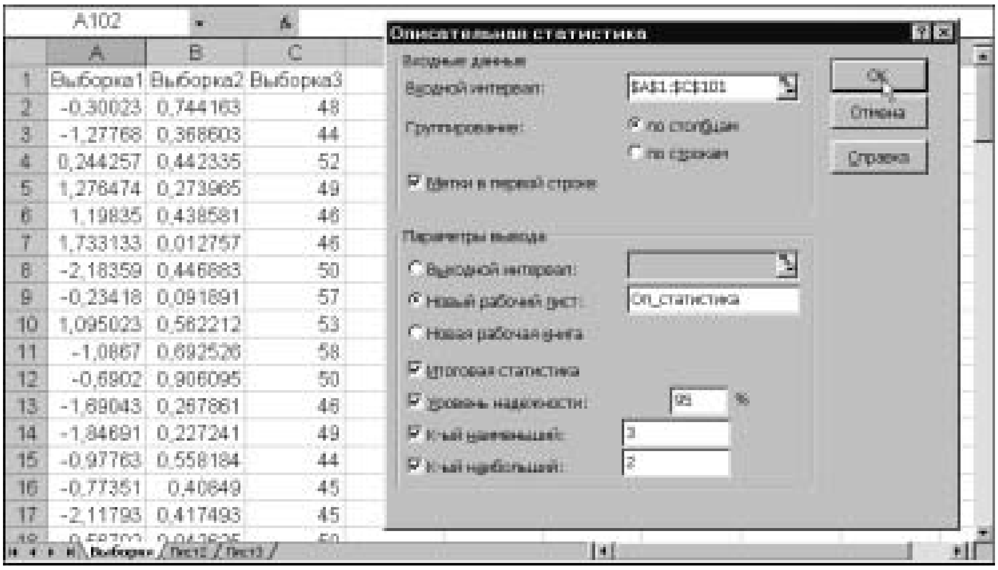
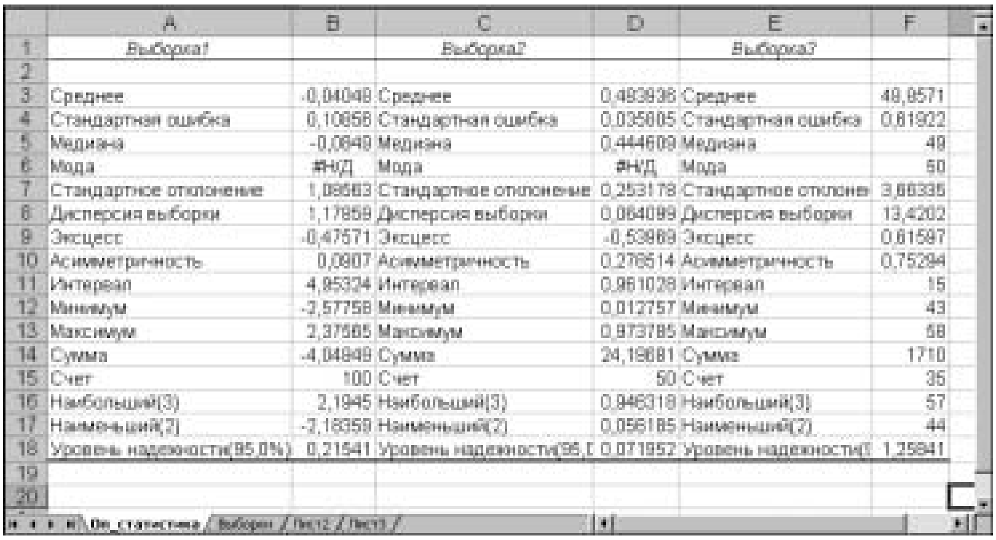


Рис. 5.3. *Три выборки и диалоговое окно* Описательная статистика

Отметим, что в данном случае имеются выборки разных размеров. Средство Описательная статистика правильно определяет размеры выборок, игнорируя пустые ячейки. На рис. 5.4 показан рабочий лист с результатами расчетов.

В табл. 5.2 перечислены вычисляемые средством Описательная статистика статистические характеристики выборок, а также функции, которые возвращают те же самые характеристики.



	A	B	C	D	E	F
1	Выборка1		Выборка2		Выборка3	
2						
3	Среднее	-0,04048	Среднее	0,493936	Среднее	48,8571
4	Стандартная ошибка	0,10856	Стандартная ошибка	0,035805	Стандартная ошибка	0,61922
5	Медиана	-0,0849	Медиана	0,444609	Медиана	49
6	Мода	#Н/Д	Мода	#Н/Д	Мода	50
7	Стандартное отклонение	1,08563	Стандартное отклонение	0,253178	Стандартное отклонен	3,66335
8	Дисперсия выборки	1,17859	Дисперсия выборки	0,064099	Дисперсия выборки	13,4202
9	Эксцесс	-0,47571	Эксцесс	-0,53969	Эксцесс	0,61597
10	Асимметричность	0,0907	Асимметричность	0,276514	Асимметричность	0,75294
11	Интервал	4,95324	Интервал	0,961028	Интервал	15
12	Минимум	-2,57758	Минимум	0,012757	Минимум	43
13	Максимум	2,37565	Максимум	0,973785	Максимум	58
14	Сумма	-4,04849	Сумма	24,19681	Сумма	1710
15	Счет	100	Счет	50	Счет	35
16	Наибольший(3)	2,1945	Наибольший(3)	0,946318	Наибольший(3)	57
17	Наименьший(2)	-2,18359	Наименьший(2)	0,056185	Наименьший(2)	44
18	Уровень надежности(95,0%)	0,21541	Уровень надежности(95,0	0,071952	Уровень надежности(	1,25841
19						
20						

Рис. 5.4. Результаты работы средства Описательная статистика

Таблица 5.2. Значения, вычисляемые средством Описательная статистика

Значение	*Описание*
Среднее	Выборочное среднее $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Функция СРЗНАЧ
Стандартная ошибка	Оценка среднеквадратического отклонения выборочного среднего; вычисляется по формуле $\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2}$
Медиана	Значение медианы, т.е. квантиля порядка 0,5. Функция МЕДИАНА
Мода	Значение моды. Вычисляется так же, как и функцией МОДА (см. раздел 4.11.3), — если нет одинаковых выборочных значений, то возвращается значение ошибки #Н/Д
Стандартное отклонение	Оценка среднеквадратического отклонения генеральной совокупности $s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$. Функция СТАНДОТКЛОН
Дисперсия выборки	Оценка дисперсии генеральной совокупности $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Функция ДИСП
Эксцесс	Выборочный коэффициент эксцесса (см. раздел 2.3.4). Функция ЭКСЦЕСС
Асимметричность	Выборочный коэффициент асимметрии (см. раздел 2.3.4). Функция СКОС

Значение	*Описание*
Интервал	Размах выборки. Вычисляется как разность между максимальным и минимальным выборочными значениями
Минимум	Минимальное выборочное значение. Функция МИН
Максимум	Максимальное выборочное значение. Функция МАКС
Сумма	Сумма выборочных значений. Функция СУММ
Счет	Объем выборки. Функция СЧЁТ
Наибольший (К)	К-е наибольшее значение. Если К = 1, то выводится максимальное выборочное значение. Функция НАИБОЛЬШИЙ
Наименьший (К)	К-е наименьшее значение. Если К = 1, то выводится минимальное выборочное значение. Функция НАИМЕНЬШИЙ
Уровень надежности (Х%)	Граница доверительного интервала для неизвестного математического ожидания с доверительным уровнем Х%; доверительный интервал строится как выборочное среднее плюс-минус данное значение. Граница вычисляется с помощью распределения Стьюдента (см. раздел 2.3.6), т.е. здесь неявно используется предположение о нормальности распределения генеральной совокупности. Поэтому к данному показателю следует относиться осторожно, особенно при малых выборках

5.1.1. Опции диалогового окна Описательная статистика

Установка флажка опции Итоговая статистика указывает, что в итоговом отчете этого средства будут вычислены все статистические характеристики выборки, за исключением границы доверительного интервала для среднего и К-х наибольших и наименьших значений, для которых имеются отдельные опции Уровень надежности, К-ый наименьший и К-ый наибольший. Если флажок опции Итоговая статистика не установлен, то выводится только то, что задается с помощью опций Уровень надежности, К-ый наименьший и К-ый наибольший.

Опция Уровень надежности указывает, надо ли вычислять границу доверительного интервала для среднего. В поле ввода рядом с этой опцией задается доверительный уровень в процентах.

В полях ввода рядом с опциями К-ый наибольший и К-ый наименьший указываются порядки выводимых наибольшего и наименьшего значений. Если эти порядки равны 1, то выводятся соответственно максимальное и минимальное выборочные значения.

5.2. Гистограмма

Это средство полезно для первичного анализа распределения выборки и построения гистограмм (столбцовых диаграмм эмпирических плотностей вероятностей). В качестве исходных данных нужно указать входной диапазон, содержащий выборочные значения, и интервал карманов. Интервал карманов определяет границы для столбцов гистограммы. Средство Гистограмма подсчитывает число выборочных значений, попавших в каждый карман (эти числа в выходных данных называются Частота), и по этим числам строит гистограмму. Далее последовательно суммируются частоты (подсчитываются так называемые накапливающие суммы), эти суммы делятся на объем выборки и умножаются на 100. Получается то, что здесь называется Интегральный процент. На самом деле, если убрать проценты (т.е. накапливающие суммы нормировать не на 100%, а на 1), это просто эмпирическая функция распределения. Средство Гистограмма предоставляет возможность вывести значения интегрального процента в виде графика. В качестве дополнительной возможности предусмотрена сортировка частот по убыванию и построение гистограммы по этим отсортированным частотам.

5.2.1. Опции диалогового окна Гистограмма

Диалоговое окно Гистограмма показано на рис. 5.5. В области Входные данные задаются адрес диапазона ячеек с выборочными значениями (поле ввода Входной интервал) и адрес диапазона, содержащего границы карманов (поле ввода Интервал карманов). Границы карманов должны быть представлены в порядке возрастания. При подсчете количества попаданий выборочных значений в карманы в число попавших в данный карман включаются значения, равные нижней границе кармана и меньшие верхней границы кармана. Если не указывать интервал границ карманов, будут автоматически созданы равновеликие интервалы, количество которых определяется по формуле Стерджесса $k = [1 + 3{,}22 \ln(n)]$ ($[x]$ — целая часть числа x). (Более подробно о построении интервалов речь идет в разделе 8.3.2.)

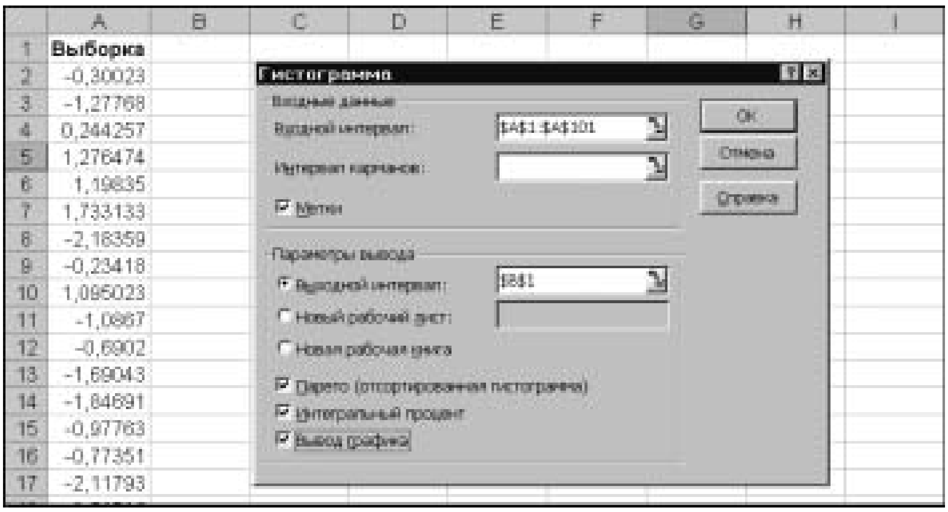
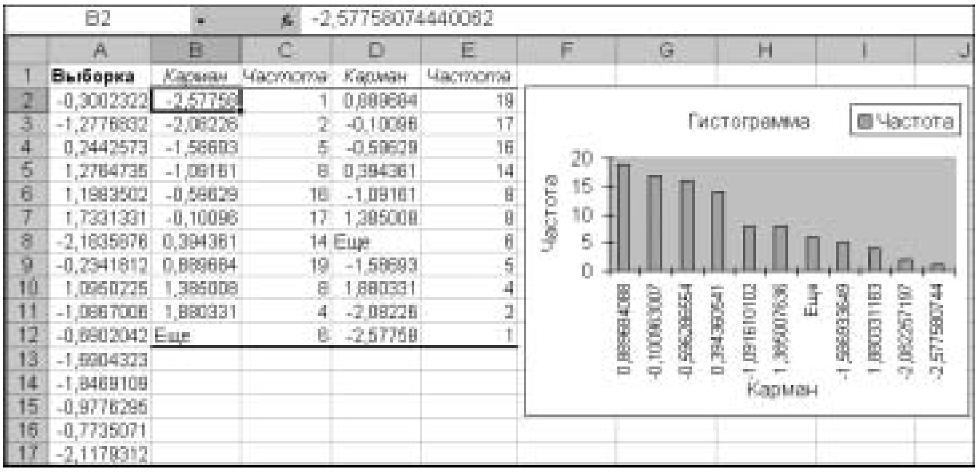


Рис. 5.5. Исходные данные и диалоговое окно Гистограмма

Рассмотрим опции Парето (отсортированная гистограмма), Интегральный процент и Вывод графика из области Параметры вывода.

Если установлен только флажок опции Парето (отсортированная гистограмма), то выводятся таблица частот и таблица отсортированных в порядке убывания частот. Если также установлен флажок опции Вывод графика, выводится гистограмма отсортированных частот, как показано на рис. 5.6.

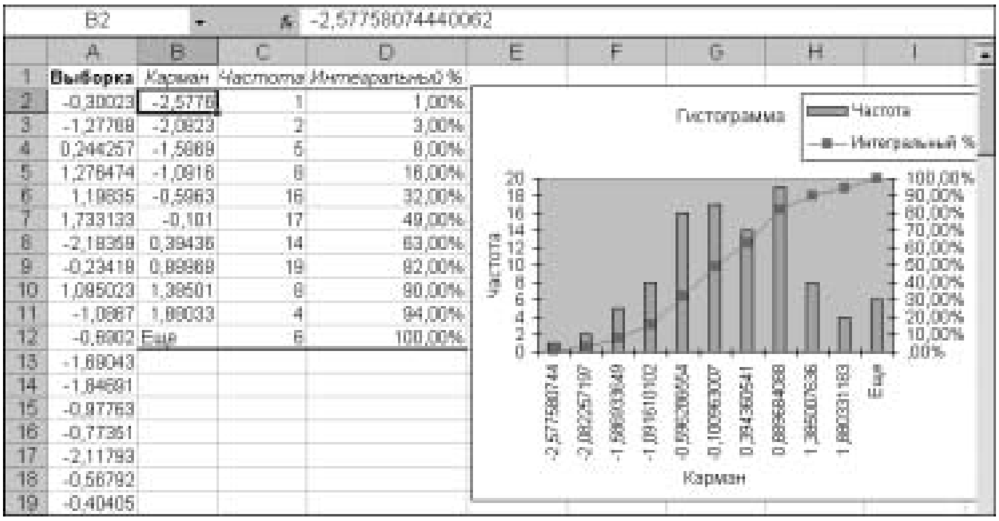


B2 = -2,57758074440082

	A	B	C	D	E
1	Выборка	Карман	Частота	Карман	Частота
2	-0,3002322	-2,57758	1	0,889684	19
3	-1,2776832	-2,08226	2	-0,10096	17
4	0,2442573	-1,58693	5	-0,59629	16
5	1,2764735	-1,09161	8	0,394361	14
6	1,1983502	-0,59629	16	-1,09161	8
7	1,7331331	-0,10096	17	1,385008	8
8	-2,1835876	0,394361	14	Еще	6
9	-0,2341812	0,889684	19	-1,58693	5
10	1,0950225	1,385008	8	1,880331	4
11	-1,0867006	1,880331	4	-2,08226	2
12	-0,6902042	Еще	6	-2,57758	1
13	-1,6904323				
14	-1,8469109				
15	-0,9776295				
16	-0,7735071				
17	-2,1179312				

Рис. 5.6*. Отсортированная гистограмма*

Если установлен только флажок опции Интегральный процент, то выводится таблица, содержащая частоты и значения интегрального процента. Если еще установлен флажок опции Вывод графика, эти данные также отображаются графически, как показано на рис. 5.7.



B2 = -2,57758074440062

	A	B	C	D
1	Выборка	Карман	Частота	Интегральный %
2	-0,30023	-2,5776	1	1,00%
3	-1,27768	-2,0823	2	3,00%
4	0,244257	-1,5869	5	8,00%
5	1,276474	-1,0916	8	16,00%
6	1,19835	-0,5963	16	32,00%
7	1,733133	-0,101	17	49,00%
8	-2,18359	0,39436	14	63,00%
9	-0,23418	0,88968	19	82,00%
10	1,095023	1,38501	8	90,00%
11	-1,0867	1,88033	4	94,00%
12	-0,6902	Еще	6	100,00%
13	-1,69043			
14	-1,84691			
15	-0,97763			
16	-0,77361			
17	-2,11793			
18	-0,56792			
19	-0,40405			

Рис. 5.7*. Гистограмма частот и график интегрального процента*

Если установлены флажки опций Парето (отсортированная гистограмма) и Интегральный процент, то выводятся две таблицы: одна содержит неотсортированные частоты и интегральные проценты, вторая — отсортированные частоты и соответствующие интегральные проценты (рис. 5.8). Если также установлен флажок опции Вывод графика, выводятся гистограмма и график интегрального процента, построенные по отсортированным частотам.

	A	B	C	D	E	F	G
1	**Выборка**	*Карман*	*Частота*	*Интегральный %*	*Карман*	*Частота*	*Интегральный %*
2	-0,30023	-2,577581	1	1,00%	0,88968	19	19,00%
3	-1,27768	-2,082257	2	3,00%	-0,101	17	36,00%
4	0,244257	-1,586934	5	8,00%	-0,5963	16	52,00%
5	1,276474	-1,09161	8	16,00%	0,39436	14	66,00%
6	1,19835	-0,596287	16	32,00%	-1,0916	8	74,00%
7	1,733133	-0,100963	17	49,00%	1,38501	8	82,00%
8	-2,18359	0,3943605	14	63,00%	Еще	6	88,00%
9	-0,23418	0,8896841	19	82,00%	-1,5869	5	93,00%
10	1,095023	1,3850076	8	90,00%	1,88033	4	97,00%
11	-1,0867	1,8803312	4	94,00%	-2,0823	2	99,00%
12	-0,6902	Еще	6	100,00%	-2,5776	1	100,00%
13	-1,69043						
14	-1,84691						

Рис. 5.8. *Выходные данные (две таблицы)*

Наконец, если установлен флажок только опции Вывод графика, выводятся таблица частот (не отсортированная) и гистограмма.

