Министерство сельского хозяйства РФ
Федеральное государственное образовательное учреждение
высшего профессионального образования
«Мичуринский государственный аграрный университет»

Кафедра информатики

Утверждено
Протокол № 6
экономического факультета
от 14 февраля 2007

ПРАКТИКУМ

для самостоятельной работы студентов
<u>Часть 2</u>

Excel в математических *и* статистических расчетах

Мичуринск-наукоград 2007

Методическое пособие разработано старшим преподавателем кафедры информатики **М.А. Ильченко**, старшим лаборантом **Л.С. Струковой** на основании государственных образовательных стандартов высшего профессионального образования. Методическое пособие рекомендовано студентам, изучающим математику, преподавателям, а также аспирантам, сталкивающимся с необходимостью математической обработки данных.

Рецензент:

Г.Б. Ширяева ст. преподаватель кафедры математического моделирования экономических систем

Рассмотрено на заседании кафедры
Протокол № 63 от 14 декабря 2006г.

Содержание

4. Элементы математического анализа

К математическому анализу относят совокупность разделов математики, посвященных исследованию функций методами дифференциального и интегрального исчислений.

Производная

Производной функции $y = f(x)$ называется предел отношения приращения функции к приращению независимой переменной при стремлении последнего к нулю (если этот предел существует):

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x - \Delta x) - f(x)}{\Delta x}$$

Геометрический смысл производной заключается в том, что производная $f'(x_0)$ есть угловой коэффициент (тангенс угла наклона a) касательной, проведенной к кривой $y = f(x)$ в точке x_0, то есть $k = f'(x_0)$ (рис. 4.1).

Обычно производная характеризует скорость изменения различных функций. Например, скорость движения — это производная от пути по времени s'_t. Экономическим приложением производной, например, может быть нахождение производительности труда в момент t_0 по функции количества произведенной продукции $u=u(t)$.

С ее помощью находятся такие показатели, как предельные затраты, предельная выручка, предельный спрос, предельная производительность и др. *Внешним признаком наличия связи с производной в экономических приложениях является присутствие термина «предельный».*

Предельные затраты — это производная от затрат по выпуску продукции.

Предельная производительность — это производная от выпуска продукции по затратам данного ресурса.

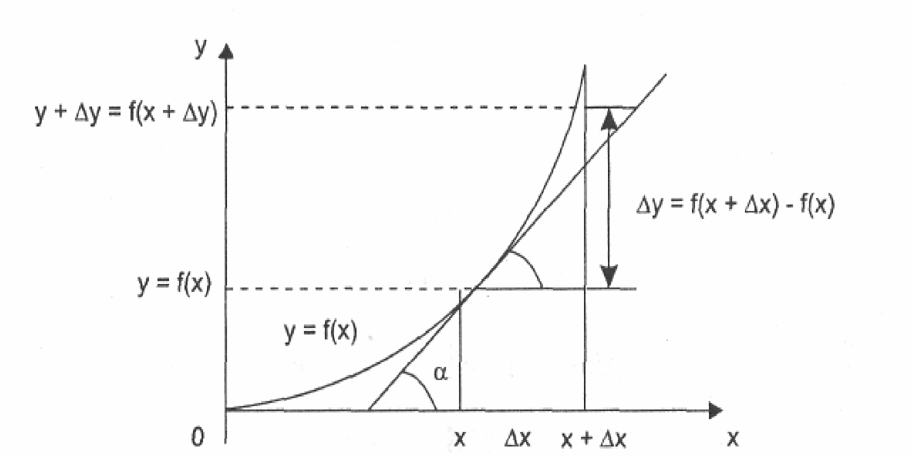


Рис. 4.1. Геометрический смысл производной

<u>*Предельный спрос*</u> — это производная от спроса по цене.

Поскольку производная *f'(x)* сама является функцией, то она также может иметь производную. Продолжая дальше, можно сказать, что производной *n-го* порядка называется производная от производной *(n - 1)-го* порядка. Обозначаются производные высших порядков как: $f''(x)$ - второго порядка (или вторая производная), $f'''(x)$ — третьего порядка (или третья производная). Для обозначения производных более высокого порядка используются арабские цифры в скобках или римские цифры $(f^{IV}(x))$.

Продифференцировать функцию (найти ее производную) можно, используя таблицу производных простейших функций и правила дифференцирования:

Таблица производных простейших функций

$f(x)$	$c = \text{const}$	x	x^n	$\sqrt{x}$	$\sin x$	$\cos x$
$f'(x)$	0	1	$n * x^{n-1}$	$\frac{1}{2\sqrt{x}}$	$\cos x$	$-\sin x$
$f(x)$	$\mathrm{tg} x$	a^x	e^x	$\log_a x$	$\ln x$	
$f'(x)$	$\frac{1}{\cos^2 x}$	$a^x * \ln a$	e^x	$\frac{1}{x * \ln a}$	$\frac{1}{x}$	

Правила дифференцирования:

- производная суммы функций:

$$(f(x) + g(x) + \ldots + h(x))' = f'(x) + g'(x) + \ldots + h'(x)$$

- производная суммы функций равна сумме производных

$$(k * f(x))' = k * f'(x)$$

- постоянный коэффициент можно выносить за знак производной.

- производная произведения:

$$(f(x)*g(x))' = f'(x)*g(x)+F(x)*g'(x)$$

- производная дроби (частного, отношения):

$$\left(\frac{u(x)}{v(x)}\right)' = \frac{u'(x)*v(x)-u(x)*v'(x)}{v^2(x)}$$

- производная сложной функции:

$$(f(g(x)))' = f'(g(x))*g'(x)$$

Однако в ряде случаев простейшими правилами дифференцирования воспользоваться достаточно сложно. Например, когда функция задана таблично или получена в результате наблюдений. В этих случаях прибегают к численным методам дифференцирования.

Численное дифференцирование очень чувствительно к ошибкам, вызванным неточностью исходных данных, отбрасываемых членов ряда и т. д., и поэтому должно применяться с осторожностью.

Существуют различные формулы численного дифференцирования. Из них простейшими являются явные трехточечные формулы, в частности:

$$y_0' \approx \frac{1}{2\Delta x}(-y_{-1} + y_1) \tag{4.1.}$$

Здесь y_{-1}, y_0 и y_1, — три последовательные точки, Δx — достаточно малый шаг дискретизации. Погрешность вычисления производной Δ по формуле (4.1) может быть оценена из выражения:

$$\Delta = \frac{\Delta x^2}{6} y''(x)$$

где $x_{-1} < x < x_1$

Когда экспериментальные данные зашумлены более сильно (имеют большие погрешности) используют методы дифференцирования по большему числу точек. Примером может являться так называемая формула численного дифференцирования после сглаживания:

$$y_k' \approx \frac{1}{10\Delta x}(-2y_{k-2} - y_{k-1} + y_{k+1} + 2y_{k+2})$$

Рассмотрим примеры численного нахождения производной.