5.3. Генерация случайных чисел

Это средство предназначено для генерирования значений случайных чисел, имеющих заданное распределение, т.е. для получения случайных выборок. Средство имеет возможность генерировать случайные числа, имеющие следующие распределения.

- **Равномерное.** Генерируется последовательность равномерно распределенных случайных чисел в заданном интервале, для чего необходимо указать верхнюю и нижнюю границы интервала.
- **Нормальное.** Генерируется последовательность случайных чисел, подчиняющихся нормальному распределению. Задается математическое ожидание и среднеквадратическое отклонение.
- **Бернулли.** Генерируется последовательность случайных чисел, принимающих только значение 0 или 1, в зависимости от заданной вероятности успеха (исхода “1”). (О распределении Бернулли речь идет в разделе 1.4.2.)
- **Биномиальное.** Генерируется последовательность случайных чисел, равная количеству исходов “1” в n независимых испытаниях. В результате каждого из них с вероятностью p может произойти исход “1” и с вероятностью $(1 - p)$ — исход “0” (см. раздел 1.4.3). Здесь необходимо задать число испытаний n и вероятность p.
- **Пуассона.** Генерируется последовательность случайных чисел, подчиняющихся распределению Пуассона с заданным параметром λ. (О распределении Пуассона речь идет в разделе 1.4.4.)
- **Модельное.** При выборе этого распределения на самом деле генерируются не случайные числа, а повторяющаяся последовательность членов арифметической прогрессии, причем члены прогрессии также могут повторяться заданное число раз. Для этого распределения задаются интервал изменения

членов арифметической прогрессии, шаг прогрессии, число повторений членов прогрессии и число повторений этой последовательности чисел.

- **Дискретное.** Генерируется последовательность случайных чисел, подчиняющихся заданному дискретному распределению. Для задания этого распределения необходимо указать диапазон ячеек, состоящий из двух столбцов: в первом столбце содержатся значения, а во втором — вероятности каждого значения. Сумма вероятностей во втором столбце должна быть равна 1.

5.3.1. Опции диалогового окна Генерация случайных чисел

Диалоговое окно Генерация случайных чисел при задании различных распределений имеет ряд одинаковых элементов, но наличие некоторых других опций зависит от выбранного типа распределения. Выбор распределения осуществляется в раскрывающемся списке Распределение.

Рассмотрим сначала общие элементы всех диалоговых окон Генерация случайных чисел.

В поле ввода Число переменных указывается количество генерируемых выборок. Каждая выборка располагается в отдельном столбце. Максимальное количество выборок — 256 (по количеству столбцов в рабочем листе Excel). Если это число не введено, то будет сгенерирована одна случайная выборка, или, если в поле Выходной интервал указан диапазон ячеек, в котором будут располагаться сгенерированные значения, будут заполнены все столбцы этого диапазона.

В поле ввода Число случайных чисел задается количество выборочных значений (т.е. объем генерируемых выборок), одно и то же для всех выборок. Если это число не введено, то будет сгенерировано одно значение, или, если в поле Выходной интервал указан диапазон ячеек, в котором будут располагаться сгенерированные значения, будут заполнены все строки этого диапазона.

В большинстве диалоговых окон Генерация случайных чисел (кроме окон для модельного и дискретного распределений) имеется поле ввода Случайное рассеивание. Число, введенное в это поле, задает начальное значение, которое будет использовано в алгоритме генерации случайных чисел. Обычно это поле оставляют пустым. Однако, чтобы генерировать одинаковые последовательности случайных чисел, необходимо ввести число из диапазона от 1 до 32 767 (допускаются только целые числа). Тогда в будущем можно получить тот же набор выборочных значений, если в это поле снова ввести то же самое начальное значение.

Все диалоговые окна Генерация случайных чисел имеют область Параметры; опции этой области зависят от типа выбранного распределения. Назначение большинства этих опций очевидно, но некоторые требуют пояснений.

Равномерное распределение. Диалоговое окно Генерация случайных чисел для этого распределения показано на рис. 5.9.

Здесь в области Параметры надо задать только верхнюю и нижнюю границы, в пределах которых сосредоточено распределение.

Нормальное распределение. Диалоговое окно Генерация случайных чисел для этого распределения показано на рис. 5.10.

В области Параметры задаются значения среднего (математического ожидания) и стандартное (среднеквадратическое отклонение). Для стандартного нормального распределения среднее равно 0, а стандартное отклонение — 1.

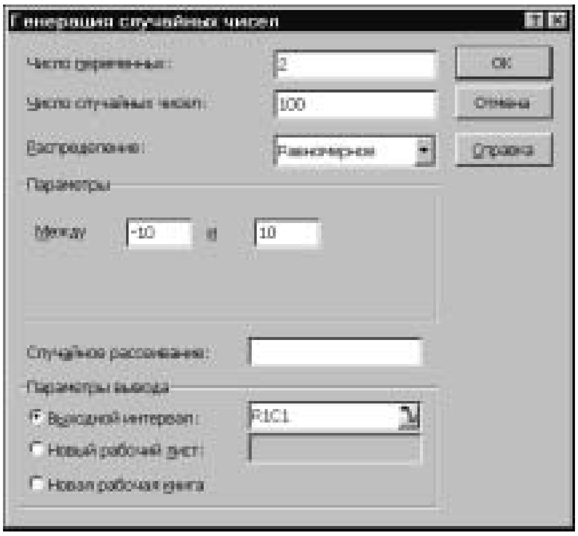


Рис. 5.9. Диалоговое окно для генерирования равномерно распределенных случайных чисел

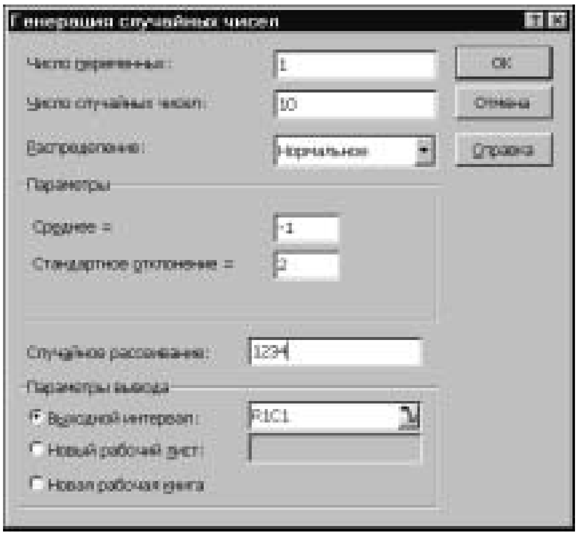


Рис. 5.10. Диалоговое окно для генерирования нормально распределенных случайных чисел

Распределение Бернулли. Диалоговое окно Генерация случайных чисел для данного случая показано на рис. 5.11.

Здесь в области Параметры задается только один параметр — вероятность p.

Биномиальное распределение. Диалоговое окно Генерация случайных чисел для этого распределения показано на рис. 5.12.

Для этого распределения задаются значения вероятности p и количество испытаний n.

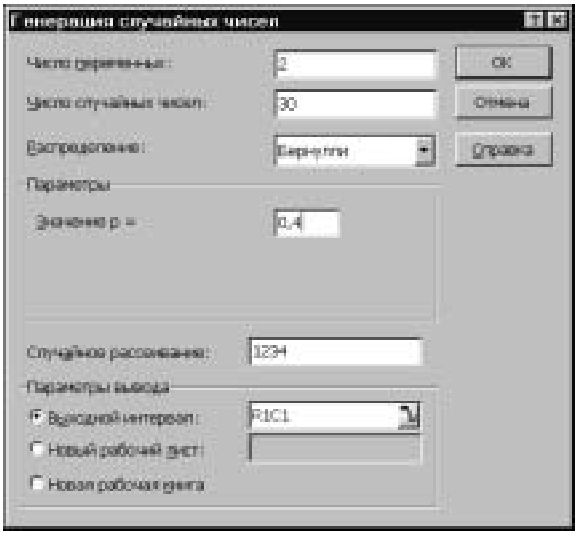


Рис. 5.11. Диалоговое окно для генерирования случайных чисел, имеющих распределение Бернулли

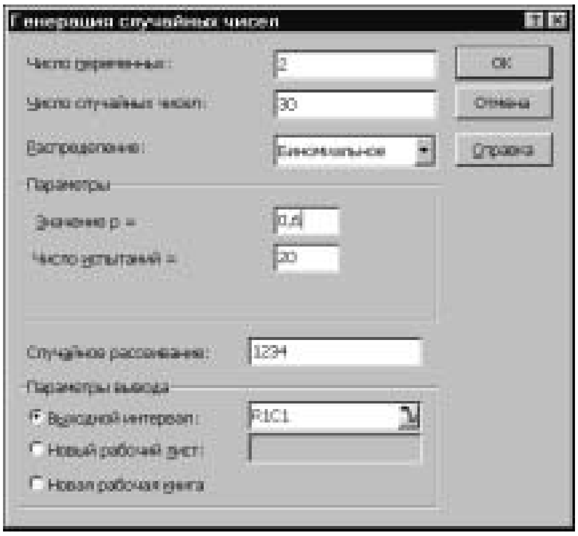


Рис. 5.12. Диалоговое окно для генерирования случайных чисел, имеющих биномиальное распределение

Распределение Пуассона. Диалоговое окно Генерация случайных чисел для данного случая показано на рис. 5.13.

Здесь в области Параметры задается только один параметр Лямбда.

Модельное распределение. Диалоговое окно Генерация случайных чисел для этого случая показано на рис. 5.14.

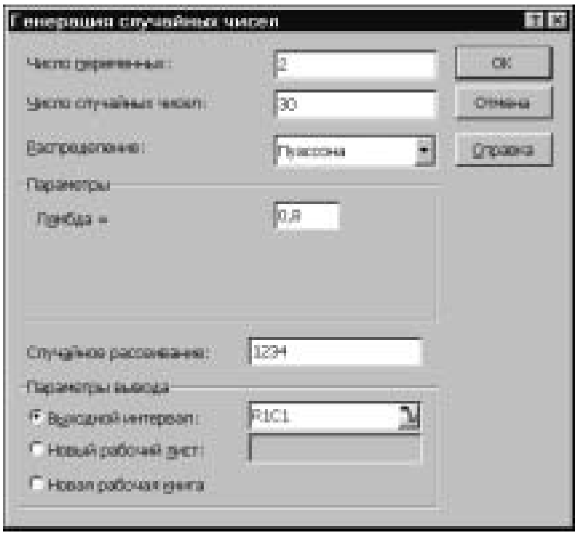


Рис. 5.13. *Диалоговое окно для генерирования случайных чисел, имеющих распределение Пуассона*

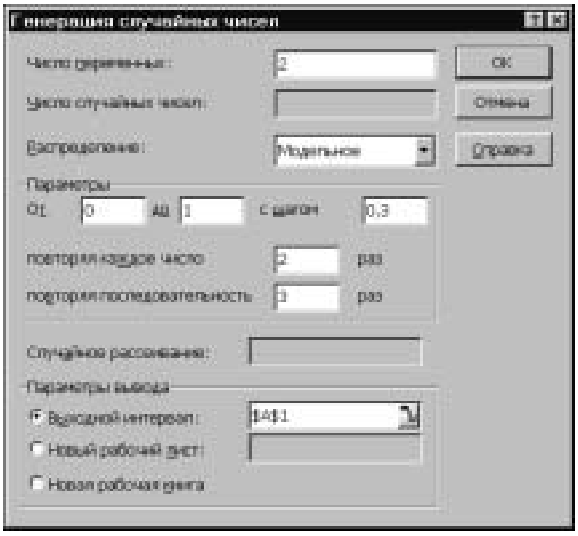
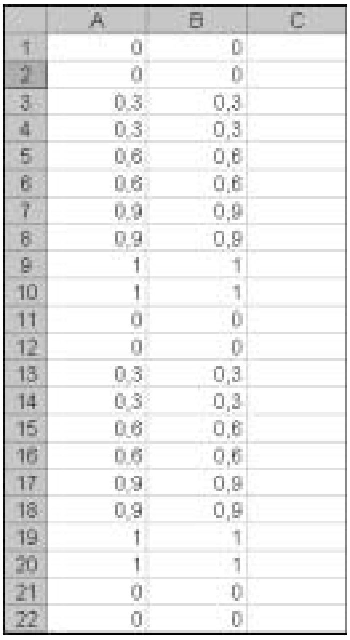


Рис. 5.14. *Диалоговое окно для генерирования заданных чисел (модельное распределение)*

Здесь задаются нижняя и верхняя границы чисел, шаг прогрессии, число повторений значений в последовательности и число повторений последовательности. На рис. 5.15 показаны сгенерированные числа с модельным распределением, параметры которого заданы на рис. 5.14.



	A	B	C
1	0	0	
2	0	0	
3	0,3	0,3	
4	0,3	0,3	
5	0,6	0,6	
6	0,6	0,6	
7	0,9	0,9	
8	0,9	0,9	
9	1	1	
10	1	1	
11	0	0	
12	0	0	
13	0,3	0,3	
14	0,3	0,3	
15	0,6	0,6	
16	0,6	0,6	
17	0,9	0,9	
18	0,9	0,9	
19	1	1	
20	1	1	
21	0	0	
22	0	0	

Рис. 5.15. *Сгенерированные числа*

Дискретное распределение. Диалоговое окно Генерация случайных чисел для этого типа распределения вместе с необходимыми входными данными показано на рис. 5.16.

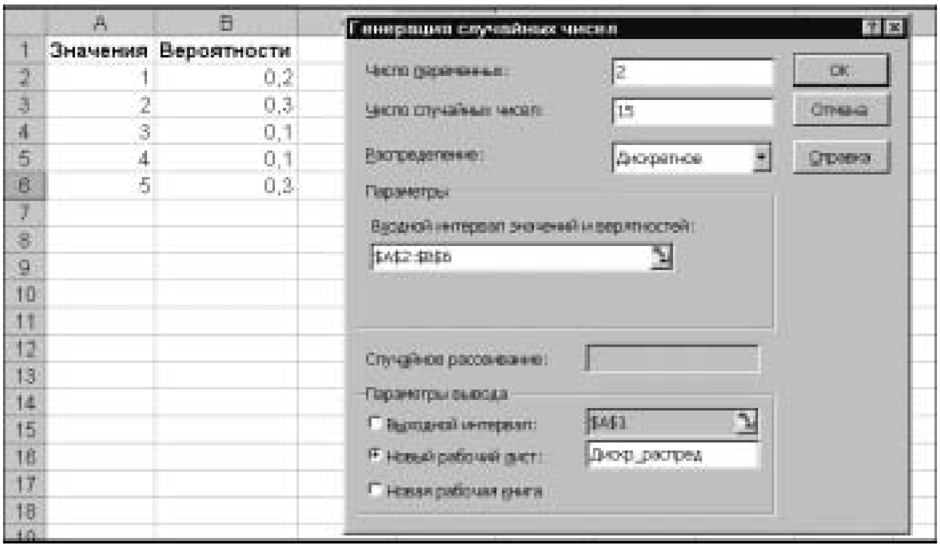
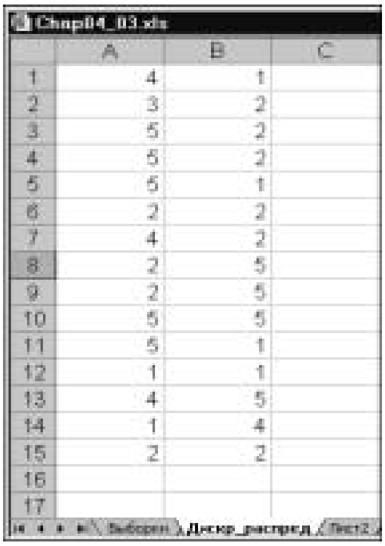


Рис. 5.16. *Диалоговое окно для генерирования случайных чисел, имеющих заданное дискретное распределение*



Chap04_03.xls

	A	B	C
1	4	1	
2	3	2	
3	5	2	
4	5	2	
5	5	1	
6	2	2	
7	4	2	
8	2	5	
9	2	5	
10	5	5	
11	5	1	
12	1	1	
13	4	5	
14	1	4	
15	2	2	
16			
17			

Выборки / Дискр_распред / Лист2

Рис. 5.17. *Сгенерированные числа*

Для задания дискретного распределения в поле Входной интервал значений и вероятностей необходимо указать адрес диапазона ячеек, содержащий значения случайной величины и соответствующие им вероятности. Диапазон должен состоять из двух столбцов: левого, содержащего значения, и правого, содержащего вероятности, как показано на рис. 5.16. Сумма вероятностей должна быть равна 1. На рис. 5.17 представлены сгенерированные числа с распределением, параметры которого заданы на рис. 5.16.

В заключение отметим, что в Excel имеются и другие средства генерирования случайных выборок, например функции СЛЧИС и СЛУЧМЕЖДУ (см. раздел 4.13). Подробно задача генерирования значений случайных величин рассмотрена в главе 7.

5.4. Выборка

Это средство из исходного числового множества выбирает указанное количество чисел, причем либо случайным образом, либо с заданным периодом (например, каждое второе или каждое десятое число). Такую операцию выбора числовых значений из заданного множества можно трактовать как создание выборки заданного объема, если исходное множество рассматривать как генеральную совокупность. Подобная операция часто составляет один из этапов предварительной обработки данных. Например, если исходная выборка слишком велика для обработки или построения диаграмм либо если исходные данные содержат периодическую составляющую, то можно создать выборку, содержащую значения только из отдельных частей периода.

5.4.1. Опции диалогового окна Выборка

Диалоговое окно Выборка показано на рис. 5.18. Адрес диапазона ячеек, содержащий исходный набор числовых значений, задается в поле Входной интервал. Если этот диапазон состоит из нескольких столбцов, то значения сначала будут извлекаться из первого столбца, затем из второго столбца и т.д. Средство Выборка откажется работать (выведет соответствующее окно предупреждения), если среди исходных данных имеются нечисловые значения.

В области Метод выборки необходимо указать, каким способом будут выбираться значения из исходного множества. Если установлен переключатель Периодический, то из исходного множества будет выбрано каждое n-е значение; число n задается в поле ввода Период. Количество выбранных значений будет равно количеству значений в исходном диапазоне, деленному на значение в поле Период. Если установлен переключатель Случайный, значения из исходного множества выбираются случайным образом; количество выбираемых значений задается в поле Число выборок.