Пример. Функция затрат, определенная экспериментально, имеет вид:

Объем	Затраты
1	2,693147
1,1	2,941937
1,2	3,188457
1,3	3,432909
1,4	3,675469
1,5	3,916291
1,6	4,155511
1,7	4,393252
1,8	4,629619
1,9	4,864711
2,0	5,098612
2,1	5331402
2,2	5,563151
2,3	5,793922

Найти предельные издержки производства при объеме выпуска $x_1 = 2$.

Решение. Для решения задачи необходимо найти производную затрат по объему производства в точке *x = 2.* Для этого:

1. Вводим в рабочий лист Excel в диапазон A1:B3 значения точек с объемом выпуска *1,9-2,1:*

A	B
1,9	4,864711
2	5,098612
2,1	5,331402

2. В ячейку B4 вводим формулу дифференцирования (4.1): *=(B3 - B1)/0,2.* После нажатия клавиши Enter получаем значение производной в точке *x = 2 (2,33455).*

Отметим, что истинное значение производной в данном случае равно $2\frac{1}{3}$.

Пример. Найти производную функции $y = \sin x$ на промежутке $x \in [0;\ 6{,}2]$ при шаге дискретизации $\Delta x = 0{,}2$. Построить график функции и ее производной.

Решение.

1. Для решения задачи, прежде всего, необходимо ввести данные в рабочую таблицу. Вводим в ячейку A1 слово *аргумент*. Затем в ячейку A2 — первое значение аргумента — *0* (левую границу диапазона). Далее в ячейку A3 введем второе значение: левая граница плюс шаг дискретизации — *0,2*. Теперь необходимо скопировать формулу в ячейки A4:A33 (выделяем блок ячеек A2:A3 и за правый нижний угол протягиваем до ячейки A33, там появляется значение 6,2). Значения аргумента введены.

2. Далее требуется вводить значения функции (в примере синуса). В ячейку B1 заносим слово *синус* и устанавливаем табличный курсор в ячейку B2. Здесь должно оказаться значение синуса, соответствующее значению аргумента в ячейке A2. Для получения значения синуса воспользуемся специальной функцией: нажимаем на панели инструментов кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций - шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Математические. Справа в поле Функция выбираем SIN. Нажимаем кнопку ОК. Появляется диалоговое окно SIN. Наведя указатель мыши на серое поле окна, при нажатой левой кнопке сдвигаем поле вправо, чтобы открыть столбец данных (A). Указываем значение аргумента синуса щелчком мыши на ячейке A2. Нажимаем кнопку ОК. В ячейке B2 появляется *0*. Теперь необходимо скопировать функцию из ячейки B2 в ячейки B3:B33. Для этого устанавливаем табличный курсор в ячейку B2, и за правый нижний угол протягиванием копируем в ячейки B3:B33. Значения синуса получены.

3. Теперь по введенным в рабочую таблицу данным необходимо найти значения производной. Для этого в ячейку C1 вводим слово *производная*. В ячейку C3 вводим формулу дифференцирования (4.1): *=(B4 - B2)/(2*0,2)*. Протягиванием (за правый нижний угол) копируем ее из ячейки C3 в ячейки C4:C32 (в ячейках C2 и C33 значения производной не определены, так как не заданы значения синуса в ячейках B1 и B34). Получены значения производной.

4. Далее по полученным данным строим диаграмму. Щелчком указателя мыши на кнопке на панели инструментов вызываем Мастер диаграмм. В появившемся диалоговом окне выбираем

тип диаграммы — График, вид — левый верхний. После нажатия кнопки Далее указываем диапазон данных — *B 1:C33* (с помощью мыши). Проверяем положение переключателя Ряды в: столбцах. Выбираем вкладку Ряд и с помощью мыши вводим диапазон подписей оси *X: A2:A33*. Нажав кнопку Далее, вводим названия осей *X* и У: *Аргумент* и *Значения*, соответственно. Нажимаем кнопку Готово. Появляется диаграмма, изображенная на рис. 4.2.

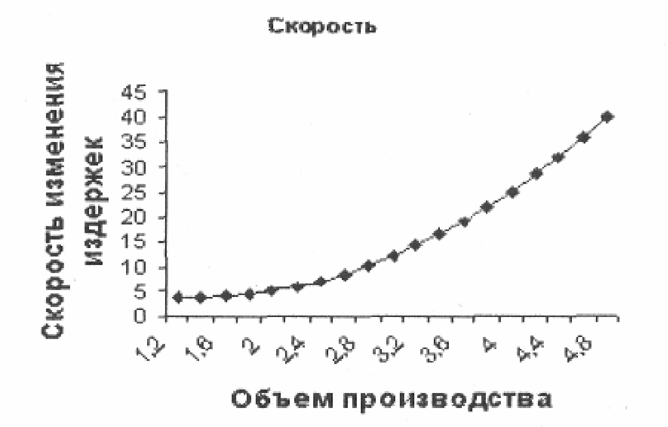


Рис. 4.2. Диаграмма синуса и его производной из примера

Как нетрудно заметить, график производной на рис. 4.2 совпадает с графиком косинуса, который и является точным значением производной от функции синуса.

УПРАЖНЕНИЯ.

1. Зависимость спроса на товар от цены выражается формулой:

$$d(p) = \frac{100}{p+1}$$

Построить график функции этой зависимости в диапазоне $p \in [1\ ;\ 3]$ с шагом $\Delta p = 0{,}1$. С какой скоростью изменяется спрос при цене $p = 2$?

2. Зависимость полных издержек производства k от объема производства x выражается формулой:

$$k = x^3 - 4x^2 + 9x.$$

Изобразите графически изменение издержек с изменением объема производства в диапазоне $x \in [1;\ 5]$ с шагом $\Delta x = 0{,}2$. Постройте график скорости изменения издержек в диапазоне же $x \in [1{,}2;\ 4{,}8]$ с шагом $\Delta x = 0{,}2$.