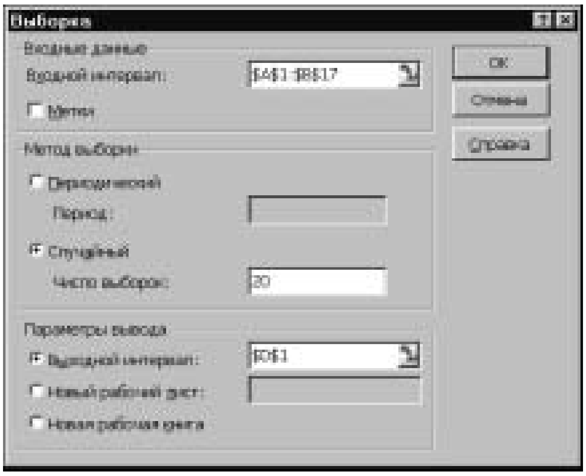


Рис. 5.18. *Диалоговое окно* Выборка

5.5. Ранг и персентиль

Это средство позволяет создать таблицу, содержащую порядковый и процентный ранги для каждого значения в заданном наборе данных, при этом значения упорядочиваются в *порядке убывания*. На рис. 5.19 показаны диалоговое окно Ранг и персентиль и исходные данные, на рис. 5.20 — результат применения этого средства. Итоговая таблица содержит порядковый номер выборочного значения, столбец выборочных значений, отсортированных в порядке убывания, столбец рангов и столбец процентных рангов этих значений, причем наибольшему значению присваивается ранг 1 и процентный ранг 100%, а наименьшему — наибольший ранг и процентный ранг, равный 0%.

Если имеется группа совпадающих значений, то им присваиваются одинаковые ранги, равные рангу первого числа из группы совпадающих значений. Значению, следующему за этой группой, присваивается ранг, больший ранга совпадающих значений на число этих одинаковых значений. Процентный ранг T_i для выборочного значения x_i рассчитывается по формуле $T_i = \frac{n - R_i}{n - 1} \cdot 100\%$, где R_i — ранг значения x_i, рассчитанный при условии упорядочивания данных по убыванию, n — объем выборки.

5.6. Двухвыборочный z-тест для средних

Это средство применяется для проверки гипотезы о равенстве (неравенстве) математических ожиданий двух независимых генеральных совокупностей, имеющих нормальное распределение, при известных дисперсиях этих распределений (см. раздел 2.4.2). Пусть имеются две независимые выборки $x_1, x_2, \ldots, x_n$ и $y_1, y_2, \ldots, y_m$ объемом соответственно n и m, извлеченные из совокупностей, имеющих нормальные распределения с известными дисперсиями σ_1^2 и σ_2^2 и неизвестными математическими ожиданиями соответственно μ_1 и μ_2. Проверяется нулевая

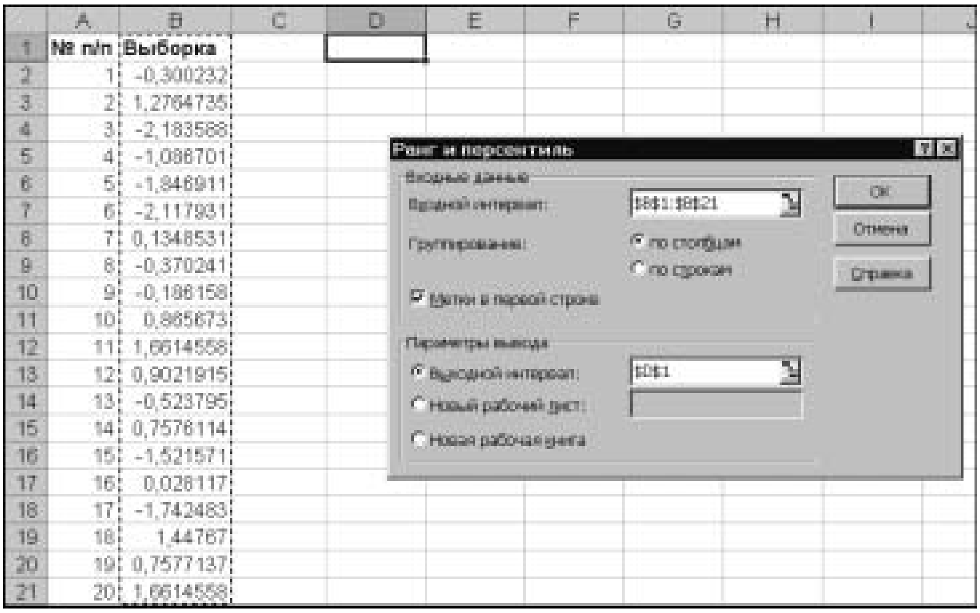
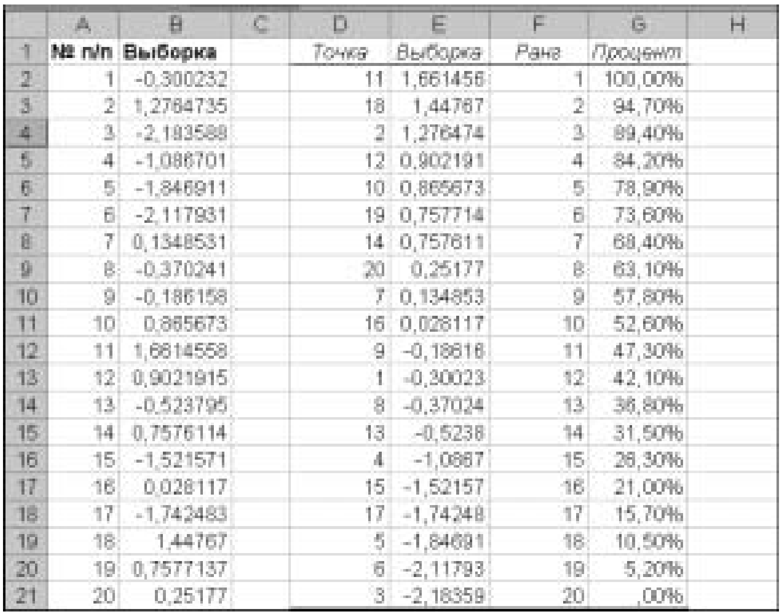


Рис. 5.19. Исходные данные и диалоговое окно Ранг и персентиль



	A	B	C	D	E	F	G	H
1	№ п/п	Выборка		Точка	Выборка	Ранг	Процент	
2	1	-0,300232		11	1,661456	1	100,00%	
3	2	1,2764735		18	1,44767	2	94,70%	
4	3	-2,183588		2	1,276474	3	89,40%	
5	4	-1,086701		12	0,902191	4	84,20%	
6	5	-1,846911		10	0,865673	5	78,90%	
7	6	-2,117931		19	0,757714	6	73,60%	
8	7	0,1348531		14	0,757611	7	68,40%	
9	8	-0,370241		20	0,25177	8	63,10%	
10	9	-0,186158		7	0,134853	9	57,80%	
11	10	0,865673		16	0,028117	10	52,60%	
12	11	1,6614558		9	-0,18616	11	47,30%	
13	12	0,9021915		1	-0,30023	12	42,10%	
14	13	-0,523795		8	-0,37024	13	36,80%	
15	14	0,7576114		13	-0,5238	14	31,50%	
16	15	-1,521571		4	-1,0867	15	26,30%	
17	16	0,028117		15	-1,52157	16	21,00%	
18	17	-1,742483		17	-1,74248	17	15,70%	
19	18	1,44767		5	-1,84691	18	10,50%	
20	19	0,7577137		6	-2,11793	19	5,20%	
21	20	0,25177		3	-2,18359	20	,00%	

Рис. 5.20. Результат вычислений

гипотеза H_0: $\mu_1 - \mu_2 = \delta$ (δ задано). Z-тест позволяет проверить гипотезу H_0 против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$. Критериальная статистика вычисляется по формуле

$$z = \frac{\overline{x} - \overline{y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_1^2}{m}}},$$

где $\overline{x}$ и $\overline{y}$ — выборочные средние соответственно первой и второй выборок.

Для выборок из нормально распределенных генеральных совокупностей критериальная статистика z имеет стандартное нормальное распределение. Поэтому при заданном уровне значимости α критическая область строится на основе стандартного нормального распределения — вычисляется квантиль t порядка $1 - \alpha$ для проверки гипотезы о равенстве либо квантиль t порядка $1 - \alpha/2$ для проверки гипотез неравенства. Нулевая гипотеза о равенстве принимается, если $|z| < t$ (в противном случае отвергается); гипотеза H_0 при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $z < t$; и при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $-t < z$.

Рассмотрим пример. Имеется две выборки[1] объемом соответственно 50 и 20 значений, показанные на рис. 5.21. Обе имеют нормальное распределение, первая — стандартное (т.е. $\mu_1 = 0$ и $\sigma_1^2 = 1$), а для второй — $\mu_2 = 1$ и $\sigma_2^2 = 2$. Проверим с помощью средства Двухвыборочный z-тест для средних нулевую гипотезу, что $\mu_2 - \mu_1 = 1{,}5$ для разных случаев конкурирующих гипотез. Заполненное диалоговое окно для этого примера также показано на рис. 5.21.

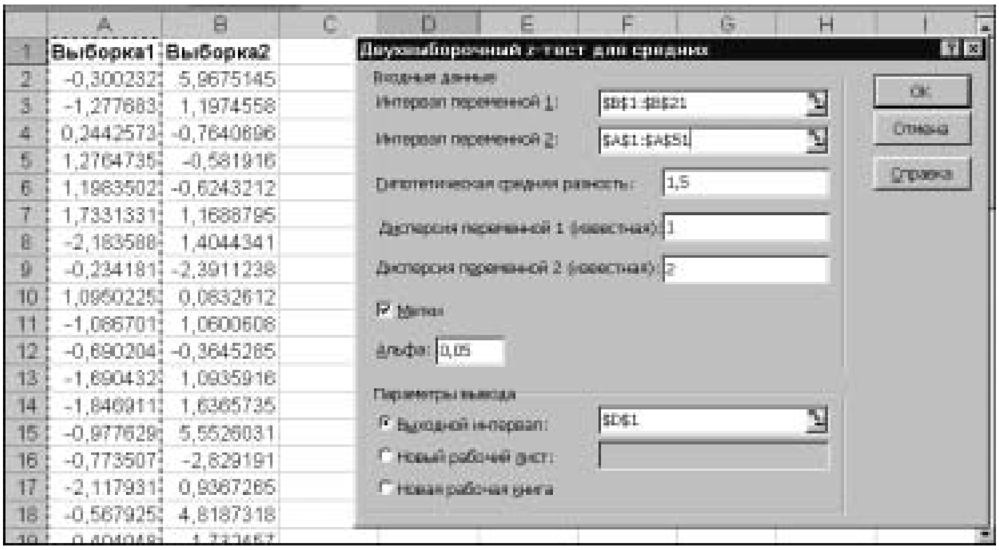


Рис. 5.21. *Исходные данные и диалоговое окно* Двухвыборочный z-тест для средних

Отметим, что средство требует, чтобы δ, значение которого задается в поле Гипотетическая средняя разность, было неотрицательно. Поэтому первым (в поле ввода Интервал переменной 1) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле Интервал переменной 2

---

[1] *Выборки получены с помощью средства* Генерация случайных чисел.

указывается адрес второй выборки. В полях ввода Дисперсия переменной 1 и Дисперсия переменной 2 вводятся значения дисперсий соответственно первой и второй выборок. В поле Альфа вводится значение уровня значимости α. Результат вычислений средства Двухвыборочный z-тест для средних показан на рис. 5.22.

	A	B	C	D	E	F
1	Выборка1	Выборка2		Двухвыборочный z-тест для средних		
2	-0,300232	5,9675145				
3	-1,277683	1,1974558			Выборка2	Выборка1
4	0,2442573	-0,7640696		Среднее	0,992495873	-0,10700326
5	1,2764735	-0,581916		Известная дисперсия	2	1
6	1,1963502	-0,6243212		Наблюдения	20	50
7	1,7331331	1,1688795		Гипотетическая разность средних	1,5	
8	-2,163588	1,4044341		z	-1,15614643	
9	-0,234181	-2,3911238		P(Z<=z) одностороннее	0,123810689	
10	1,0950225	0,0632612		z критическое одностороннее	1,644853476	
11	-1,086701	1,0600608		P(Z<=z) двухстороннее	0,247621378	
12	-0,690204	-0,3645285		z критическое двухстороннее	1,959962787	
13	-1,690432	1,0935916				
14	-1,846911	1,6365735				
15	-0,977629	5,5526031				
16	-0,773507	-2,829191				
17	-2,117931	0,9367265				
18	-0,567925	4,8187318				

Рис. 5.22. *Результат вычислений*

В итоговой таблице приводятся следующие данные.

- Среднее — выборочные средние выборок.
- Известная дисперсия — дисперсии выборок, которые указаны в диалоговом окне.
- Наблюдения — объемы выборок.
- Гипотетическая разность средних — значение δ, которое задано в диалоговом окне.
- z — значение критериальной статистики.
- P(Z<=z) одностороннее — вероятность $P(X \le z)$, где X — случайная величина, распределенная по стандартному нормальному закону, z — подсчитанное значение критериальной статистики.
- z критическое одностороннее — значение квантиля порядка $1 - \alpha/2$.
- P(Z<=z) двухстороннее — вероятность $P(|X| \le |z|)$, где X — случайная величина, распределенная по стандартному нормальному закону, z — подсчитанное значение критериальной статистики.
- z критическое двухстороннее — значение квантиля порядка $1 - \alpha$.

Как видно из результатов расчета, в данном примере нет оснований отвергать нулевую гипотезу при любых конкурирующих гипотезах.

Статистическая функция ZTECT (см. раздел 4.8.1) вычисляет вероятность P(Z<=z) двухстороннее.

5.7. Двухвыборочный t-тест с одинаковыми дисперсиями

Это средство реализует критерий проверки гипотезы о равенстве (неравенстве) математических ожиданий распределений двух независимых генеральных совокупностей, имеющих нормальные распределения с неизвестными дисперсиями в предположении, что дисперсии равны. Этот критерий, называемый *t*-тестом или тестом Стьюдента, подробно описан в разделе 2.4.2.

Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_0: $\mu_1 - \mu_2 = \delta$ (δ задано) против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$ (μ_1 и μ_2 — неизвестные математические ожидания выборок). Исходные данные и заполненное диалоговое окно Двухвыборочный t-тест с одинаковыми дисперсиями показаны на рис. 5.23. Выборки извлечены из нормально распределенных генеральных совокупностей с одной и той же дисперсией, равной 1, и математическими ожиданиями 0 и 1 соответственно[2]. Проверим гипотезу, что $\mu_2 - \mu_1 = 2$ (на самом деле $\mu_2 - \mu_1 = 1$).

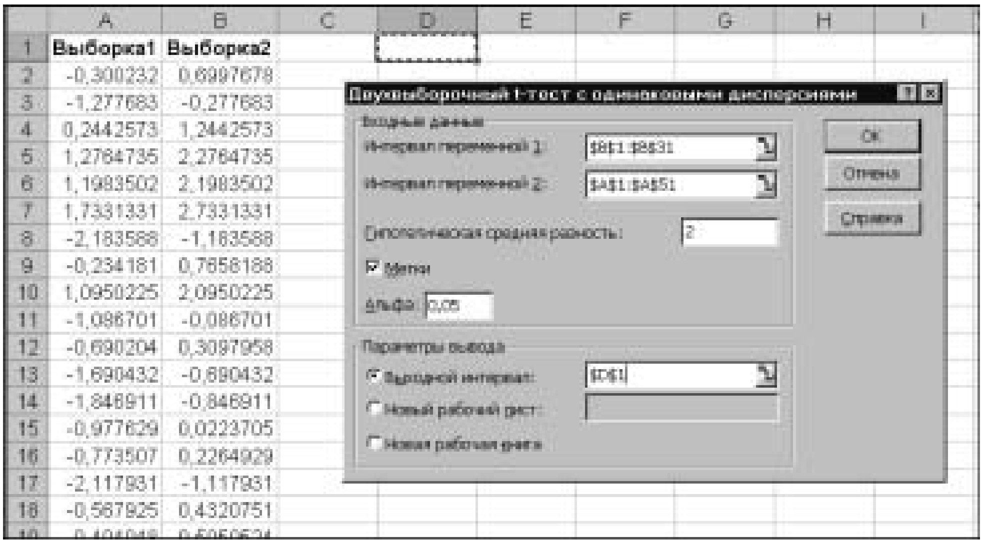


Рис. 5.23. Исходные данные и диалоговое окно Двухвыборочный t-тест с одинаковыми дисперсиями

Отметим, что средство требует, чтобы δ, значение которого задается в поле Гипотетическая средняя разность, было неотрицательно. Поэтому первым (в поле ввода Интервал переменной 1) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле Интервал переменной 2 указывается адрес второй выборки. (Диапазоны должны состоять из одного столбца или одной строки.) В поле Альфа вводится значение уровня значимости α. Результат вычислений средства Двухвыборочный t-тест с одинаковыми дисперсиями показан на рис. 5.24.

[2] *Выборки получены с помощью средства* Генерация случайных чисел.

	A	B	C	D	E	F
1	**Выборка1**	**Выборка2**		Двухвыборочный t-тест с одинаковыми дисперсиями		
2	-0,300232	0,6997678				
3	-1,277683	-0,277683			*Выборка2*	*Выборка1*
4	0,2442573	1,2442573		Среднее	0,8528339	-0,107003
5	1,2764735	2,2764735		Дисперсия	1,4100187	1,3535148
6	1,1983502	2,1983502		Наблюдения	30	50
7	1,7331331	2,7331331		Объединенная дисперсия	1,3745227	
8	-2,183588	-1,183588		Гипотетическая разность средних	2	
9	-0,234181	0,7658188		df	78	
10	1,0950225	2,0950225		t-статистика	-3,841723	
11	-1,086701	-0,086701		P(T<=t) одностороннее	0,0001237	
12	-0,690204	0,3097958		t критическое одностороннее	1,6646254	
13	-1,690432	-0,690432		P(T<=t) двухстороннее	0,0002474	
14	-1,846911	-0,846911		t критическое двухстороннее	1,9908475	
15	-0,977629	0,0223705				
16	-0,773507	0,2264829				
17	-2,117931	-1,117931				
18	-0,567925	0,4320751				

Рис. 5.24. *Результат вычислений*

В итоговой таблице приводятся следующие данные.