Определенный интеграл

Пусть на отрезке $[a, b]$ задана функция $y = f(x)$, и отрезок $[a,b]$ разбит на n элементарных отрезков точками $a = x_0 < x_1 < x_2 < \ldots < x_n = b$:

На каждом отрезке $[x_{i-1}, x_i]$ разбиения выбрана некоторая точка x_i и положено, что $\Delta x_i = x_i - x_{i-1}$ где $i = 1, 2, \ldots, n$. Тогда сумма вида:

$$\sum_{i=1}^{n} f(x_i)\Delta x_i \tag{4.2}$$

называется интегральной суммой для функции $y = f(x)$ на $[a,b]$. Интегральная сумма зависит как от способа разбиения отрезка $[a, b]$ точками $x_0, x_1, \ldots, x_n$, так и от выбора точек $x_1, x_2, \ldots, x_N$ на каждом из отрезков разбиения $\Delta x_i = x_i - x_{i-1}, i-1,2,\ldots,n$. Обозначим через $\max \Delta x_i$; максимальную из длин отрезков $[x_{i-1}, x_i]$, где $i=1,2,\ldots n$. Тогда определенным интегралом от функции $y = f(x)$ на $[a, b]$ называется предел интегральной суммы при стремлении $\max \Delta x_i$ к нулю, если он существует, конечен и не зависит от способа выбора точек и точек $x_1, x_2, \ldots$ и точек $x_1, x_2, \ldots$. Определенный интеграл обозначается

$$\int_a^b f(x)dx$$

а функция $f(x)$ называется интегрируемой на отрезке $[a, b]$, то есть:

$$\int f(x)dx = \lim_{\max \Delta x_1 \to 0} \sum_{i=1}^{n} f(x_i)\Delta x_i \tag{4.3}$$

При этом число a называется нижним пределом определенного интеграла, число b — его верхним пределом, функция $f(x)$ — подынтегральной функцией, выражение

$f(x)dx$ — подынтегральным выражением, а задача о нахождении $\int_a^b f(x)dx$ -интегрированием функции $f(x)$ на отрезке $[a, b]$.

Геометрический смысл определенного интеграла заключается в следующем.

Если функция $y = f(x)$ неотрицательна на отрезке $[a, b]$, где $a < b$, то $\int_a^b f(x)dx$ численно равен площади S под кривой $y = f(x)$ на $[a, b]$ (рис. 4.3). Действительно, отдельное слагаемое интегральной суммы (4.2) равно площади s_i прямоугольника со сторонами Δx_i и $f(x_i)$ (согласно определению значение определенного интеграла не зависит от способа выбора точек $x_1, x_2, \ldots$) где $i = 1, 2, \ldots, n$ (рис. 4.3). Поэтому вся интегральная сумма (4.2) равна площади $S_л = S_1 + S_2 + \ldots + S_n$ под ломаной, образованной на каждом из отрезков $[x_{i-1}, x_i]$ прямыми, параллельными оси абсцисс. При стремлении max Δx_i к нулю ломаная неограниченно приближается к исходной кривой и площадь под ломаной переходит в площадь под кривой $S_л = S$.

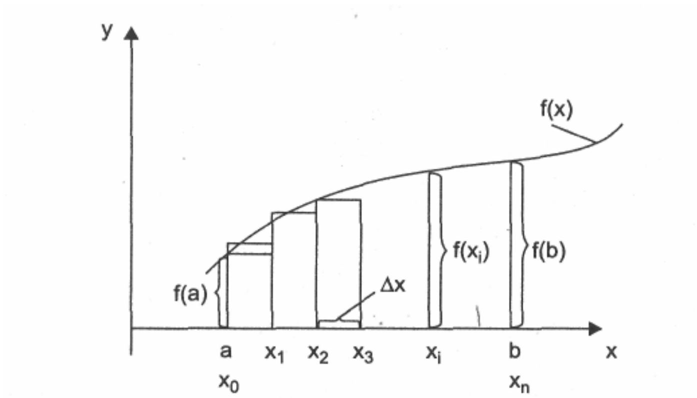


Рис. 4.3. Диаграмма, поясняющая геометрический смысл определенного интеграла

В экономических приложениях, например, определенный интеграл может выражать объем произведенной продукции *(u)* при известной функции производительности труда (*f(t)*):

$$u = \int_0^T f(t)dt$$

Обычно для нахождения определенного интеграла используется формула Ньютона-Лейбница:

$$\int_a^b f(x)dx = F(b) - F(a) \qquad (4.4)$$

где *F(a)* и *F(b)* первообразные для *f(x)* в точках *a* и *b*. Первообразной функцией для функции *f(x)* на промежутке X называется функция *F(x)*, если в каждой точке *x* этого промежутка *F'(x) = f(x)*.

Однако применение формулы (4.4) на практике связано с существенными трудностями, возникающими при нахождении первообразной в случае усложнения подынтегральной функции. Поэтому в приложениях используют так называемые *численные методы, позволяющие найти приближенное значение искомого интеграла с требуемой точностью*. Этот подход оказывается особенно предпочтительным при использовании компьютеров для нахождения интегралов.

Существует значительное количество численных методов вычисления интегралов. Они основаны на разных способах нахождения площади под кривой *f(x)* (рис. 4.3.):

- *как суммы элементарных прямоугольников — метод прямоугольников:*

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} f(x_i)\Delta x \qquad (4.5)$$

- *как суммы элементарных трапеций — метод трапеций:*

$$\int_a^b f(x)dx \approx (\sum_{i-1}^{n-1} f(x_i) + \frac{f(x_0) + f(x_n)}{2})\Delta x \qquad (4.6)$$

Существуют также метод Симпсона и ряд других.

Формула метода прямоугольников (4.5) получается, если отрезок интегрирования [*a*, *b*] разбить на *n* равных частей длиной

$$\Delta x = \frac{b-a}{n}$$

На каждом из отрезков разбиения $[x_{i-1}, x_i]$ участок кривой *y* - *f(x)* заменяется отрезком прямой, параллельным оси абсцисс. Тогда:

$$\int_a^b f(x)dx \approx S_1 + S_2 + \ldots + S_n$$

где $S_1, S_2, \ldots, S_n$ — площади прямоугольников на каждом из отрезков разбиения. Отдельное слагаемое S_i равно площади прямоугольника со сторонами Δx и $f(x_i)$, где i = 1, 2, ..., n (рис. 4.3). Отсюда получается формула (4.5). Метод прямоугольников является простейшим, но и наименее точным.

Более точно определенный интеграл может быть вычислен по формуле трапеций (4.6). В этом случае, в отличие от метода прямоугольников, на каждом из отрезков разбиения [x_{i-1}, x_i] участок кривой *y=f(x)* заменяется хордами, стягивающими концевые точки. Тогда, отдельное слагаемое интегральной суммы S_i,- равно площади трапеции с основаниями *f(xi)* и *f(*$_{xi}$*-1)* и высотой Δx, где i= 1, 2, ..., n. То есть:

$$S_i = \frac{f(x_{i-1}) + f(x_i)}{2} \Delta x$$

Складывая площади элементарных трапеций и приводя подобные члены, получаем формулу (4.6).

Погрешность (Δ) вычисления определенного интеграла по формуле трапеций (S(n)):

$$\Delta = \left| \int_a^b f(x)dx - S(n) \right|$$

может быть оценена из выражения:

$$\Delta \le \frac{(b-a)^3}{12n^2} M_2$$

где M_2 — максимальное значение модуля второй производной *f''(x)* подынтегральной функции *y -f(x)* на *[a, Ь]*. Рассмотрим вычисления интегралов по методу прямоугольников и методу трапеций.