- Среднее — выборочные средние для каждой выборки.
- Дисперсия — несмещенные выборочные оценки дисперсий выборок.
- Наблюдения — объемы выборок.
- Гипотетическая разность средних — значение δ, которое задано в диалоговом окне.
- Объединенная дисперсия — "средняя" оценка дисперсии; рассчитывается по формуле $s^2 = \dfrac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$, где n и m — объемы выборок, s_i^2 — оценки дисперсий (их значения приводятся в строке Дисперсия).
- df — число степеней свободы; вычисляется как $n + m - 2$.
- t-статистика — значение критериальной статистики; вычисляется по формуле $t = \dfrac{\sqrt{n+m-2}(\overline{x} - \overline{y} - \delta)}{\sqrt{\dfrac{n+m}{nm}}\sqrt{(n-1)s_1^2 + (m-1)s_2^2}}$, имеет распределение Стьюдента с df степенями свободы.
- P(T<=t) одностороннее — вероятность $P(X \le t)$, где X — случайная величина, имеющая распределение Стьюдента с df степенями свободы, t — подсчитанное значение критериальной статистики.
- t критическое одностороннее — значение квантиля $t_{кр2}$ порядка $1 - \alpha$ распределения Стьюдента с df степенями свободы.

- P(T<=t) двухстороннее — вероятность $P(|X| \leq |t|)$, где X — случайная величина, имеющая распределение Стьюдента с df степенями свободы, t — подсчитанное значение критериальной статистики.
- t критическое двухстороннее — значение квантиля $t_{кр1}$ порядка $1 - \alpha/2$ распределения Стьюдента с df степенями свободы.

Нулевая гипотеза H_0: $\mu_1 - \mu_2 = \delta$ принимается, если $|t| < t_{кр1}$ (в противном случае отвергается); гипотеза H_0 при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $t < t_{кр2}$; при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $t_{кр2} < t$.

Как видно из результатов расчета, в данном примере нулевую гипотезу следует отвергнуть при любых конкурирующих гипотезах.

Статистическая функция ТТЕСТ при значении аргумента Тип = 2 (см. раздел 3.8.2) вычисляет вероятности P(T<=t) двухстороннее и P(T<=t) одностороннее.

5.8. Двухвыборочный t-тест с различными дисперсиями

Это средство реализует критерий проверки гипотезы о равенстве (неравенстве) математических ожиданий распределений двух независимых генеральных совокупностей, имеющих нормальные распределения с неизвестными и различными дисперсиями. Этот критерий также называется t-тестом или тестом Стьюдента для неравных дисперсий, либо критерием Фишера–Беренса и подробно описывается в разделе 2.4.2.

Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_0: $\mu_1 - \mu_2 = \delta$ (δ задано) против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$ (μ_1 и μ_2 — неизвестные математические ожидания выборок). Повторим тест на примере данных из предыдущего раздела, т.е. выборки извлечены из нормально распределенных генеральных совокупностей с одной и той же дисперсией, равной 1, и математическими ожиданиями соответственно 0 и 1. Проверим гипотезу, что $\mu_2 - \mu_1 = 2$ (на самом деле $\mu_2 - \mu_1 = 1$). Исходные данные и заполненное диалоговое окно Двухвыборочный t-тест с различными дисперсиями показаны на рис. 5.25.

Отметим, что средство требует, чтобы δ, значение которого задается в поле Гипотетическая средняя разность, было неотрицательно. Поэтому первым (в поле ввода Интервал переменной 1) задается адрес диапазона ячеек, содержащий выборку с бо́льшим математическим ожиданием, а затем в поле Интервал переменной 2 указывается адрес второй выборки. (Диапазоны должны состоять из одного столбца или одной строки.) В поле Альфа вводится значение уровня значимости α. Результат вычислений средства Двухвыборочный t-тест с различными дисперсиями показан на рис. 5.26.

В итоговой таблице приводятся следующие данные.

- Среднее — выборочные средние для каждой выборки.
- Дисперсия — несмещенные выборочные оценки дисперсий выборок.

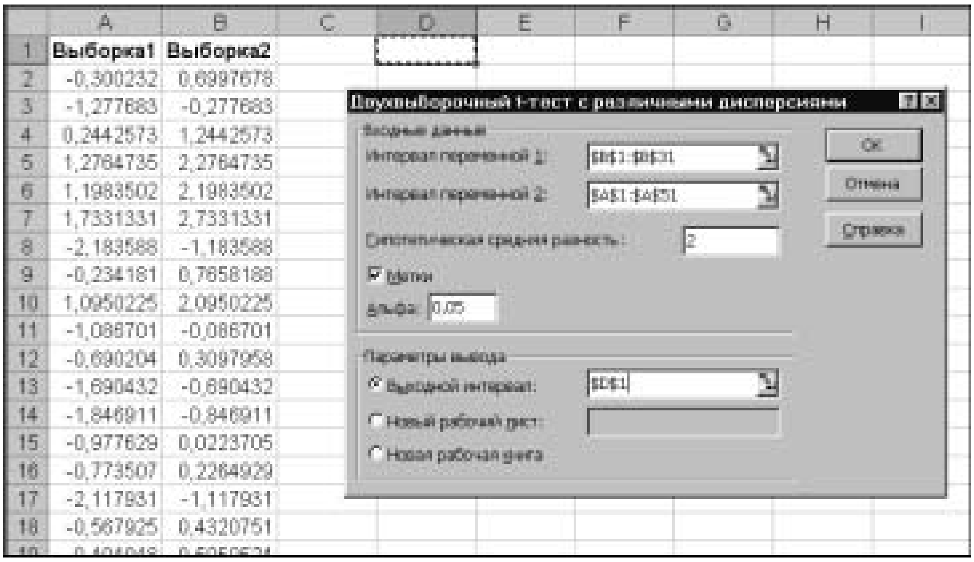


Рис. 5.25. *Исходные данные и диалоговое окно* ***Двухвыборочный t-тест с различными дисперсиями***

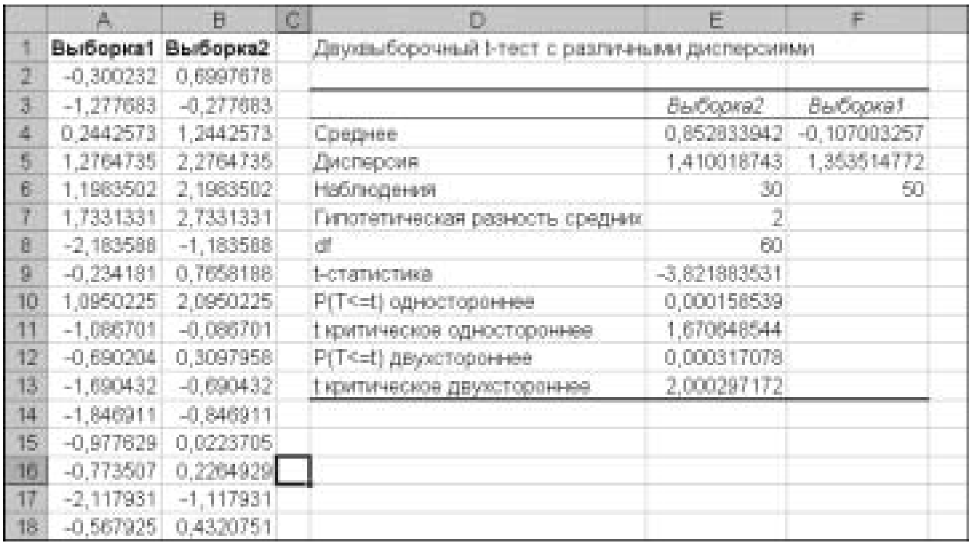


Рис. 5.26. *Результат вычислений*

- Наблюдения — объемы выборок.
- Гипотетическая разность средних — значение δ, которое задано в диалоговом окне.

- df — число степеней свободы; вычисляется по формуле $\dfrac{\left(\dfrac{s_1^2}{n}+\dfrac{s_2^2}{m}\right)^2}{\dfrac{(s_1^2/n)^2}{n-1}+\dfrac{(s_2^2/m)^2}{m-1}}$, где

  s_1^2 и s_2^2 — несмещенные оценки дисперсий (их значения приводятся в строке Дисперсия), n и m — объемы соответственно первой и второй выборок.
- t-статистика — значение критериальной статистики; вычисляется по формуле $t=\dfrac{\overline{x}-\overline{y}-\delta}{\sqrt{\dfrac{s_1^2}{n}+\dfrac{s_2^2}{m}}}$, имеет распределение, близкое к распределению Стьюдента с df степенями свободы.
- P(T<=t) одностороннее — вероятность $P(X \le t)$, где X — случайная величина, имеющая распределение Стьюдента с df степенями свободы, t — подсчитанное значение критериальной статистики.
- t критическое одностороннее — значение квантиля $t_{кр2}$ порядка $1-\alpha$ распределения Стьюдента с df степенями свободы.
- P(T<=t) двухстороннее — вероятность $P(|X| \le |t|)$, где X — случайная величина, имеющая распределение Стьюдента с df степенями свободы, t — подсчитанное значение критериальной статистики.
- t критическое двухстороннее — значение квантиля $t_{кр1}$ порядка $1-\alpha/2$ распределения Стьюдента с df степенями свободы.

Нулевая гипотеза H_0: $\mu_1 - \mu_2 = \delta$ принимается, если $|t| < t_{кр1}$ (в противном случае отвергается); гипотеза H_0 при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $t < t_{кр2}$; при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $t_{кр2} < t$.

Как видно из результатов расчета, в данном примере нулевую гипотезу следует отвергнуть при любых конкурирующих гипотезах.

Статистическая функция ТТЕСТ при значении аргумента Тип = 3 (см. раздел 3.8.2) вычисляет вероятности P(T<=t) двухстороннее и P(T<=t) одностороннее.

5.9. Парный двухвыборочный t-тест для средних

Это средство реализует критерий проверки гипотезы о равенстве (неравенстве) математических ожиданий распределений двух зависимых выборок, имеющих нормальные распределения. Этот критерий также называется t-тестом или тестом Стьюдента для парных наблюдений и подробно описан в разделе 2.4.2.

Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_0: $\mu_1 - \mu_2 = \delta$ (δ задано) против разных конкурирующих гипотез: H_1: $\mu_1 \ne \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$ (μ_1 и μ_2 — неизвестные математические ожидания выборок). Рассмотрим пример, когда выборки извлечены из нормально распределенных генеральных совокупностей с математическими ожиданиями соответственно 0 и 1.

Проверим гипотезу, что $\mu_2 - \mu_1 = 1{,}5$ (на самом деле $\mu_2 - \mu_1 = 1$). Исходные данные и заполненное диалоговое окно Парный двухвыборочный t-тест для средних показаны на рис. 5.27.

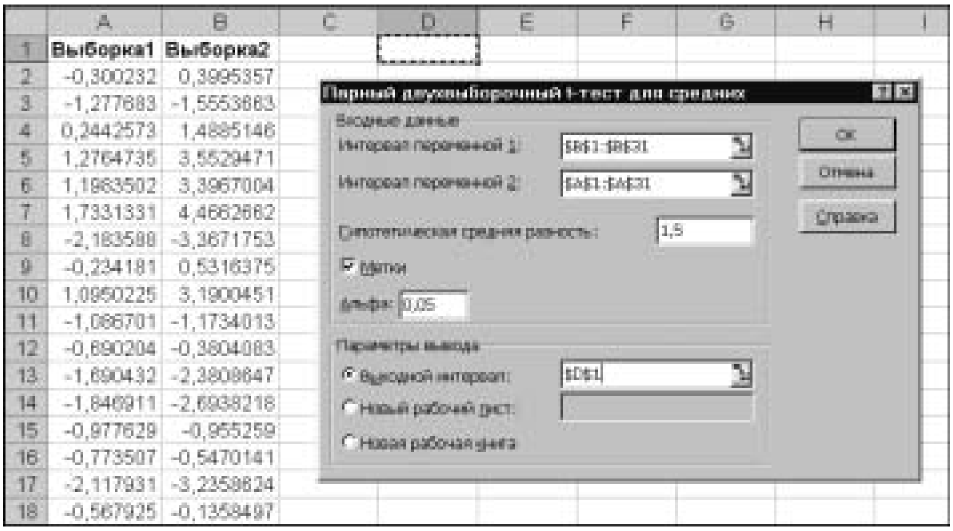


Рис. 5.27. *Исходные данные и диалоговое окно Парный двухвыборочный t-тест для средних*

Отметим, что средство требует, чтобы δ, значение которого задается в поле Гипотетическая средняя разность, было неотрицательно. Поэтому первым (в поле ввода Интервал переменной 1) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле Интервал переменной 2 указывается адрес второй выборки. (Диапазоны должны состоять из одного столбца или одной строки.) В поле Альфа вводится значение уровня значимости α. Результат вычислений средства Парный двухвыборочный t-тест для средних показан на рис. 5.28.

В итоговой таблице приводятся следующие данные.

- Среднее — выборочные средние для каждой выборки.
- Дисперсия — несмещенные выборочные оценки дисперсий выборок.
- Наблюдения — объемы выборок.
- Корреляция Пирсона — выборочный коэффициент корреляции; вычисляется по формуле $r = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}}$.
- Гипотетическая разность средних — значение δ, которое задано в диалоговом окне.

	A	B	C	D	E	F
1	**Выборка1**	**Выборка2**		Парный двухвыборочный t-тест для средних		
2	-0,300232	0,3995357				
3	-1,277683	-1,5553663			*Выборка2*	*Выборка1*
4	0,2442573	1,4885146		Среднее	0,937821973	-0,147166058
5	1,2764735	3,5529471		Дисперсия	9,684347868	1,410018743
6	1,1983502	3,3967004		Наблюдения	30	30
7	1,7331331	4,4662862		Корреляция Пирсона	0,489593926	
8	-2,183588	-3,3671753		Гипотетическая разность средних	1,5	
9	-0,234181	0,5316375		df	29	
10	1,0950225	3,1900451		t-статистика	-0,831355674	
11	-1,086701	-1,1734013		P(T<=t) одностороннее	0,206282928	
12	-0,690204	-0,3804083		t критическое одностороннее	1,699127097	
13	-1,690432	-2,3808647		P(T<=t) двухстороннее	0,412565857	
14	-1,846911	-2,6938218		t критическое двухстороннее	2,045230758	
15	-0,977629	-0,955259				
16	-0,773507	-0,5470141				
17	-2,117931	-3,2358624				

Рис. 5.28. *Результат вычислений*

- df — число степеней свободы, равное $n - 1$.
- t-статистика — значение критериальной статистики; вычисляется по формуле $t = \dfrac{\bar{d} - \delta}{S_n / \sqrt{n}}$, где $\bar{d} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$, $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - y_i - \bar{d})^2$, и имеет распределение Стьюдента с df степенями свободы.
- P(T<=t) одностороннее — вероятность $P(X \le t)$, где X — случайная величина, имеющая распределение Стьюдента с df степенями свободы, t — подсчитанное значение критериальной статистики.
- t критическое одностороннее — значение квантиля $t_{кр2}$ порядка $1 - \alpha$ распределения Стьюдента с df степенями свободы.
- P(T<=t) двухстороннее — вероятность $P(|X| \le |t|)$, где X — случайная величина, имеющая распределение Стьюдента с df степенями свободы, t — подсчитанное значение критериальной статистики.
- t критическое двухстороннее — значение квантиля $t_{кр1}$ порядка $1 - \alpha/2$ распределения Стьюдента с df степенями свободы.

Нулевая гипотеза H_0: $\mu_1 - \mu_2 = \delta$ принимается, если $|t| < t_{кр1}$ (в противном случае отвергается); гипотеза H_0 при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $t < t_{кр2}$; при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $t_{кр2} < t$.

Как видно из результатов расчета, в данном примере нулевую гипотезу следует принять при любых конкурирующих гипотезах.

Статистическая функция ТТЕСТ при значении аргумента Тип = 1 (см. раздел 3.8.2) вычисляет вероятности P(T<=t) двухстороннее и P(T<=t) одностороннее.

5.10. Двухвыборочный F-тест для дисперсий

Это средство реализует критерий Фишера проверки равенства дисперсий двух независимых выборок из нормально распределенных генеральных совокупностей с дисперсиями соответственно σ_1^2 и σ_2^2. Критерий подробно описан в разделе 2.4.2.

Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_0: $\sigma_1^2 = \sigma_2^2$ против конкурирующей гипотезы H_1: $\sigma_1^2 \neq \sigma_2^2$. Рассмотрим пример, когда выборки извлечены из нормально распределенных генеральных совокупностей с равными дисперсиями 1,5. Исходные данные и заполненное диалоговое окно Двухвыборочный F-тест для дисперсий показаны на рис. 5.29.

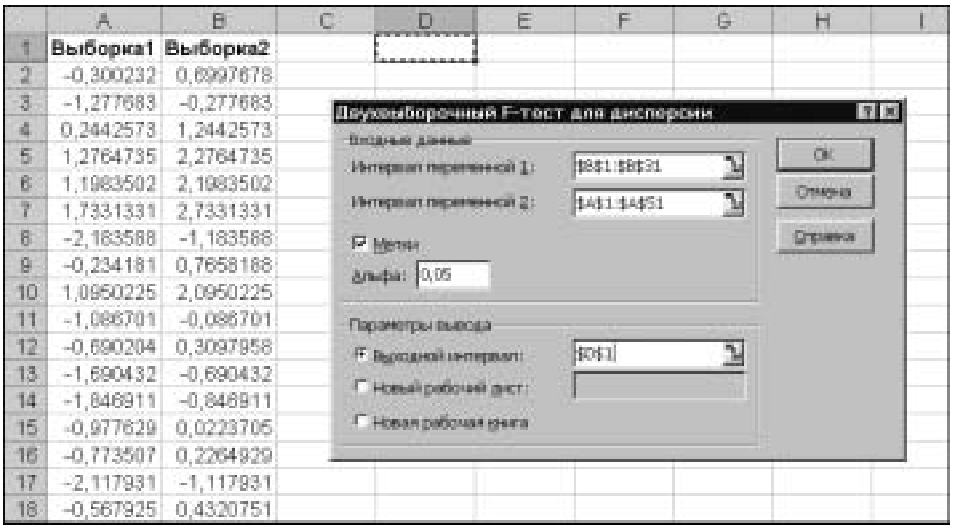


***Рис. 5.29.** Исходные данные и диалоговое окно Двухвыборочный F-тест для дисперсий*

Отметим, что первой (в поле Входной интервал 1) должна задаваться выборка, имеющая бо́льшую дисперсию. В поле Альфа вводится значение уровня значимости α. Результат вычислений средства Двухвыборочный F-тест для дисперсий показан на рис. 5.30.

В итоговой таблице приводятся следующие данные.

- Среднее — выборочные средние для каждой выборки.
- Дисперсия — несмещенные выборочные оценки дисперсий выборок.
- Наблюдения — объемы выборок.
- df — числа степеней свободы, равные $n - 1$ и $m - 1$; n и m — объемы выборок.
- F — значение критериальной статистики, вычисляемой по формуле $F = \frac{S_x^2}{S_y^2}$, где $S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $S_y^2 = \frac{1}{m}\sum_{i=1}^{m}(y_i - \overline{y})^2$, и имеющей F-распределение со степенями свободы $k_1 = n - 1$ и $k_2 = m - 1$ (о F-распределении речь идет в разделе 1.5.7).

	A	B	C	D	E	F
1	**Выборка1**	**Выборка2**		Двухвыборочный F-тест для дисперсии		
2	-0,300232	0,6997678				
3	-1,277683	-0,277683			*Выборка2*	*Выборка1*
4	0,2442573	1,2442573		Среднее	0,852833942	-0,107003257
5	1,2764735	2,2764735		Дисперсия	1,410018743	1,353514772
6	1,1983502	2,1983502		Наблюдения	30	50
7	1,7331331	2,7331331		df	29	49
8	-2,183588	-1,183588		F	1,041746106	
9	-0,234181	0,7658188		P(F<=f) одностороннее	0,439884678	
10	1,0950225	2,0950225		F критическое одностороннее	1,698879259	
11	-1,086701	-0,086701				
12	-0,690204	0,3097958				
13	-1,690432	-0,690432				
14	-1,846911	-0,846911				
15	-0,977629	0,0223705				
16	-0,773507	0,2264929				

Рис. 5.30. *Результат вычислений*

- P(F<=f) одностороннее — вероятность $P(X \le F)$, где X — случайная величина, имеющая F-распределение с df степенями свободы, F — подсчитанное значение критериальной статистики.
- F критическое одностороннее — значение квантиля t порядка $1 - \alpha$ F-распределения с df степенями свободы.

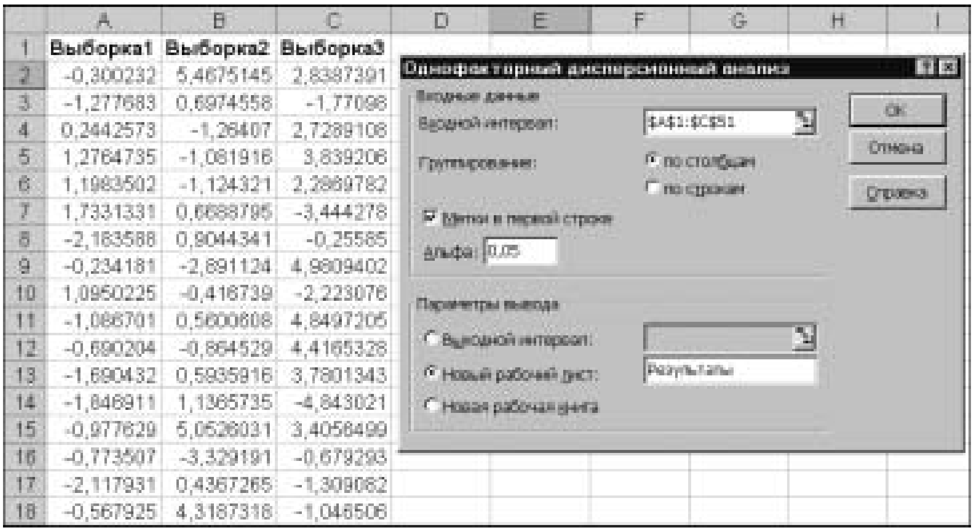
Нулевая гипотеза H_0 принимается, если $F < t$ (в противном случае отвергается). Как видно из результатов расчета, в данном примере нулевую гипотезу следует принять.

Статистическая функция ФТЕСТ (см. раздел 4.8.3) вычисляет удвоенную вероятность P(F<=f) одностороннее.

5.11. Однофакторный дисперсионный анализ

Это средство реализует критерий проверки гипотезы о равенстве математических ожиданий нескольких независимых выборок, построенный на основе дисперсионного анализа. Однофакторный дисперсионный анализ описан в разделе 3.4.2. Здесь покажем применение средства Однофакторный дисперсионный анализ и опишем его выходные данные.

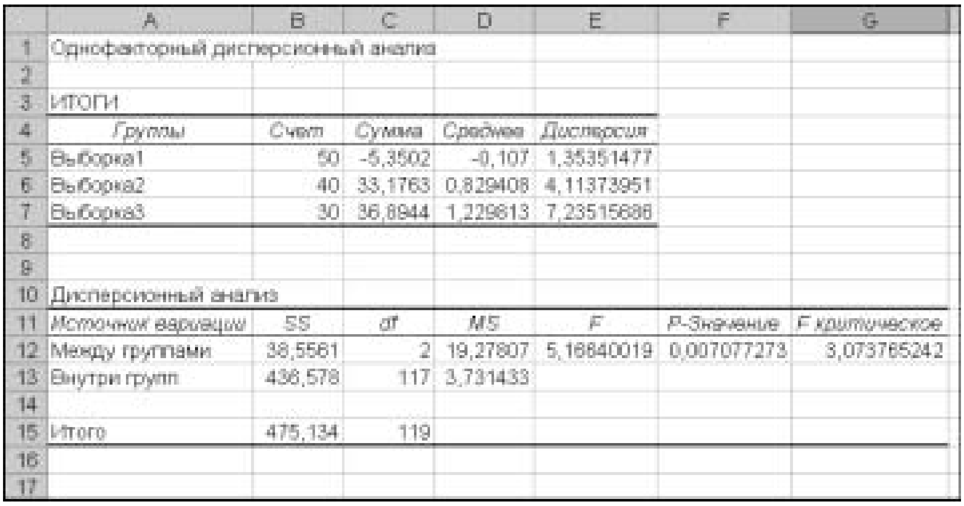
На рис. 5.31 показаны три выборки, имеющие нормальное распределение с математическими ожиданиями 0, 0,5 и 1 и среднеквадратическими отклонениями 1, 2 и 3 соответственно. Объемы выборок — 50, 40 и 30 значений. (Выборки сгенерированы с помощью средства Генерация случайных чисел.) На рис. 5.31 также показано заполненное диалоговое окно Однофакторный дисперсионный анализ. Обращаем внимание, что все три выборки задаются в виде одного диапазона ячеек. В случае, когда выборки имеют разные размеры, диапазон задается в соответствии с наибольшей выборкой и неизбежно содержит пустые ячейки. Но средство правильно определяет объемы выборок. Также отметим, что в данном случае результаты анализа будут выводиться на отдельный рабочий лист с именем Результаты, который автоматически вставится в текущую рабочую книгу.



	A	B	C
1	Выборка1	Выборка2	Выборка3
2	-0,300232	5,4675145	2,8387391
3	-1,277683	0,6974558	-1,77096
4	0,2442573	-1,26407	2,7289108
5	1,2764735	-1,061916	3,839206
6	1,1983502	-1,124321	2,2869782
7	1,7331331	0,6688795	-3,444278
8	-2,163588	0,9044341	-0,25565
9	-0,234181	-2,891124	4,9809402
10	1,0950225	-0,416739	-2,223076
11	-1,086701	0,5600608	4,8497205
12	-0,690204	-0,864529	4,4165328
13	-1,690432	0,5935916	3,7801343
14	-1,846911	1,1365735	-4,843021
15	-0,977629	5,0526031	3,4058499
16	-0,773507	-3,329191	-0,679293
17	-2,117931	0,4367265	-1,309062
18	-0,567925	4,3187318	-1,046506

Рис. 5.31. *Исходные данные и диалоговое окно* Однофакторный дисперсионный анализ

На рис. 5.32 показаны результаты, выводимые средством Однофакторный дисперсионный анализ. Они представлены в виде двух таблиц, озаглавленных ИТОГИ и Дисперсионный анализ. В таблице ИТОГИ выводятся основные статистические характеристики выборок: в столбце Счет — объемы выборок, в столбце Сумма — суммы выборочных значений, в столбцах Среднее и Дисперсия — соответственно выборочные средние и дисперсии.



	A	B	C	D	E	F	G
1	Однофакторный дисперсионный анализ						
2							
3	ИТОГИ						
4	*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
5	Выборка1	50	-5,3502	-0,107	1,35351477		
6	Выборка2	40	33,1763	0,829408	4,11373951		
7	Выборка3	30	36,8944	1,229813	7,23515886		
8							
9							
10	Дисперсионный анализ						
11	*Источник вариации*	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
12	Между группами	38,5561	2	19,27807	5,16840019	0,007077273	3,073765242
13	Внутри групп	436,578	117	3,731433			
14							
15	Итого	475,134	119				
16							
17							

Рис. 5.32. *Результат вычислений*

Значения в первых четырех столбцах таблицы Дисперсионный анализ повторяют значения из дисперсионной таблицы (см. раздел 3.4.2). В столбце *SS* приведены суммы квадратов (межгрупповая, внутригрупповая и полная); в столбце *df* —

значения степеней свободы, а в столбце *MS* — дисперсии, межгрупповая и внутригрупповая. В столбце *F* записано значение критериальной статистики, в столбце *P-Значение* — значение вероятности $P(X \geq x)$, где X — случайная величина, имеющая F-распределение с df степенями свободы (о F-распределении речь идет в разделе 1.5.7). В столбце *F критическое* приводится критическое значение t, рассчитанное в соответствии с заданным уровнем значимости (параметр Альфа). Формулы для вычисления всех перечисленных значений приведены в разделе 3.4.2.

Нулевая гипотеза о равенстве математических ожиданий всех выборок принимается, если выполняется неравенство *F* ≤ *F критическое*. В нашем примере эту гипотезу следует отвергнуть.

5.12. Двухфакторный дисперсионный анализ с повторениями

Двухфакторный дисперсионный анализ описан в разделе 3.5.3. Здесь рассмотрим структуру входных данных для работы с этим средством и опишем выходные результаты. Структура входных данных представлена на рис. 5.33 (обозначения и пояснения приведены в разделе 3.5.3): в строке 1 показаны обозначения уровней фактора β; в столбце A — обозначения уровней фактора γ; в данном случае имеется три выборки, поэтому под общим обозначением уровней фактора γ записаны три строки числовых данных. Таким образом, в диапазоне, например, C8:C10 содержатся выборочные значения, соответствующие второму уровню фактора β и третьему уровню фактора γ.

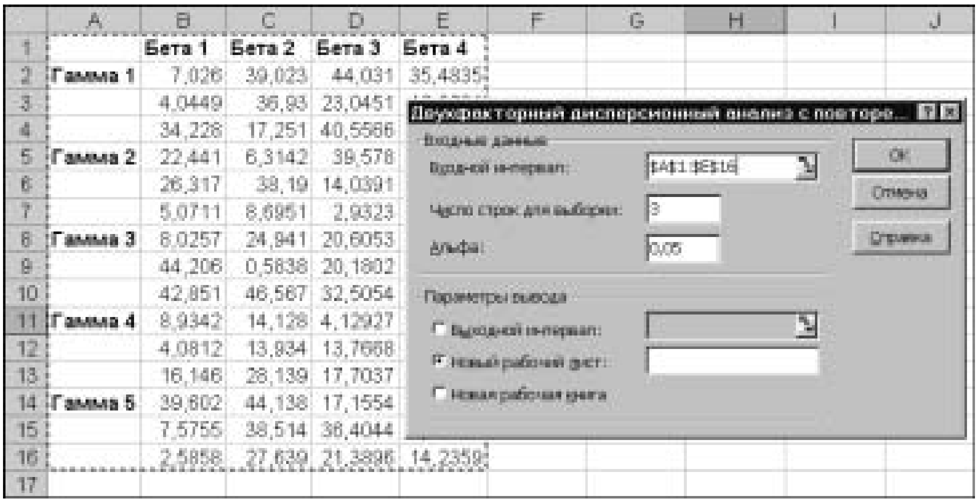


Рис. 5.33. *Исходные данные и диалоговое окно Двухфакторный дисперсионный анализ с повторениями*

Диалоговое окно рассматриваемого здесь средства показано на рис. 5.33. В поле Входной интервал указывается диапазон ячеек, содержащий входные данные, включая заголовки. В поле Число строк для выборки указывается количество рассматриваемых выборок, в данном случае введено число 3. В поле Альфа, как обычно, указывается значение уровня значимости.

На рис. 5.34 показаны выходные результаты работы данного средства, выведенные на отдельный рабочий лист. Выходные результаты сгруппированы в несколько таблиц. В первой таблице, озаглавленной ИТОГИ и состоящей из нескольких подтаблиц (по количеству уровней фактора γ), приводятся статистические характеристики выборочных значений, соответствующих каждому сочетанию уровней фактора β и фактора γ: количество выборочных значений (строка Счет), сумма выборочных значений (строка Сумма), выборочное среднее (строка Среднее) и выборочная дисперсия (строка Дисперсия). На рис. 5.34 показана такая подтаблица для первого уровня фактора γ (таблица обозначена как Гамма 1), другие подобные подтаблицы, соответствующие другим уровням фактора γ, на этом рисунке не показаны. В столбце Итого подтаблиц выводятся такие же статистические характеристики выборочных значений, соответствующие одному уровню фактора γ: количество выборочных значений, выборочное среднее и выборочная дисперсия (вычисляется по всем значениям данного уровня относительно общего среднего). В конце таблицы ИТОГИ выводится подтаблица Итого, в которой приведены те же характеристики, но подсчитанные по выборочным значениям для каждого уровня фактора β.

	A	B	C	D	E	F	G
1	Двухфакторный дисперсионный анализ с повторениями						
3	ИТОГИ	Бета 1	Бета 2	Бета 3	Бета 4	Итого	
4	Гамма 1						
5	Счет	3	3	3	3	12	
6	Сумма	45,299209	93,204725	107,63278	54,691573	300,8282832	
7	Среднее	15,099736	31,068242	35,877592	18,230524	25,0690236	
8	Дисперсия	276,64969	144,27967	126,5216	276,94289	231,244209	
34	Итого						
35	Счет	15	15	15	15		
36	Сумма	273,13636	384,98774	348,02225	322,19117		
37	Среднее	18,209024	25,66585	23,201483	21,479412		
38	Дисперсия	237,85875	217,7876	184,11801	244,1962		
41	Дисперсионный анализ						
42	Источник вариации	SS	df	MS	F	P-Значение	F критическое
43	Выборка	416,05599	4	104,014	0,5097581	0,728830298	2,606972327
44	Столбцы	441,70963	3	147,23654	0,7215858	0,545029484	2,838746127
45	Взаимодействие	3517,5466	12	293,12888	1,4365839	0,190048828	2,003459509
46	Внутри	8161,8313	40	204,04578			
47							
48	Итого	12537,143	59				
49							

Рис. 5.34. *Выходные результаты работы средства Двухфакторный дисперсионный анализ с повторениями*

В нижней части выходных результатов приведена дисперсионная таблица (обозначения и вычисляющие формулы даны в разделе 3.5.3). Здесь в первом столбце, обозначенном SS, выведены суммы квадратов: соответственно SS_1, SS_2, SS_3, SS_4 и в строке Итого — SS. В столбце df приведены степени свободы сумм квадратов, а в столбце MS — значения соответствующих дисперсий. В столбце F вычислены значения критериальных статистик, т.е. отношения дисперсий s_1^2, s_2^2, s_3^2 к дисперсии s_4^2.

В столбце P-Значение вычисляются вероятности $P(X \geq F)$, где X — случайная величина, имеющая F-распределение со степенями свободы, значения которых приведены в столбце df: первое значение степени свободы — из соответствующей строки этого столбца, а второе — всегда из четвертой строки, F — значение из столбца F. Например, значение в ячейке F43 (см. рис. 5.34), можно вычислить по формуле Excel =FРАСП(E43;C43;C46). Эти значения используются для проверки гипотез о значимом влиянии факторов или их взаимного влияния: если вероятность больше заданного уровня значимости, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается.

В столбце F критическое вычисляются критические значения, соответствующие заданному в диалоговом окне Двухфакторный дисперсионный анализ с повторениями уровню значимости α. Эти значения вычисляются как квантили порядка $1 - \alpha$ F-распределения со степенями свободы, значения которых определяются так же, как при вычислении вероятностей из столбца P-Значение. Например, значение в ячейке G43 (см. рис. 5.34) можно вычислить по формуле Excel =FРАСПОБР(0,05;C43;C46). Эти значения используются для проверки гипотез о значимом влиянии факторов или их взаимного влияния: если значение в этом столбце больше значения в столбце F той же строки, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается. Здесь принимаются все три нулевые гипотезы об отсутствии влияния факторов β и γ и их взаимного влияния. Однако значение в столбце F третьей строки (соответствует взаимному влиянию факторов) значительно больше аналогичных значений для отдельных факторов, и на это необходимо обратить внимание.

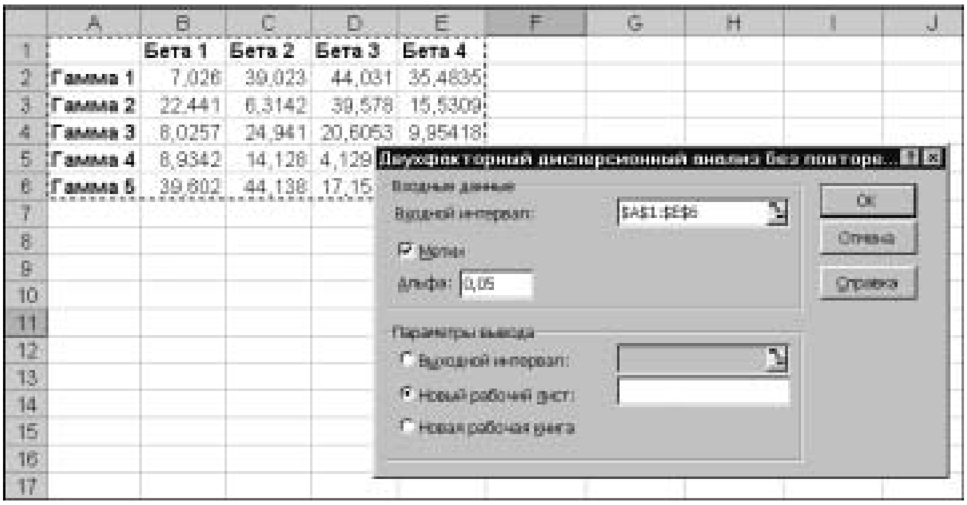
5.13. Двухфакторный дисперсионный анализ без повторений

Двухфакторный дисперсионный анализ описан в разделе 3.5.3. Структура входных данных показана на рис. 5.35 (обозначения и пояснения даны в разделе 3.5.3): в строке 1 приводятся обозначения уровней фактора β; в столбце A — обозначения уровней фактора γ; в диапазоне, обозначенном этими заголовками, введены числовые данные.

Диалоговое окно этого средства показано на рис. 5.35. В поле Входной интервал указывается диапазон ячеек, содержащий входные данные; если в этот диапазон включены заголовки строк и столбцов, то следует установить флажок опции Метки. В поле Альфа указывается значение уровня значимости.