Пример. Методом прямоугольников и методом трапеций найти $\int_0^3 x^2dx$ с шагом $\Delta x = 0{,}1$. Заметим, что этот интеграл легко может быть вычислен аналитически:

$$\int_0^3 x^2dx = \frac{x^3}{3}\Big|_0^3 = 9 - 0 = 9$$

Решение 1. Для нахождения определенного интеграла методом прямоугольников необходимо ввести значения подынтегральной функции *f(x)* в рабочую таблицу Excel в диапазоне $x \in [0; 3]$ с заданным шагом $\Delta X = 0{,}1$.

1. Открываем чистый рабочий лист (команда Вставка > Лист).
2. Составляем таблицу данных (*x* и $f(x)$). Пусть первый столбец будет значениями *x, а* второй соответствующими показателями *f(x).* Для этого в ячейку A1 вводим слово *Аргумент,* а в ячейку B1 — слово *Функция.* В ячейку A2 вводится первое значение аргумента — левая граница диапазона *(0).* В ячейку A3 вводится второе значение аргумента — левая граница диапазона плюс шаг построения *(0,1).* Затем, выделив блок ячеек A2:A3, автозаполнением получаем все значения аргумента (за правый нижний угол блока протягиваем до ячейки A32, до значения *x* = 3).
3. Далее вводим значения подынтегральной функции. В ячейку B2 необходимо записать ее уравнение. Для этого табличный курсор необходимо установить в ячейку B 2 и с клавиатуры ввести формулу *=A2^2* (при английской раскладке клавиатуры). Нажимаем клавишу Enter. В ячейке B2 появляется *0.* Теперь необходимо скопировать функцию из ячейки B2. Автозаполнением копируем эту формулу в диапазон B2:B32.

В результате должна быть получена таблица данных для нахождения интеграла.

4. Теперь в ячейке B33 может быть найдено приближенное значение интеграла. Для этого в ячейку B33 вводим формулу *=0,1 *,* затем вызываем Мастер функций (нажатием на панели

инструментов кнопки Вставка функции (f_x)). В появившемся диалоговом окне Мастер функций - шаг 1 из 2 слева в поле Категория выбираем Математические. Справа в поле Функция — функцию Сумм. Нажимаем кнопку ОК. Появляется диалоговое окно Сумм. В рабочее поле мышью вводим диапазон суммирования *B3:B32* (рис. 4.4).

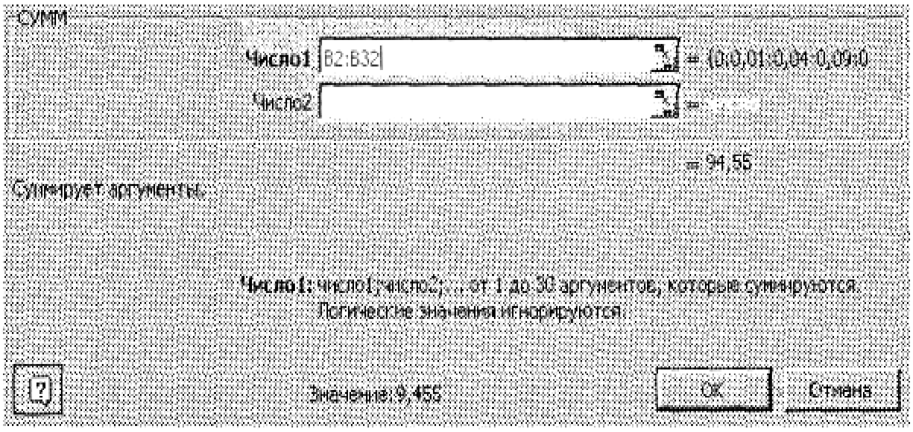


Рис. 4.4. Пример заполнения диалогового окна СУММ

Нажимаем кнопку ОК. В ячейке В33 появляется приближенное значение искомого интеграла *(9,455).*

Сравнивая полученное приближение с истинным значением интеграла (9), можно видеть, что ошибка приближения метода прямоугольников в данном случае довольно значительна — *0,455.*

Решение 2. Для нахождения определенного интеграла методом трапеций, как и в случае использования метода прямоугольников, значения подынтегральной функции $f(x)$ должны быть введены в рабочую таблицу Excel в диапазоне $x \in$ [0; 3] с заданным шагом Δx=0,1. Поэтому этапы 1-3 полностью аналогичны этапам предыдущего решения. Поскольку таблица данных для нахождения интеграла уже введена, приступаем сразу к этапу 4:

Теперь в ячейке В34 может быть найдено приближенное значение интеграла по методу трапеций. Для этого в ячейку В34 вводим формулу = *0,1*((B2 + B32)/2 +,* затем вызываем Мастер функций (нажатием на панели инструментов кнопки Вставка функции (f_x)). В появившемся диалоговом окне Мастер функций - шаг 1 из 2 слева в поле Категория выбираем Математические.

Справа в поле Функция — функцию Сумм. Нажимаем кнопку ОК. Появляется диалоговое окно Сумм. В рабочее поле мышью вводим диапазон суммирования *B3:B31*. Нажимаем кнопку ОК и еще раз ОК. В ячейке В34 появляется приближенное значение искомого интеграла *(9,005)*.

Сравнивая полученное приближение с истинным значением интеграла *(9)*, можно увидеть, что ошибка приближения метода трапеций в данном случае вполне приемлемая — *0,005*.

Пример. Методом трапеций найти $\int_0^{3,1} Sinxdx$ с шагом $\Delta x = 0,1$.

(Точное значение $\int_0^{p} Sinxdx = 2$.)

Решение. Для нахождения определенного интеграла методом трапеций значения подынтегральной функции f(x) должны быть введены в рабочую таблицу Excel в диапазоне xI[0;3,1] с заданным шагом $\Delta x = 0,1$.

1. Открываем чистый рабочий лист (команда Вставка > Лист).

2. Составляем таблицу данных *(x* и *f(x))*. Пусть первый столбец будет значениями *x*, а второй соответствующими показателями *f(x)*. Для этого в ячейку А1 вводим слово *Аргумент*, а в ячейку В1 — слово *Функция*. В ячейку А2 вводится первое значение аргумента — левая граница диапазона *(0)*. В ячейку А3 вводится второе значение аргумента — левая граница диапазона плюс шаг построения *(0,1)*. Затем, выделив блок ячеек А2:А3, автозаполнением получаем все значения аргумента (за правый нижний угол блока протягиваем до ячейки А33, до значения *x* = 3,1).