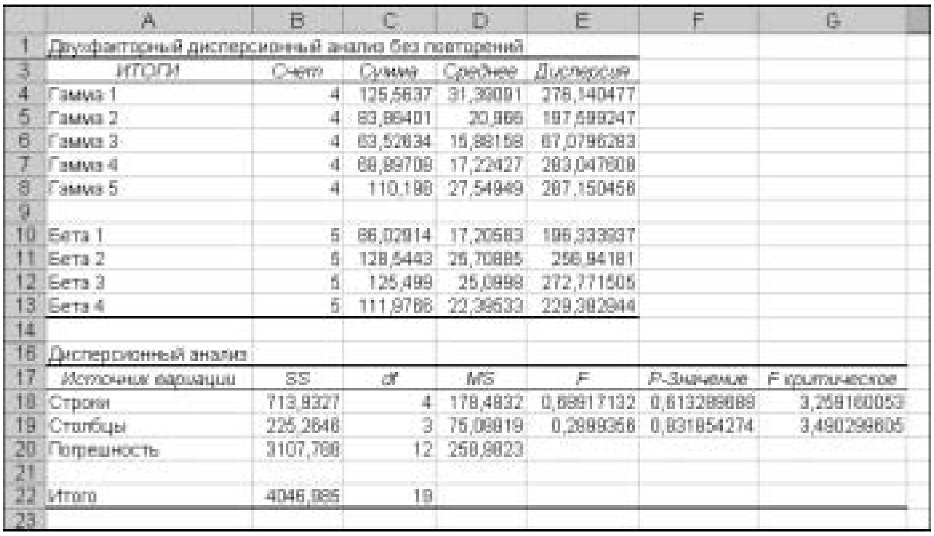
На рис. 5.36 представлены выходные результаты работы данного средства, выведенные на отдельный рабочий лист. Выходные результаты сгруппированы в две таблицы. В первой таблице, озаглавленной ИТОГИ, приводятся статистические характеристики выборочных значений, соответствующих каждому уровню фактора β (группировка по столбцам) и каждому уровню фактора γ (группировка по строка): количество выборочных значений (столбец Счет), сумма выборочных значений (столбец Сумма), выборочное среднее (столбец Среднее) и выборочная дисперсия (столбец Дисперсия).

В нижней части выходных результатов приведена дисперсионная таблица (обозначения и вычисляющие формулы даны в разделе 3.5.3). Здесь в первом столбце, обозначенном SS, выведены суммы квадратов: соответственно SS_1, SS_2, SS_3 и в строке Итого — SS. В столбце df приведены степени свободы сумм квадратов, а в столбце MS — значения соответствующих дисперсий. В столбце F вычислены значения критериальных статистик, т.е. отношения дисперсий s_1^2 и s_2^2 к дисперсии s_3^2.



	Бета 1	Бета 2	Бета 3	Бета 4
Гамма 1	7,026	39,023	44,031	35,4835
Гамма 2	22,441	6,3142	39,578	15,5309
Гамма 3	8,0257	24,941	20,6053	9,95418
Гамма 4	8,9342	14,128	4,129	
Гамма 5	39,602	44,138	17,15	

Рис. 5.35. *Исходные данные и диалоговое окно* *Двухфакторный дисперсионный анализ без повторений*



Двухфакторный дисперсионный анализ без повторений

ИТОГИ	Счет	Сумма	Среднее	Дисперсия
Гамма 1	4	125,5637	31,39091	278,140477
Гамма 2	4	83,86401	20,966	197,699247
Гамма 3	4	63,52634	15,88158	67,0796283
Гамма 4	4	68,89709	17,22427	283,047808
Гамма 5	4	110,188	27,54949	287,150458
Бета 1	5	86,02914	17,20583	196,333937
Бета 2	5	128,5443	25,70885	256,94181
Бета 3	5	125,499	25,0998	272,771505
Бета 4	5	111,9766	22,39533	229,382944

Дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Строки	713,9327	4	178,4832	0,68917132	0,613289888	3,259160053
Столбцы	225,2646	3	75,08819	0,2899358	0,831854274	3,490299605
Погрешность	3107,788	12	258,9823			
Итого	4046,985	19				

Рис. 5.36. *Выходные результаты работы средства* *Двухфакторный дисперсионный анализ без повторений*

В столбце P-Значение вычисляются вероятности $P(X \geq F)$, где X — случайная величина, имеющая F-распределение со степенями свободы, значения которых приведены в столбце df: первое значение степени свободы — из соответствующей строки этого столбца, а второе — всегда из третьей строки, F — значение из столбца F. Например, значение в ячейке E18 (см. рис. 5.36) можно вычислить по формуле Excel =FРАСП(D18;C18;C20). Эти значения используются для проверки гипотез о значимом влиянии факторов: если вероятность больше заданного уровня значимости, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается.

В столбце F критическое вычисляются критические значения, соответствующие заданному в диалоговом окне Двухфакторный дисперсионный анализ без повторений уровню значимости α. Эти значения вычисляются как квантили порядка $1 - \alpha$ F-распределения со степенями свободы, значения которых определяются так же, как при вычислении вероятностей из столбца P-Значение. Например, значение в ячейке G18 (см. рис. 5.36) можно вычислить по формуле Excel =FРАСПОБР(0,05;C18;C20). Эти значения используются для проверки гипотез о значимом влиянии факторов или их взаимного влияния: если значение в этом столбце больше значения в столбце F той же строки, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается. Здесь принимаются обе нулевые гипотезы об отсутствии влияния факторов β и γ.

5.14. Корреляция

Это средство вычисляет корреляционную матрицу компонентов многомерной выборки. Диагональные элементы матрицы равны единице, а внедиагональные — коэффициентам корреляции соответствующих компонентов (о коэффициентах корреляции речь идет в разделе 1.2.5). На рис. 5.37 показаны многомерная выборка, имеющая совместное нормальное распределение, причем первая пара компонентов зависима с коэффициентом корреляции 0,5. С таким же коэффициентом корреляции зависимы третий и четвертый компоненты выборки. Первая и вторая пара компонентов между собой независимы.

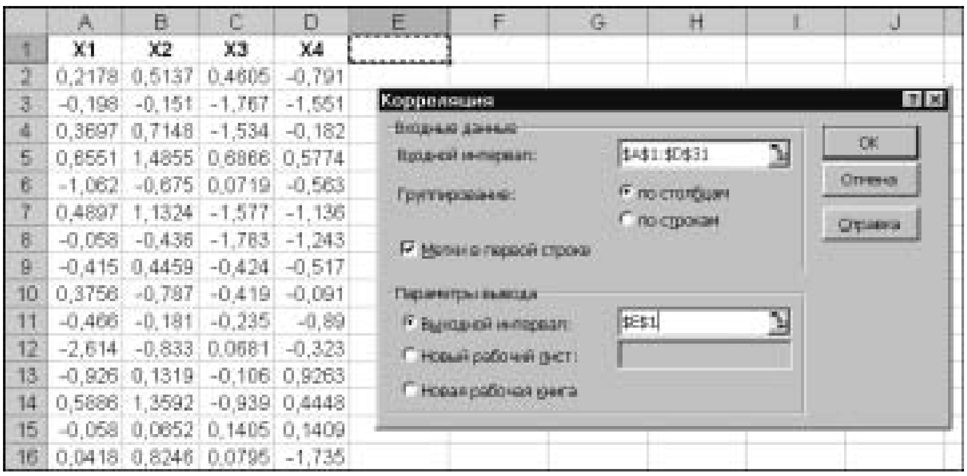


Рис. 5.37. Исходные данные и диалоговое окно Корреляция

Внедиагональные элементы корреляционной матрицы рассчитываются по стандартным формулам: коэффициент корреляции r_{xy} между компонентами x и y многомерной выборки вычисляется как

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}}, \text{ где } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \ \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \ n \text{ — объем выборки.}$$

Отметим, что эти же вычисления выполняет функция КОРРЕЛ (см. раздел 4.10.2).

На рис. 5.38 показан результат применения средства Корреляция. Поскольку корреляционная матрица симметрична, выводится только нижняя ее половина.

	A	B	C	D	E	F	G	H	I
1	X1	X2	X3	X4		X1	X2	X3	X4
2	0,2178	0,5137	0,4605	-0,791	X1	1			
3	-0,198	-0,151	-1,767	-1,551	X2	0,539796	1		
4	0,3897	0,7148	-1,534	-0,162	X3	-0,11155	-0,05404	1	
5	0,6551	1,4855	0,6866	0,5774	X4	0,07675	0,079159	0,49465	1
6	-1,062	-0,675	0,0719	-0,563					
7	0,4897	1,1324	-1,577	-1,136					
8	-0,058	-0,436	-1,783	-1,243					
9	-0,415	0,4459	-0,424	-0,517					
10	0,3756	-0,787	-0,419	-0,091					
11	-0,466	-0,181	-0,235	-0,89					
12	-2,614	-0,833	0,0681	-0,323					

Рис. 5.38. *Результат применения средства* Корреляция

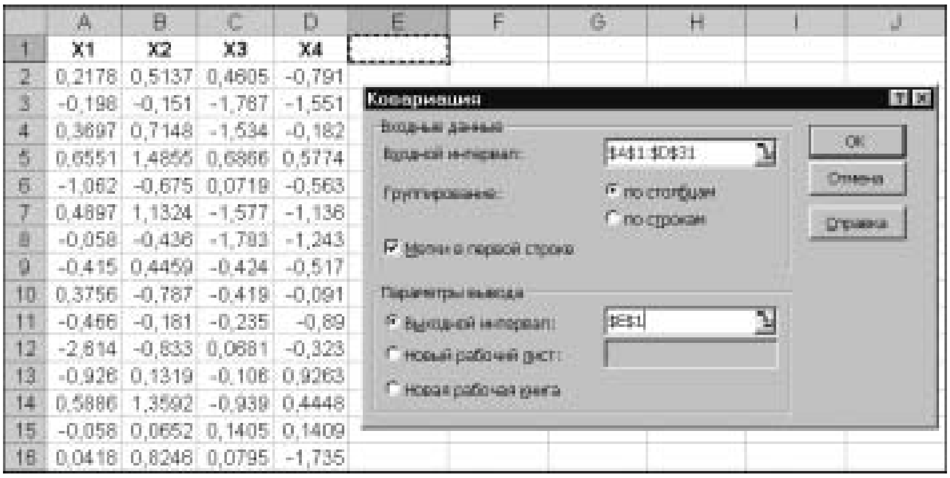
5.15. Ковариация

Это средство вычисляет ковариационную матрицу компонентов многомерной выборки. Диагональные элементы матрицы равны выборочным дисперсиям, а внедиагональные — ковариациям соответствующих компонентов (о ковариациях речь идет в разделе 1.2.5). На рис. 5.39 показана многомерная выборка, имеющая совместное нормальное распределение, причем первая пара компонентов зависима с коэффициентом корреляции 0,5. С таким же коэффициентом корреляции зависимы третий и четвертый компоненты выборки. Первая и вторая пары компонентов между собой независимы.

Внедиагональные элементы ковариационной матрицы рассчитываются по формулам: ковариация cov(X, Y) между компонентами x и y многомерной выборки вычисляется как

$$\operatorname{cov}(X, Y) = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}), \text{ где } \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \ \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \ n \text{ — объем выборки.}$$

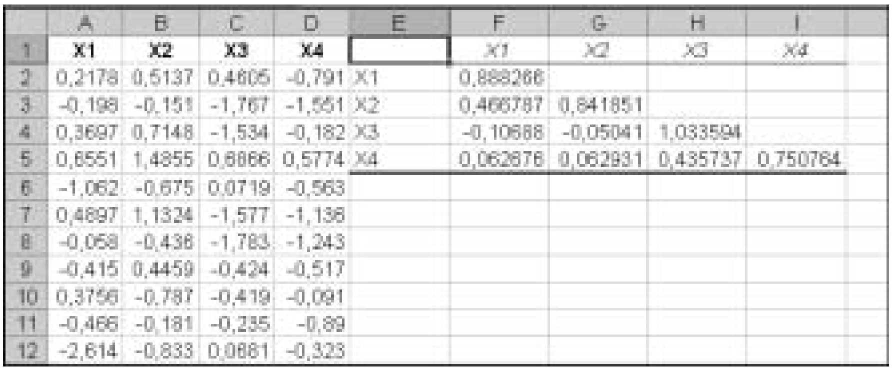
Отметим, что эти же вычисления выполняет функция КОВАР (см. раздел 4.10.1). Диагональные элементы матрицы — выборочные дисперсии — вычисляются по стандартным формулам $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Выборочную дисперсию также вычисляют функции ДИСПР и ДИСПРА (см. раздел 4.5.2).



	A	B	C	D
1	X1	X2	X3	X4
2	0,2178	0,5137	0,4605	-0,791
3	-0,198	-0,151	-1,787	-1,551
4	0,3697	0,7148	-1,534	-0,182
5	0,6551	1,4855	0,6866	0,5774
6	-1,062	-0,675	0,0719	-0,563
7	0,4897	1,1324	-1,577	-1,136
8	-0,058	-0,436	-1,783	-1,243
9	-0,415	0,4459	-0,424	-0,517
10	0,3756	-0,787	-0,419	-0,091
11	-0,466	-0,181	-0,235	-0,89
12	-2,614	-0,833	0,0681	-0,323
13	-0,926	0,1319	-0,106	0,9263
14	0,5886	1,3592	-0,939	0,4448
15	-0,058	0,0652	0,1405	0,1409
16	0,0418	0,8246	0,0795	-1,735

Рис. 5.39*. Исходные данные и диалоговое окно* Ковариация

На рис. 5.40 показан результат применения средства Ковариация. Поскольку ковариационная матрица симметрична, выводится только нижняя ее половина.



	A	B	C	D	E	F	G	H	I
1	X1	X2	X3	X4		X1	X2	X3	X4
2	0,2178	0,5137	0,4605	-0,791	X1	0,888266			
3	-0,198	-0,151	-1,787	-1,551	X2	0,466787	0,841851		
4	0,3697	0,7148	-1,534	-0,182	X3	-0,10688	-0,05041	1,033594	
5	0,6551	1,4855	0,6866	0,5774	X4	0,062876	0,062931	0,435737	0,750784
6	-1,062	-0,675	0,0719	-0,563					
7	0,4897	1,1324	-1,577	-1,136					
8	-0,058	-0,436	-1,783	-1,243					
9	-0,415	0,4459	-0,424	-0,517					
10	0,3756	-0,787	-0,419	-0,091					
11	-0,466	-0,181	-0,235	-0,89					
12	-2,614	-0,833	0,0681	-0,323					

Рис. 5.40*. Результат применения средства* Ковариация

5.16. Регрессия

Задачи регрессионного анализа описаны в разделе 3.4. Покажем, что для проведения регрессионного анализа может сделать средство Регрессия. В отдельных таблицах оно вычисляет (рис. 5.42 и 5.43) следующее:

- методом наименьших квадратов — коэффициенты линейной (относительно этих коэффициентов) функции регрессии; вид функции регрессии определяется структурой исходных данных (подробнее об этом речь идет ниже);

- коэффициент детерминации и связанные с ним величины (таблица Регрессионная статистика);
- дисперсионную таблицу и критериальную статистику для проверки значимости регрессии (таблица Дисперсионный анализ);
- для каждого коэффициента регрессии — среднеквадратическое отклонение и другие его статистические характеристики, позволяющие проверить значимость этого коэффициента и построить для него доверительные интервалы;
- значения функции регрессии и *остатки* — разности между исходными значениями переменной Y и вычисленными значениями функции регрессии (таблица Вывод остатка);
- вероятности, соответствующие упорядоченным по возрастанию значениям переменной Y (таблица Вывод вероятности).

Кроме того, средство Регрессия строит три типа графиков, которые будут показаны ниже.

Пусть входной интервал X состоит из k диапазонов-столбцов, содержащих значения $\{x_{i1}\}$, $\{x_{i2}\}$, ..., $\{x_{ik}\}$ переменных X_1, X_2, ..., X_k. В каждом диапазоне содержится одинаковое количество значений. Входной интервал Y, состоящий из одного диапазона-столбца, должен содержать такое же количество значений. Средство Регрессия вычисляет коэффициенты функции регрессии вида

$$Y = m_1X_1 + m_2X_2 + \ldots + m_kX_k + b.$$

Это уравнение линейной множественной регрессии, если переменные X_i независимы. На основе данного уравнения, используя соответствующие значения переменных X_i, можно получить множество других уравнений регрессии. Например, если в качестве переменных X_i взять значения одной переменной X в степени i (т.е. $X_i = X^i$), получим уравнение полиномиальной регрессии

$$Y = m_1X + m_2X^2 + \ldots + m_kX^k + b.$$

На рис. 5.41 показан рабочий лист с исходными данными: входной интервал X состоит из пяти столбцов. В первом столбце представлены значения переменной X_1, во втором — квадраты значений переменной X_1, в третьем — значения второй переменной X_2, в четвертом — квадраты значений переменной X_2, в пятом — произведения значений переменных X_1 и X_2. Таким образом, в данном случае Регрессия будет вычислять значения коэффициентов уравнения регрессии вида

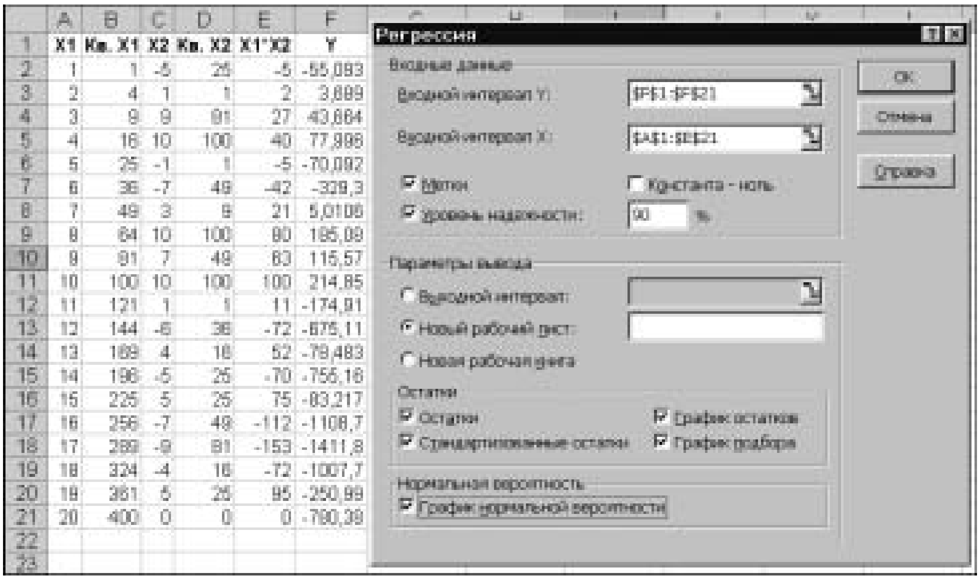
$$Y = m_1X_1 + m_2X_1^2 + m_3X_2 + m_4X_2^2 + m_5X_1X_2 + b.$$

Отметим, что значения зависимой переменной Y в столбце F получены по формуле

$$Y = X_1 - 2X_1^2 + 0{,}5X_2 - X_2^2 + 5X_1X_2 + \varepsilon.$$

Здесь случайная переменная ε имеет стандартное нормальное распределение. (О моделировании случайных величин речь идет в главе 7.)