3. Далее вводим значения подынтегральной функции. В ячейку В 2 необходимо записать ее уравнение (в примере синуса). Для этого устанавливаем табличный курсор в ячейку В 2. Здесь должно оказаться значение синуса, соответствующее значению аргумента в ячейке А2. Для получения значения синуса воспользуемся специальной функцией: нажимаем на панели инструментов кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций - шаг 1 из 2 слева в поле Категория

указаны виды функций. Выбираем Математические. Справа в поле Функция — функцию SIN. Нажимаем кнопку ОК. Появляется диалоговое окно SIN. Наведя указатель мыши на серое поле окна, при нажатой левой кнопке сдвигаем поле вправо, чтобы открыть столбец данных (А). Указываем значение аргумента синуса щелчком мыши на ячейке А2. Нажимаем кнопку ОК. В ячейке В 2 появляется *0.* Теперь необходимо скопировать функцию из ячейки В2. Автозаполнением копируем эту формулу в диапазон В3:В33. В результате должна быть получена таблица данных для нахождения интеграла. 4. Теперь в ячейке В34 может быть найдено приближенное значение интеграла по методу трапеций. Для этого в ячейку В34 вводим формулу = *0,1*((В2 + В33)/2 +,* затем вызываем Мастер функций (нажатием на панели инструментов кнопки Вставка функции (f_x)). В появившемся диалоговом окне Мастер функций-шаг 1 из 2 слева в поле Категория выбираем Математические. Справа в поле Функция — функцию Сумм. Нажимаем кнопку ОК. Появляется диалоговое окно Сумм. В рабочее поле мышью вводим диапазон суммирования *В3:В32.* Нажимаем кнопку ОК и еще раз ОК. В ячейке В34 появляется приближенное значение искомого интеграла (*1,997).* Сравнивая полученное приближение с истинным значением интеграла (2), можно видеть, что ошибка приближения метода трапеций в данном случае вполне приемлемая для практики — Δ = 0,003.

УПРАЖНЕНИЯ.

1. Методом прямоугольников и методом трапеций найти следующие интегралы:

1) $\int_0^2 x dx$ при $\Delta x = 0.1$

2) $\int_1^{1.5} \frac{dx}{x}$ при $\Delta x = 0.1$

2. Найти объем продукции, произведенной за 4 года, если функция производительности труда имеет вид $f(x) = (1 + t)e^{3t}$, методом трапеций с шагом $\Delta t = 0,2$.

Ряды

При решении прикладных математических задач приходится рассматривать суммы, составленные из большого количества слагаемых. Задача суммирования бесконечного множества слагаемых решается в теории рядов.

Числовые последовательности

Числовой последовательностью называется бесконечное множество чисел u_1, u_2, . . ., u_n, расположенных в определенном порядке одно за другим. Числа, входящие в последовательность, называются ее *членами*. Последовательность считается заданной, если известен закон ее образования, то есть правило, по которому можно определить любой член последовательности. *Примерами последовательностей являются натуральный ряд, арифметическая прогрессия, геометрическая прогрессия и* т. д.

Пределом последовательности u_1, u_2, ...,u_n называется число U, *к* которому числа u_n подходят сколь угодно близко, если такое число U существует для данной последовательности. Предел последовательности обозначается как

$$U = \lim_{n\to\infty} u_n$$

Широко распространенными числовыми последовательностями являются арифметическая и геометрическая прогрессии.

Арифметической прогрессией называется такая числовая последовательность, в которой каждый последующий член получается из предыдущего прибавлением определенного числа a, называемого разностью прогрессии. Если a > О, прогрессия называется возрастающей, если a < О — убывающей.

$$u_n = u_1 + (n-1)a \quad \text{и} \quad S_n = \frac{n(u_1+u_n)}{2} \tag{4.7}$$

Здесь s_n — сумма первых n членов прогрессии.

Геометрической прогрессией называется такая числовая последовательность, в которой каждый последующий член получается из предыдущего умножением его на определенное число q, называемое знаменателем (основанием) прогрессии. Если q > 1, прогрессия называется возрастающей, если \q\ < 1, то убывающей.

$$u_n = u_1 q^{n-1} \quad \text{и} \quad S_n = \frac{u_1(q^n - 1)}{q - 1} \qquad (4.8)$$

Здесь s_n — сумма первых *n* членов прогрессии.

В MS Excel члены последовательностей и их суммы удобнее вычислять не по формулам (4.7) и (4.8), а напрямую, поочередно член за членом. Кроме того, для нахождения членов арифметической или геометрической прогрессии существует специальная процедура Прогрессия. Для ее реализации необходимо:

1. Ввести значение первого элемента прогрессий в выбранную ячейку.
2. Выделить блок ячеек под требуемое количество членов прогрессии (либо в дальнейшем потребуется указать значение последнего элемента).
3. Выполнить команду меню Правка > Заполнить > Прогрессия.
4. В появившемся диалоговом окне Прогрессия указать тип и параметры формируемой последовательности значений (рис. 4.5).

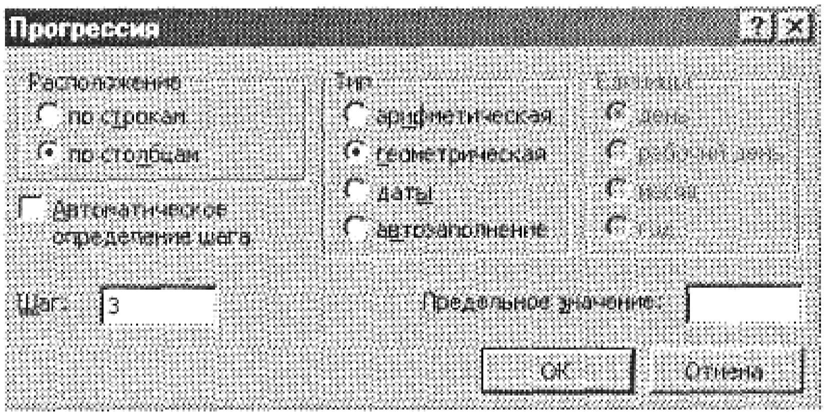


Рис. 4.5. Пример задания параметров прогрессии

Пример. Найти сумму первых 10 членов арифметической прогрессии с u_1 = *3* и разностью прогрессии *a* = *2*.

Решение.

1. Находим члены прогрессии. Для этого в ячейку A1 вводим значение первого члена — *3*. В ячейку A2 вводим значение второго члена прогрессии — *5 (3 + 2)*. Выделяем блок A1:A2 и протягиванием за правый нижний угол до ячейки A10 — автозаполнением находим остальные члены арифметической прогрессии.

2. Вычисляем частичную сумму. Устанавливаем табличный курсор в ячейку АН. Нажимаем кнопку Автосумма на панели инструментов Стандартная. Указателем мыши вводим диапазон суммирования A1:A10. Нажимаем клавишу Enter.

3. В результате в ячейке A11 должна оказаться сумма первых 10 членов арифметической прогрессии — *120.*

Пример. Построить первые 11 членов геометрической прогрессии с первым членом u_1 = 4 и со знаменателем q = 3.