Диалоговое окно средства Регрессия показано на рис. 5.41. В поле Входной интервал Y вводится адрес диапазона, содержащего значения зависимой переменной Y. Диапазон должен состоять из одного столбца. В поле Входной интервал X вводится адрес диапазона, содержащего значения переменной X. Диапазон должен состоять из одного или нескольких столбцов, но не более чем из 16 столбцов. Если указанные в полях Входной интервал Y и Входной интервал X диапазоны включают заголовки столбцов, то необходимо установить флажок опции Метки — эти заголовки будут использованы в выходных таблицах, сгенерированных средством Регрессия.



	A	B	C	D	E	F
1	X1	Кв. X1	X2	Кв. X2	X1*X2	Y
2	1	1	-5	25	-5	-55,093
3	2	4	1	1	2	3,889
4	3	9	9	81	27	43,864
5	4	16	10	100	40	77,998
6	5	25	-1	1	-5	-70,092
7	6	36	-7	49	-42	-329,3
8	7	49	3	9	21	5,0106
9	8	64	10	100	80	185,09
10	9	81	7	49	63	115,57
11	10	100	10	100	100	214,85
12	11	121	1	1	11	-174,91
13	12	144	-6	36	-72	-675,11
14	13	169	4	16	52	-79,483
15	14	196	-5	25	-70	-756,16
16	15	225	5	25	75	-83,217
17	16	256	-7	49	-112	-1108,7
18	17	289	-9	81	-153	-1411,8
19	18	324	-4	16	-72	-1007,7
20	19	361	5	25	95	-250,99
21	20	400	0	0	0	-780,39
22						
23						

Рис. 5.41. *Исходные данные и диалоговое окно* Регрессия

Флажок опции Константа - ноль следует установить, если в уравнении регрессии константа *b* принудительно полагается равной нулю. Опция Уровень надежности устанавливается тогда, когда необходимо построить доверительные интервалы для коэффициентов регрессии с доверительным уровнем, отличным от 0,95, который используется по умолчанию. После установки флажка опции Уровень надежности становится доступным поле ввода, в котором вводится новое значение доверительного уровня.

В области Остатки имеются четыре опции: Остатки, Стандартизованные остатки, График остатков и График подбора. Если установлена хотя бы одна из них, то в выходных результатах появится таблица Вывод остатка, в которой будут выведены значения функции регрессии и остатки — разности между исходными значениями переменной Y и вычисленными значениями функции регрессии. Значения этой таблицы и возможности каждой из опций показаны ниже.

В области Нормальная вероятность имеется одна опция — График нормальной вероятности; ее установка порождает в выходных результатах таблицу Вывод вероятности и приводит к построению соответствующего графика.

На рис. 5.42–5.44 показаны части рабочего листа с выходными результатами средства Регрессия, которые получены на основе исходных данных, приведенных на рис. 5.41. Рассмотрим подробнее эти результаты.

В таблице Регрессионная статистика приводятся следующие данные.

- Множественный R — корень из коэффициента детерминации R^2, приведенного в следующей строке. Другое название этого показателя — индекс корреляции, или множественный коэффициент корреляции (см. раздел 3.3.1).

	A	B	C	D	E	F
1	ВЫВОД ИТОГОВ					
2						
3	*Регрессионная статистика*					
4	Множественный R	0,999998881				
5	R-квадрат	0,999997761				
6	Нормированный R-квадрат	0,999996962				
7	Стандартная ошибка	0,83074245				
8	Наблюдения	20				
9						
10	Дисперсионный анализ					
11		*df*	*SS*	*MS*	*F*	*Значимость F*
12	Регрессия	5	4315679,352	863135,8704	1250680,44	5,01903E-39
13	Остаток	14	9,661862256	0,690133016		
14	Итого	19	4315689,014			
15						

Рис. 5.42. Верхняя часть рабочего листа с выходными результатами

	A	B	C	D	E	F	G	H	I
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 90*	*Верхние 9*
17	Y-пересечение	-0,253888671	0,848126025	-0,391699463	0,7011799	-1,643962	1,1362464	-1,395419	0,887681
18	X1	1,123005378	0,138017336	8,136697954	1,124E-06	0,8269874	1,41902339	0,8799142	1,366096
19	Кв. X1	-2,005996197	0,006477434	-309,6883928	2,956E-28	-2,0198799	-1,9920935	-2,017395	-1,994577
20	X2	0,672289116	0,083366283	8,064280793	1,247E-06	0,4934861	0,85109217	0,5254553	0,819122
21	Кв. X2	-1,010435244	0,007061425	-143,0922648	1,457E-23	-1,0255806	-0,99529	-1,022873	-0,997998
22	X1*X2	4,98588398	0,007028913	709,3407926	2,704E-33	4,9708104	5,00096949	4,9735139	4,998274
23									
26	ВЫВОД ОСТАТКА					ВЫВОД ВЕРОЯТНОСТИ			
28	*Наблюдение*	*Предсказанное Y*	*Остатки*	*Стандартные остатки*		*Персентиль*	*Y*		
29	1	-54,68864596	-0,404791015	-0,567645558		2,5	-1411,7816		
30	2	3,601839089	0,087198777	0,12226916		7,5	-1108,7249		
31	3	43,68575591	-0,221342134	-0,310391966		12,5	-1007,6697		
32	4	77,25749899	0,738473754	1,03557473		17,5	-780,37807		
33	5	-71,40069086	1,30825897	1,834594557		22,5	-765,15699		
34	6	-329,3562365	0,051581852	0,072334139		27,5	-675,10838		
35	7	6,940568647	-1,929928977	-2,706373339		32,5	-329,30465		
36	8	194,8979413	0,184368329	0,258542849		37,5	-250,99418		
37	9	116,6743141	-1,106508782	-1,55167673		42,5	-174,91074		
45	17	-1411,630421	-0,131188723	-0,183968253		82,5	77,9959726		
46	18	-1007,819785	0,15010449	0,210494165		87,5	115,567805		
47	19	-251,3172928	0,323114831	0,453109609		92,5	185,08231		
48	20	-780,1882398	-0,187832093	-0,263400247		97,5	214,845179		
49									

Рис. 5.43. *Нижняя часть рабочего листа с выходными результатами*

- R-квадрат — коэффициент детерминации R^2; вычисляется как отношение регрессионной суммы квадратов (ячейка C12) к полной сумме квадратов (ячейка C14). (О коэффициенте детерминации речь идет в разделе 3.4.3.)
- Нормированный R-квадрат вычисляется по формуле $\frac{(n-1)R^2-k}{n-k-1}$, где n — количество значений переменной Y, k — количество столбцов во входном интервале переменной X.
- Стандартная ошибка — корень из остаточной дисперсии (ячейка D13).
- Наблюдения — количество значений переменной Y.

Дисперсионная таблица соответствует аналогичной таблице из раздела 3.4.3. В столбце SS приводятся суммы квадратов, в столбце df — число степеней свободы, в столбце MS — дисперсии. Строка Регрессия соответствует одноименной строке из таблицы в разделе 3.4.3, строка Остаток — строке Остатки и строка Итого — строке Полная. В дисперсионной таблице из раздела 3.4.3 приведены формулы, по которым вычисляет соответствующие значения средство Регрессия. В столбце F вычислено значение критериальной статистики для проверки значимости регрессии. Это значение вычисляется как отношение регрессионной дисперсии к остаточной (ячейки D12 и D13). В столбце Значимость F вычисляется вероятность полученного значения критериальной статистики. (Эту вероятность с помощью формул Excel можно вычислить как =FРАСП(E12;B12;B13).) Если эта вероятность меньше, например, 0,05 (заданного уровня значимости), то гипотеза о незначимости регрессии (т.е. гипотеза о том, что все коэффициенты функции регрессии равны нулю) отвергается и считается, что регрессия значима. В данном примере регрессия значима практически с любым уровнем значимости.

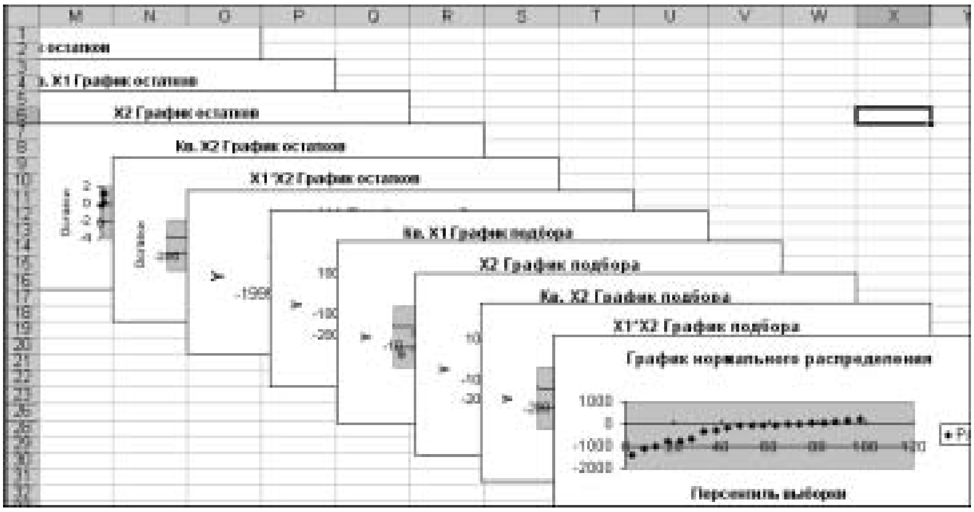


Рис. 5.44. Графики, выводимые средством Регрессия

В следующей таблице (см. рис. 5.43), в столбце Коэффициенты, записаны вычисленные значения коэффициентов функции регрессии, при этом в строке Y-пересечение записано значение свободного члена b. В столбце Стандартная ошибка вычислены среднеквадратические отклонения коэффициентов (о вычислении дисперсий коэффициентов речь идет в разделе 3.4.4). В столбце t-статистика записаны отношения значений коэффициентов к их среднеквадратическим отклонениям. Это значения критериальных статистик для проверки гипотез о значимости коэффициентов регрессии. В столбце Р-Значение вычисляются уровни значимости, соответствующие значениям критериальных статистик. (Их можно вычислить с помощью формулы Excel =СТЬЮДРАСП(ABS(D17);14;2), например, для значения в ячейке E17; второй аргумент в функции СТЬЮДРАСП вычисляется как $n - k - 1$.) Если вычисленный уровень значимости меньше заданного уровня значимости (например, 0,05), то принимается гипотеза о значимом отличии коэффициента от нуля; в противном случае принимается гипотеза о незначимом отличии коэффициента от нуля. В данном примере только коэффициент b незначимо отличается от нуля.

В столбцах Нижние 95% и Верхние 95% приводятся границы доверительных интервалов с доверительным уровнем 0,95. Эти границы вычисляются по формулам

$$\text{Нижние 95\%} = \text{Коэффициент} - \text{Стандартная ошибка} \times t_\alpha;$$

$$\text{Верхние 95\%} = \text{Коэффициент} + \text{Стандартная ошибка} \times t_\alpha.$$

Здесь t_α — квантиль порядка α распределения Стьюдента с $(n - k - 1)$ степенью свободы. В данном случае $\alpha = 0{,}95$. Аналогично вычисляются границы доверительных интервалов в столбцах Нижние 90,0% и Верхние 90,0%. Отметим, что если в диалоговом окне Регрессия не устанавливать опцию Уровень надежности, то будут повторены столбцы Нижние 95% и Верхние 95%.

Рассмотрим таблицу Вывод остатка из выходных результатов средства Регрессия. Напомним, что эта таблица появляется в выходных результатах только тогда, когда установлена хотя бы одна опция в области Остатки диалогового окна Регрессия. В столбце Наблюдение приводятся порядковые номера значений переменной Y. В столбце Предсказанное Y вычисляются значения функции регрессии $\hat{y}_i = f(x_i)$ для тех значений переменной X, которым соответствует порядковый номер i в столбце Наблюдение. В столбце Остатки содержатся разности (остатки) $\varepsilon_i = y_i - \hat{y}_i$, а в столбце Стандартные остатки — нормированные остатки, которые вычисляются как отношения $\varepsilon_i/s_\varepsilon$, где s_ε — среднеквадратическое отклонение остатков. Квадрат величины s_ε вычисляется по формуле $s_\varepsilon^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \overline{\varepsilon})^2$, где $\overline{\varepsilon}$ — среднее остатков. Здесь величину s_ε^2 можно вычислить как отношение двух значений из дисперсионной таблицы: суммы квадратов остатков (ячейка C13) и степени свободы из строки Итого (ячейка B14).

По значениям таблицы Вывод остатка средство Регрессия строит два типа графиков: графики остатков и графики подбора (если установлены соответствующие опции в области Остатки диалогового окна Регрессия). На рис. 5.45 показаны образцы этих графиков (графики немного переформатированы по сравнению с оригиналами). Они строятся для каждого компонента переменной X в отдельности. На графиках остатков отображаются остатки, т.е. разности между исходными значениями Y и вычисленными по функции регрессии для каждого значения компонента переменной X. На графиках подбора отображаются как исходные значения Y, так и вычисленные значения функции регрессии для каждого значения компонента переменной X. (На графиках подбора, представленных на рис. 5.45, эти значения практически совпадают.)

Последней таблицей выходных результатов средства Регрессия является таблица Вывод вероятности (см. рис. 5.43). Она появляется, если в диалоговом окне Регрессия установлена опция График нормальной вероятности. Значения в столбце Персентиль вычисляются следующим образом. Вычисляется шаг $h = (1/n)\times 100\%$, первое значение равно $h/2$, последнее равно $100 - h/2$. Начиная со второго значения каждое последующее значение равно предыдущему, к которому прибавлен шаг h. В столбце Y приведены значения переменной Y, упорядоченные по возрастанию. По данным этой таблицы строится так называемый график нормального распределения (рис. 5.46). Он позволяет визуально оценить степень линейности зависимости между переменными X и Y.

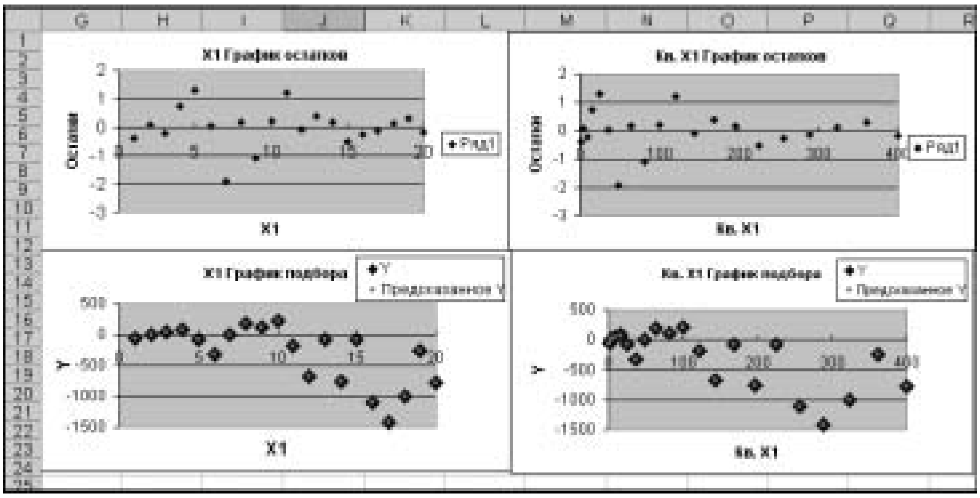


Рис. 5.45. Примеры графиков остатков и подбора

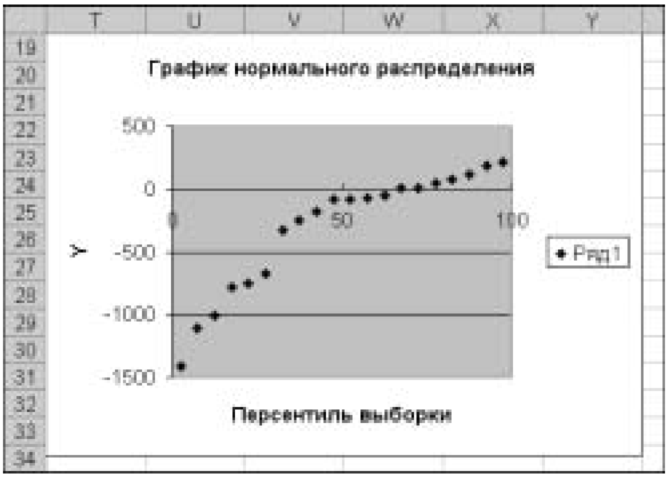


Рис. 5.46. График нормального распределения

5.17. Скользящее среднее

Метод скользящего среднего — один из наиболее широко используемых способов сглаживания значений временного ряда. Метод основан на локальном усреднении, когда за новое значение временного ряда берется среднее k последовательных значений, ближайших к заменяемому значению.

Пусть имеются дискретные наблюдения y_1, y_2, ..., y_n и задано число k наблюдений, по которым будет проводиться усреднение. Значение скользящего среднего для значения t вычисляется по формуле $y_t = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-i}$. Отметим, что по этой формуле выполняет вычисления средство Скользящее среднее, но существуют и другие способы вычисления скользящего среднего.

На рис. 5.47 показаны исходные данные, для которых будут вычисляться скользящие средние, и диалоговое окно Скользящее среднее. В поле ввода Входной интервал в качестве исходных данных задан диапазон B1:B17. Поскольку этот диапазон содержит заголовок, установлен флажок опции Метки в первой строке. В поле Интервал вводится число k — количество значений, по которым подсчитывается скользящее среднее. Если этот параметр не задан, то по умолчанию используется значение 3.

Если установлен флажок опций Вывод графика, то будет построен график, отображающий исходные значения y_i и сглаженные скользящим средним значения (рис. 5.48). Если также установлен флажок опции Стандартные погрешности, то к значениям вычисленных средних будет добавлен столбец, в котором будут записаны стандартные погрешности, вычисляемые как сумма квадратов разностей между исходными и расчетными k значения y_i, деленная на число k. Формула Excel, по которой подсчитываются стандартные погрешности, показана на рис. 5.48.

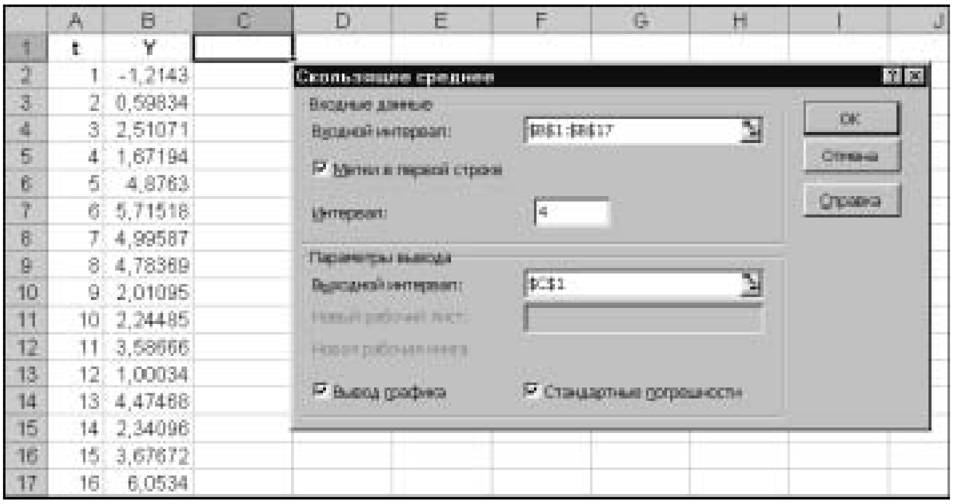
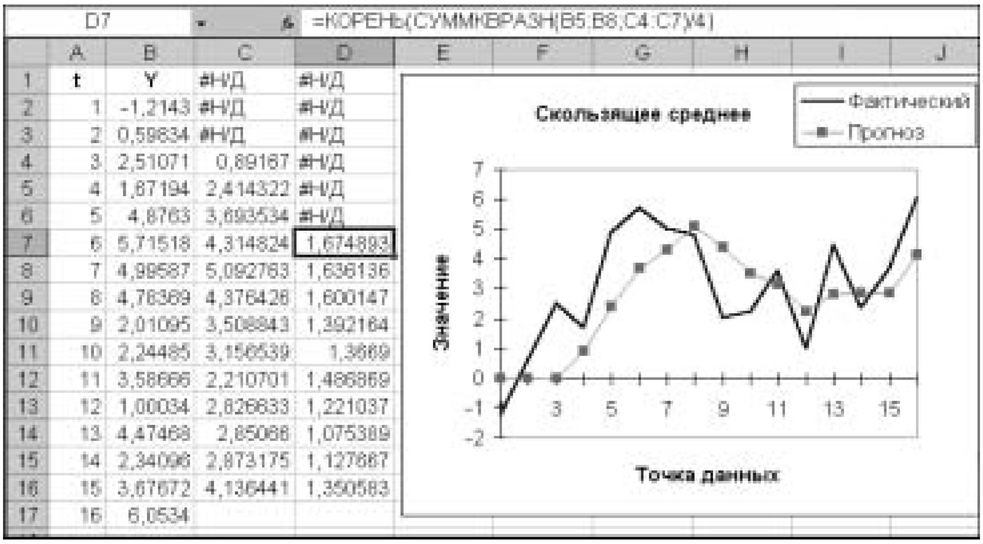


Рис. 5.47. *Исходные данные и диалоговое окно* Скользящее среднее

5.18. Экспоненциальное сглаживание

Экспоненциальное сглаживание, как и скользящее среднее (см. раздел 5.17), используется для выравнивания (сглаживания) значений временных рядов. Если имеются дискретные наблюдения y_1, y_2, ..., y_n, то сглаженные значения вычисляются по формуле $\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$, где $\hat{y}_t$ — сглаженное значение для предыдущего t, α — постоянная сглаживания, также называемая *фактором затухания* (это число из интервала (0, 1)).



	A	B	C	D
1	t	Y	#Н/Д	#Н/Д
2	1	-1,2143	#Н/Д	#Н/Д
3	2	0,59834	#Н/Д	#Н/Д
4	3	2,51071	0,89167	#Н/Д
5	4	1,67194	2,414322	#Н/Д
6	5	4,8763	3,693534	#Н/Д
7	6	5,71518	4,314824	1,674893
8	7	4,99587	5,092763	1,636136
9	8	4,78369	4,376426	1,600147
10	9	2,01095	3,508843	1,392164
11	10	2,24485	3,156539	1,3669
12	11	3,58666	2,210701	1,486869
13	12	1,00034	2,826833	1,221037
14	13	4,47468	2,65066	1,075389
15	14	2,34096	2,873175	1,127667
16	15	3,67672	4,136441	1,350583
17	16	6,0534		

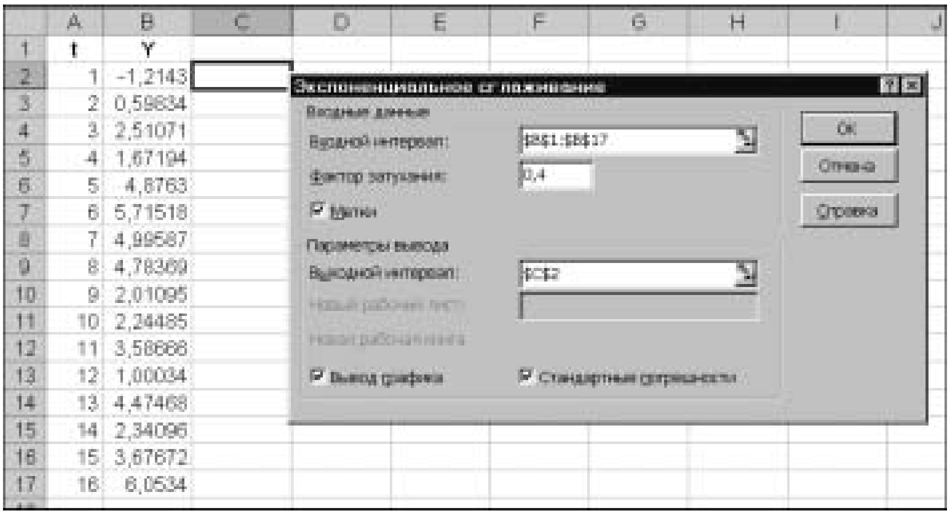
Рис. 5.48. *Результаты вычислений*

На рис. 5.49 показаны рабочий лист Excel с исходными данными (данные взяты из примера предыдущего раздела) и диалоговое окно Экспоненциальное сглаживание. В поле Входной интервал указывается адрес диапазона, содержащего значения y_i. Если этот диапазон включает заголовок, то надо установить флажок опции Метки. В поле Фактор затухания задается постоянная сглаживания; если она не задана, то по умолчанию используется значение 0,3. Установка флажков опций Вывод графика и Стандартные погрешности приводит к построению графика, на котором будут отображаться исходные и сглаженные значения (рис. 5.50), и к выводу дополнительного столбца со значениями погрешностей. Эти погрешности вычисляются как сумма квадратов разностей между тремя последовательными исходными и расчетными значениями, деленная на число 3. Формула Excel, по которой подсчитываются стандартные погрешности, показана на рис. 5.50.

5.19. Анализ Фурье

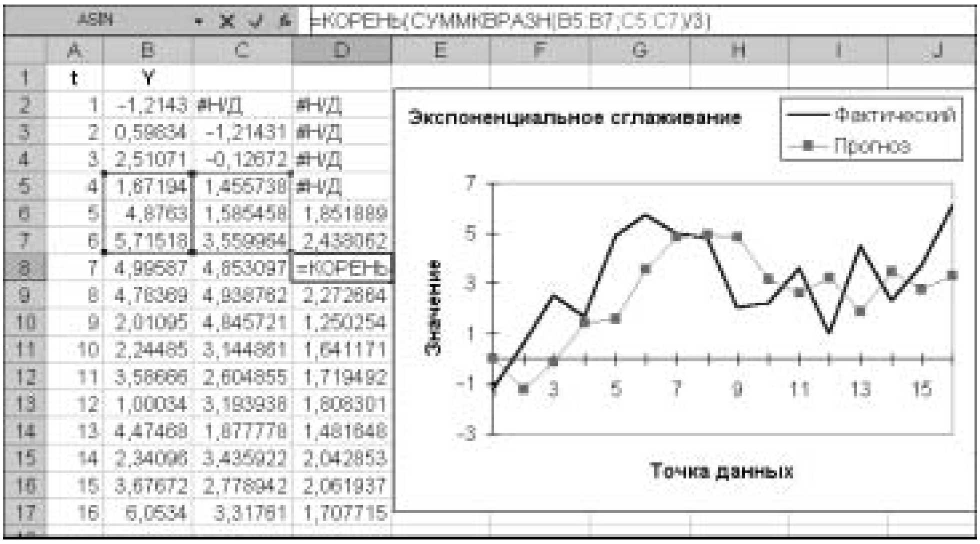
Данное средство выполняет дискретное преобразование Фурье. Это преобразование используется в анализе линейных систем и применяется к временны́м рядам для выявления периодических (спектральных) составляющих таких рядов.

Если имеются дискретные наблюдения y_1, y_2, ..., y_n, то *прямое дискретное преобразование Фурье* выполняется в соответствии с формулой $Y_k = \sum_{j=1}^{n} y_j e^{-i\frac{2\pi}{n}jk}$, $k = 0$, 1, ..., $n - 1$. Результаты преобразования Y_k являются комплексными числами, модуль которых равен амплитуде k-й спектральной составляющей (k-й гармоники), а аргумент комплексного числа Y_k равен фазе этой гармоники. Аналогично определяется *обратное дискретное преобразование Фурье* ($y_j = \sum_{k=0}^{n-1} Y_k e^{i\frac{2\pi}{n}jk}$), которое преобразует спектральное представление временного ряда в действительное.



t	Y
1	-1,2143
2	0,59834
3	2,51071
4	1,67194
5	4,8763
6	5,71518
7	4,99587
8	4,78369
9	2,01095
10	2,24485
11	3,58666
12	1,00034
13	4,47468
14	2,34096
15	3,67672
16	6,0534

Рис. 5.49. *Исходные данные и диалоговое окно* Экспоненциальное сглаживание



=КОРЕНЬ(СУММКВРАЗН(B5:B7;C5:C7)/3)

t	Y		
1	-1,2143	#Н/Д	#Н/Д
2	0,59834	-1,21431	#Н/Д
3	2,51071	-0,12672	#Н/Д
4	1,67194	1,455738	#Н/Д
5	4,8763	1,585458	1,851889
6	5,71518	3,559964	2,438062
7	4,99587	4,853097	=КОРЕНЬ
8	4,78369	4,938762	2,272664
9	2,01095	4,845721	1,250254
10	2,24485	3,144861	1,641171
11	3,58666	2,604855	1,719492
12	1,00034	3,193938	1,808301
13	4,47468	1,877778	1,481648
14	2,34096	3,435922	2,042853
15	3,67672	2,778942	2,061937
16	6,0534	3,31761	1,707715

Рис. 5.50. *Результаты вычислений*

Средство Анализ Фурье выполняет как прямое так и обратное преобразования методом *быстрого преобразования Фурье* (БПФ). Применение метода БПФ диктует условие, чтобы количество исходных значений как для прямого, так и для обратного преобразований, было равно некоторой положительной степени числа 2. Максимальное число значений, которое может обработать средство Анализ Фурье, составляет 4096 (= 2^{12}). Для применения обратного преобразования Фурье исходные значения должны быть в формате комплексных чисел $x + yi$ или $x + yj$ (i и j — обозначение мнимой единицы). Если x является отрицательным числом, перед ним ставится апостроф (').

На рис. 5.51 показаны рабочий лист с исходными данными и диалоговое окно Анализ Фурье. Результат прямого преобразования Фурье показан на рис. 5.52. Первое значение (ячейка C2) равно сумме исходных данных.

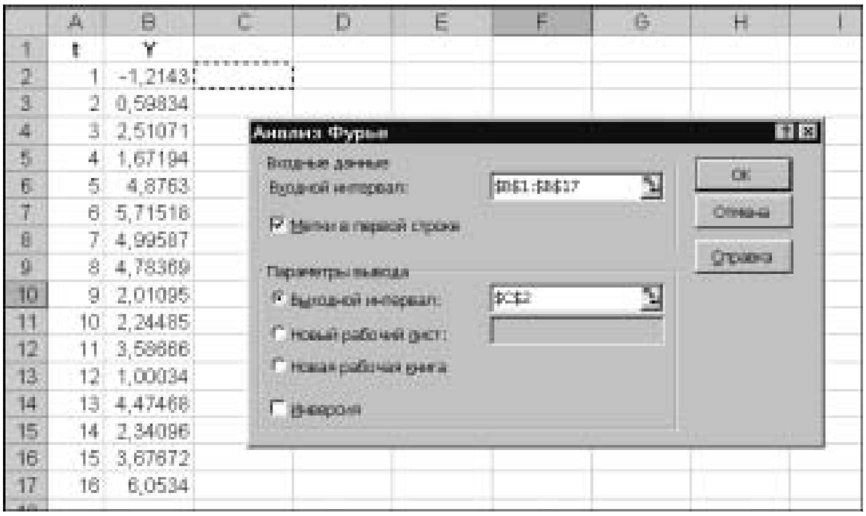


Рис. 5.51. Исходные данные и диалоговое окно Анализ Фурье

	A	B	C
1	t	Y	Y комплексное
2	1	-1,2143	49,3262968701278
3	2	0,59834	-6,30122668866927-3,1954561628609i
4	3	2,51071	-6,46705511956092+12,0347360926031i
5	4	1,67194	0,821035859520582+4,47215620909899i
6	5	4,8763	-4,6223323514806+2,61004208030321i
7	6	5,71518	-3,88437851609255+4,01286152893839i
8	7	4,99587	-10,6416229921968+6,88428999446164i
9	8	4,78369	-3,53648384452208-2,04828062788192i
10	9	2,01095	0,508894946664288
11	10	2,24485	-3,53648384452207+2,04828062788193i
12	11	3,58666	-10,6416229921968-6,88428999446163i
13	12	1,00034	-3,88437851609256-4,01286152893838i
14	13	4,47468	-4,6223323514806-2,61004208030321i
15	14	2,34096	0,82103585952057-4,472156209099i
16	15	3,67672	-6,46705511956095-12,0347360926031i
17	16	6,0534	-6,30122668866926+3,19545616286091i
18			

Рис. 5.52. Результат прямого преобразования Фурье

На рис. 5.53 показаны рабочий лист с исходными данными (результат прямого преобразования Фурье) для обратного преобразования и диалоговое окно Анализ Фурье, в котором установлен флажок опции Инверсия. Результат обратного преобразования Фурье показан на рис. 5.54; он совпадает с первоначальными данными из столбца B.

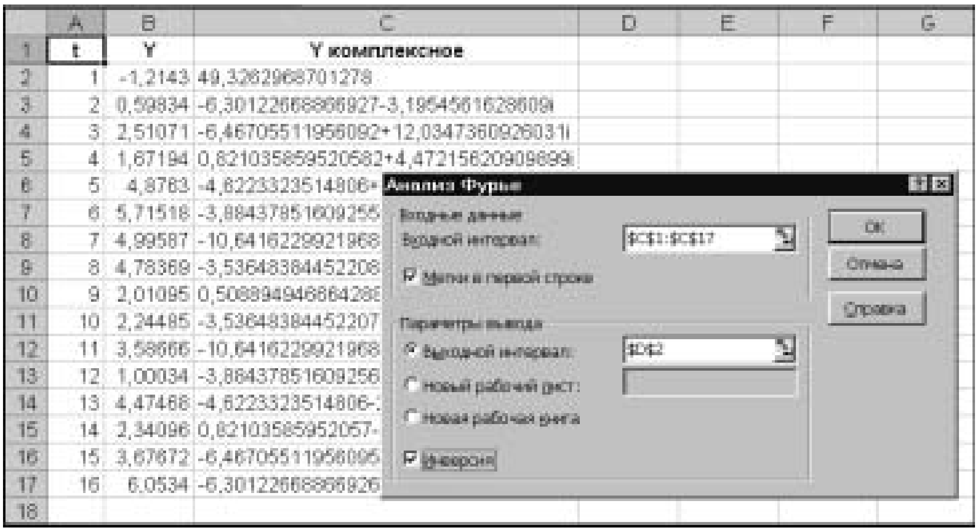
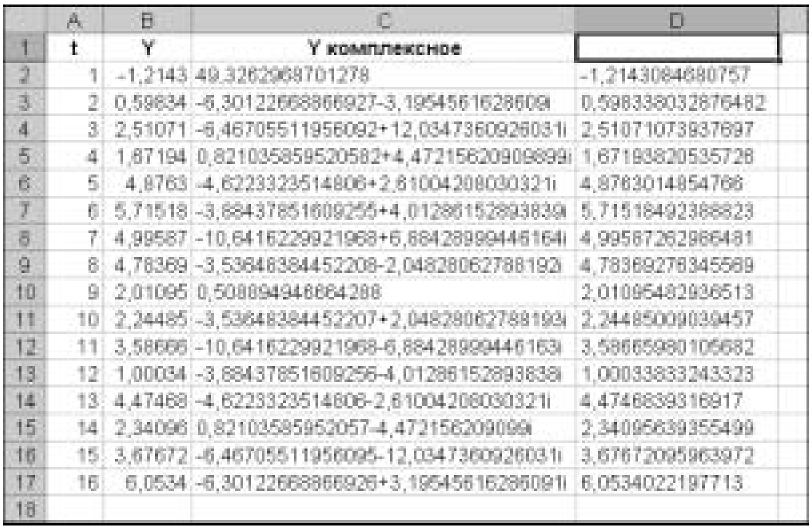


Рис. 5.53. Исходные данные для обратного преобразования Фурье и диалоговое окно Анализ Фурье



	A	B	C	D
1	t	Y	Y комплексное	
2	1	-1,2143	49,3262968701278	-1,2143084680757
3	2	0,59834	-6,30122668866927-3,1954561628609i	0,598338032876482
4	3	2,51071	-6,46705511956092+12,0347360926031i	2,51071073937697
5	4	1,67194	0,821035859520582+4,47215620909899i	1,67193820535728
6	5	4,8763	-4,6223323514806+2,61004208030321i	4,8763014854766
7	6	5,71518	-3,88437851609255+4,01286152893839i	5,71518492388823
8	7	4,99587	-10,6416229921968+6,88428999446164i	4,99587262986481
9	8	4,78369	-3,53648384452208-2,04828062788192i	4,78369276345569
10	9	2,01095	0,508894946664288	2,01095482936513
11	10	2,24485	-3,53648384452207+2,04828062788193i	2,24485009039457
12	11	3,58666	-10,6416229921968-6,88428999446163i	3,58665980105682
13	12	1,00034	-3,88437851609256-4,01286152893838i	1,00033833243323
14	13	4,47468	-4,6223323514806-2,61004208030321i	4,4746839316917
15	14	2,34096	0,82103585952057-4,472156209099i	2,34095639355499
16	15	3,67672	-6,46705511956095-12,0347360926031i	3,67672095963972
17	16	6,0534	-6,30122668866926+3,19545616286091i	6,0534022197713
18				

Рис. 5.54. Результат обратного преобразования Фурье