Решение.

1. В ячейку A1 введем значение первого члена — *4.*
2. Выделим блок ячеек A1:A11.
3. Выполним команду меню Правка > Заполнить > Прогрессия.

	A
1	4
2	12
3	36
4	108
5	324
6	972
7	2916
8	8748
9	26244
10	78732
11	236196

Рис. 4.6. Диапазон ячеек, заполненный членами геометрической прогрессии из примера

4. Заполним поля диалогового окна Прогрессия, переключатель Расположение поставим в положение по столбцам, переключатель Тип — в положение геометрическая, в поле Шаг с клавиатуры вводим значение знаменателя — *3* (рис. 4.5). Щелкаем на кнопке ОК.

5. В результате получаем диапазон ячеек, заполненный членами геометрической прогрессии (рис. 4.6).

УПРАЖНЕНИЯ.

1. Вычислить все члены арифметической прогрессии с u_1 = 4 и разностью прогрессии a = 3, не превосходящие 17.

2. Вычислить сумму 10 членов геометрической прогрессии с первым членом u_1=3 и со знаменателем q = 4.

3. Вычислить все члены геометрической прогрессии с u_1 = 2 и знаменателем прогрессии q - 2,5, не превосходящие 500.

Числовые ряды

Числовым рядом называется бесконечная последовательность чисел $u_1, u_2, \ldots, u_n, \ldots$, соединенных знаком сложения:

$$u_1 + u_2 + \ldots u_n + \ldots = \sum_{n=1}^{\infty} u_n$$

Числа $u_1, u_2, \ldots, u_n, \ldots$ называются членами ряда, а член u_n — общим или n-м членом ряда.

Ряд считается заданным, если известен его общий член $u_n = f(n)$ (n = 1, 2, ...), то есть задана функция $f(n)$ натурального аргумента.

Сумма n первых членов ряда s_n называется n-й частичной суммой ряда.

Ряд называется сходящимся, если существует конечный предел последовательности его частичных сумм, то есть:

$$\lim_{n \to \infty} S_n = S$$

Число S называется суммой ряда. Если конечного предела последовательности частичных сумм не существует, то ряд называется расходящимся. В математике существуют специальные приемы нахождения частичных сумм ряда. Применение компьютера позволяет вычислять частичные суммы напрямую. В MS Excel обычно вначале вычисляются n первых членов соответствующей числовой последовательности. Для этого вводится требуемое количество значений натурального аргумента, затем формула общего члена ряда копируется в n ячеек, после чего находятся требуемые суммы.

Пример. Необходимо вычислить сумму 12 первых членов гармонического ряда:

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots$$

Решение.

1. В ячейку A1 вводим слово *Аргумент*, в ячейку B1 — *Ряд*.

2. В диапазон A2:A13 вводим 12 значений аргумента: в ячейку A2 — число *1*, в ячейку A3 — второе значение аргумента

2, выделяем блок A2:A3 и протягиванием за правый нижний угол блока заполняем диапазон A2:A13 значениями аргумента.

3. В ячейку B2 вводим формулу общего члена ряда: *=1/A2* (при английской раскладке клавиатуры). Нажимаем клавишу Enter.

4. Протягиванием (за правый нижний угол) копируем формулу из ячейки B2 в диапазон B3:B13.

5. Проводим суммирование. Для этого, установив табличный курсор в ячейке B 14, на панели инструментов Стандартная нажимаем кнопку Автосумма и мышью указываем диапазон суммирования (B2:B13). Нажимаем клавишу Enter.

6. В ячейке B14 получаем сумму 12 первых членов гармонического ряда — *3,103211*.

УПРАЖНЕНИЯ.

1. Найти сумму первых 7 членов ряда:

$$\frac{1}{1*2}+\frac{1}{2*3}+\ldots+\frac{1}{n(n+1)}+\ldots$$

2. Найти сумму первых 10 членов ряда:

$$\sum_{n=1}^{10}\frac{n+1}{\sqrt{n^3}}$$

Функциональные ряды

В отличие от числовых рядов членами функционального ряда являются функции. Поэтому ряд, составленный из функций одной и той же переменной *x:*

$$f_1(x)+f_2(x)+\ldots+f_n(x)+\ldots=\sum_{n=1}^{\infty}f_n(x)$$

называется функциональным. Примером функционального ряда может служить степенной ряд:

$$c_0+c_1x+c_2x^2+\ldots+c_nx^n+\ldots$$

Здесь числа c_0, $c_1,\ldots,c_n$ — коэффициенты степенного ряда.

В MS Excel для функциональных рядов, имеющих прикладное значение (в основном, финансово-экономическое), существуют специальные функции, вычисляющие частичные суммы, заданные члены и другие параметры функциональных рядов.

Например, в широко известной задаче о сложных процентах при вкладе в банк u_0 денежных единиц с ежегодной выплатой x процентов годовых, функциональный ряд годовых приростов вклада будет выглядеть как

$$u_0 + u_0 x + u_0(1+x)x + \mathbf{K} + u_0(1+x)^{n-1}x + \mathbf{K} = u_0(1+x)^n \qquad (4.9)$$

Для вычисления частичных сумм этого ряда в Excel может быть использована функция *БЗ (норма; число_периодов; выплата; нз; тип)*. Здесь *норма* — процентная ставка *x*, указываемая в долях единицы; *число_периодов* — количество суммируемых членов ряда *n*, *нз* — сумма первоначального вклада u_0 (заносится со знаком -), параметры *выплата* и *тип* в рассматриваемых примерах не указываются.

Имеется несколько функций для вычисления различных параметров ряда (4.9). Так функция КПЕР позволяет определять число членов ряда *n* по его частичной сумме S_n, значению процентной ставки *x* и начальному значению u_0. Функция ПЗ вычисляет начальное значение u_0 по числу членов ряда *n*, его частичной сумме S_n и значению процентной ставки *x*.

Кроме функций, применяемых для вычислений параметров ряда (4.9), в Excel имеются функции для работы с другими функциональными рядами, обычно используемыми для анализа инвестиционных проектов, работы с ценными бумагами и расчета амортизационных платежей.

Пример. Сумма вклада в банк 12 000 руб. Банк начисляет проценты по сложной ставке 8% годовых. Определить накопленную сумму на вкладе через 11 лет.

Решение.

1. Устанавливаем курсор в свободную ячейку — A1.
2. С панели инструментов Стандартная щелчком мыши на кнопке Вставка функции вызываем Мастер функций. В поле Категория выбираем Финансовые, в поле Функция указываем БЗ. Нажимаем на кнопку ОК.

3. Заполняем поля диалогового окна БЗ. В поле Норма вводим *0,08,* в поле Число_ периодов — *11,ав* поле Нз - - *12 000* (рис. 4.7). Нажимаем кнопку ОК.

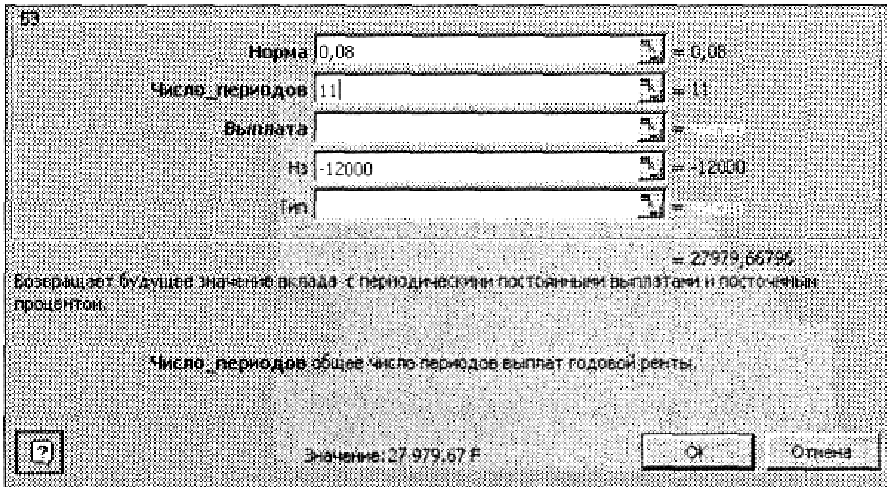


Рис. 4.7. Пример заполнения диалогового окна функции БЗ из примера

4. В результате в ячейке А1 получаем ожидаемую накопленную через 11 лет сумму - 27 575,67руб.

УПРАЖНЕНИЯ

1. Сумма вклада в банк 2000 руб. Банк начисляет проценты по сложной ставке 6% годовых. Определить накопленную сумму на вкладе через 7 лет.

2. Определить сумму вклада через 3 года, если процентная ставка 15% годовых и исходная сумма вклада 5000 руб.

Ряды Фурье

Важным частным случаем функциональных рядов являются ряды Фурье. Во многих задачах бывает необходимо заменить заданную периодическую функцию *f(x)* с периодом Т точно или приближенно тригонометрической суммой (4.10):

$$S_n(x) = \frac{a_0}{2} + a_1 Coswx + a_2 Cos2wx + \mathbf{K} + a_n Cosnwx + b_1 Sinwx + b_2 Sin2wx + \mathbf{K} + b_n Sinnwx \qquad (4.10)$$

где $w = \frac{2p}{T}$

Приближение (4.10) для функции *f*(*x*) будет наилучшим в смысле минимума средней квадратической ошибки:

$$d^2 = \int_0^T [f(x) - S_n(x)]^2 dx$$

если коэффициенты $a_к$ и $b_к$ (k = О, 1, 2,...) вычисляются по формулам Эйлера:

$$a_k = \frac{2}{T}\int_0^T f(x) Cosiwxdx$$

$$b_k = \frac{2}{T}\int_0^T f(x) Siniwxdx$$

Нахождение ряда Фурье данной функции $f(x)$ составляет задачу гармонического анализа.

Если для некоторой совокупности значений x, $s_n(x)$ при $n \to \infty$ стремится к определенному пределу $S(x)$, то для этих x имеем сходящийся ряд Фурье данной функции $f(x)$. Ряд Фурье функции $f(x)$ сходится и сумма его равна $f(x)$ в точках непрерывности $f(x)$, а в точках разрыва она равна

$$\frac{f(x-0) + f(x+0)}{2}$$

если функция $f(x)$ удовлетворяет некоторым условиям, называемым условиями Дирихле. Условия Дирихле заключаются в следующем:

- интервал, на котором функция определена, может быть разбит на конечное число интервалов, в каждом из которых функция $f(x)$ непрерывна и монотонна;
- во всякой точке разрыва $f(x)$ существуют $f(x + 0)$ и $f(x - 0)$

Непериодическая функция $f(x)$ также может быть разложена в ряд Фурье (4.10) на промежутке $0 \le x \le l$, если она удовлетворяет условиям Дирихле на этом промежутке.

В этом случае

$$w = \frac{2p}{l}$$

В MS Excel реализованы процедуры преобразования Фурье: прямая — разложение функции $f(x)$ на промежутке $0 \le x \le l$ в ряд (4.10), и обратная — по значениям коэффициентов a_k и b_k $(k = 0,1,2,...)$ ряда (4.10) восстанавливается исходная функция $f(x)$.

Процедура Анализ Фурье включена в Пакет анализа и предназначена для решения задач в линейных системах и анализа периодических данных с использованием метода быстрого преобразования Фурье (БПФ). Эта процедура поддерживает также обратные преобразования, при этом инвертирование преобразованных данных возвращает исходные данные.

Для выполнения преобразования Фурье необходимо:

1. Ввести точки анализируемой функции $f(x)$ в таблицу. Причем, вследствие использования БПФ, отсчеты должны быть взяты через равные промежутки x, их количество N должно быть четной степенью 2 и максимальное число не должно превышать 4096 $(N \le 4096)$.

2. Командой меню Сервис > Анализ данных вызвать диалоговое окно Пакета анализа и выбрать процедуру Анализ Фурье.

3. Задать параметры диалогового окна Анализ Фурье:

- входной интервал — введя ссылку на диапазон точек анализируемой функции, которые необходимо преобразовать;
- инверсия — флажок устанавливается, если выполняется обратное преобразование, возвращающее данные в выходной диапазон в виде исходной функции. Если флажок сброшен, то в выходной диапазон выводятся значения коэффициентов ряда Фурье;
- выходной интервал — введя ссылку на левую верхнюю ячейку выходного диапазона (рис. 4.8). Размер выходного диапазона будет определен автоматически, и на экран будет выведено

сообщение в случае возможного наложения выходного диапазона на исходные данные.

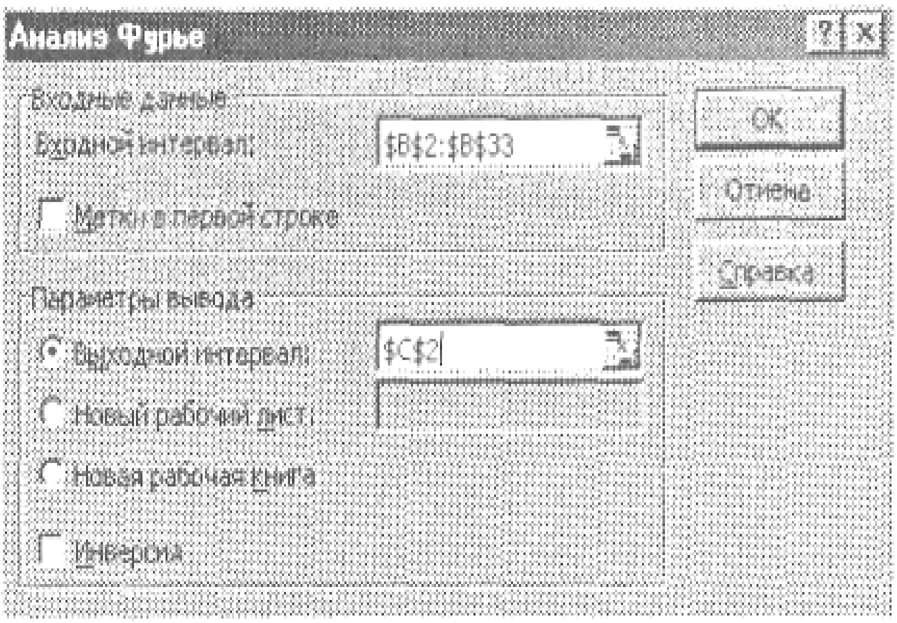


Рис. 4.8. Пример установки параметров в диалоговом окне процедуры Анализ Фурье

Результаты анализа. В результате будет выведен столбец коэффициентов в виде

$$\frac{N}{2}a_0;$$

$$\frac{N}{2}(a_{-1}+b_{-1}i);$$

$$\frac{N}{2}(a_{-2}+b_{-2}i);$$

$$\dots\dots\dots\dots\quad\dots\dots\dots\dots\quad\dots$$

$$\frac{N}{2}(a_{-N/2+1}+b_{-N/2+1}i);$$

$$\frac{N}{2}a_{N/2};$$

$$\frac{N}{2}(a_{N/2-1}+b_{N/2-1}i);$$

$$\dots\dots\dots\dots\quad\dots\dots\dots\dots\quad\dots$$

$$\frac{N}{2}(a_2+b_2i);$$

$$\frac{N}{2}(a_1+b_1i).$$

Здесь коэффициенты a_k и b_k (k = 0, 1, 2, ...) являются коэффициентами разложения функции $f(x)$ в ряд (4.10), где $w = \frac{2p}{N}$

Пример. Получить коэффициенты a_k и b_k (k = О, 1,2,...) разложения в ряд Фурье отрезка функции y = sinjΔx, j = О, 1,2, ..., 31, с шагом Δx = 0,4.

Решение.

1. Введем значения синуса в таблицу. Для этого в ячейку А1 поместим слово *Аргумент, а* в ячейку В1 — *Синус*. В ячейку А2 заносим первое значение интервала — *0*, в ячейку А3 — второе *0,4*. Выделяем блок ячеек А2:А3. За правый нижний угол выделенного блока мышью производим автозаполнение до ячейки А33 и получаем нарастающие значения аргумента до значения *12,4*. Проверяем, что число значений аргумента является целой степенью 2 ($32 = 2^5$).

Устанавливаем табличный курсор в ячейку В 2 и вызываем с панели инструментов Стандартная диалоговое окно Мастер функций (нажатием на кнопку Вставка функции). Указываем Категория — Математические, Функция — SIN. Нажимаем кнопку ОК. В рабочее поле Число диалогового окна SIN щелчком мыши на ячейке А2 вводим значение аргумента. Нажимаем кнопку ОК. Автозаполнением за правый нижний угол копируем значения синуса из ячейки В2 в диапазон В3:В33.

2. Командой меню Сервис > Анализ данных вызываем диалоговое окно Пакета анализа и выбираем процедуру Анализ Фурье.

3. Задаем параметры диалогового окна Анализ Фурье: Входной интервал — мышью вводим ссылку на диапазон точек анализируемой функции В2:В33. В поле Инверсия флажок не устанавливаем. В поле Параметры вывода переключатель ставим в положение Выходной интервал, щелчком мыши активизируем рабочее поле справа и мышью вводим ссылку на левую верхнюю ячейку выходного диапазона — С2 (рис. 4.8). Нажимаем кнопку ОК.

4. В диапазоне С2:С33 появляются значения коэффициентов a_k и b_k (рис. 4.9), причем N в ячейке С2 – a_n и далее в соответ-

ствии с последовательностью (4.11) для гармонических составляющих с частотами

$$-\frac{2p}{N\Delta x},-2\frac{2p}{N\Delta x},-3\frac{2p}{N\Delta x},\mathbf{K},2\frac{2p}{N\Delta x},\frac{2p}{N\Delta x}$$

(в примере - 0,49, -0,98, -1,47, ..., 0,98, 0,49).

	A	B	C	D	E	F	G
1	Аргумент	Синус					
2	0	0	-4,87443320573282E-002				
3	0,4	0,38942	-2,71854388197055E-002-0,378115244661136i				
4	0,8	0,71736	1,76101512258406-15,7164740222652i				
5	1,2	0,93204	-0,174797964628419+0,71781439819842i				
6	1,6	0,99957	-0,140478653886876+0,382563551999639i				
7	2	0,9093	-0,130227914826423+0,263335499922817i				
8	2,4	0,67546	-0,125580263739907+0,198640512000459i				
9	2,8	0,33499	-0,123041362017062+0,156384711869226i				
10	3,2	-0,0584	-0,121498016047588+0,125675621163486i				
11	3,6	-0,4425	-0,120494188076879+0,101716239308027i				
12	4	-0,7568	-0,119812263445973+8,20280358100675E-002i				
13	4,4	-0,9516	-0,119337221063985+6,51799112055733E-002i				
14	4,8	-0,9962	-0,119003809068714+5,02718975695147E-002i				
15	5,2	-0,8835	-0,118773261329621+3,66954982170422E-002i				
16	5,6	-0,6313	-0,118622077393989+2,40102521525438E-002i				
17	6	-0,2794	-0,118536288041853+1,18740738767287E-002i				
18	6,4	0,11655	-0,118508468336846				
19	6,8	0,49411	-0,118536288041853-1,18740738767284E-002i				
20	7,2	0,79367	-0,118622077393981-2,40102521152542E-002i				
21	7,6	0,96792	-0,118773261329622-3,66954982170042E-002i				
22	8	0,98936	-0,119003809068714-5,02718975695146E-002i				
23	8,4	0,8546	-0,119337221063985-6,51799112055729E-002i				
24	8,8	0,58492	-0,119812263445977-8,20280358100712E-002i				
25	9,2	0,22289	-0,120494188076879-0,101716239308027i				
26	9,6	-0,1743	-0,121498016047588-0,125675621163486i				
27	10	-0,544	-0,123041362017062-0,156384711869225i				
28	10,4	-0,8278	-0,125580263739906-0,198640512000459i				
29	10,8	-0,9809	-0,130227914826424-0,263335499922817i				
30	11,2	-0,9792	-0,140478653886876-0,382563551999639i				
31	11,6	-0,8228	-0,174797964628422-0,71781439819842i				
32	12	-0,5366	1,7610151225841+15,7164740222652i				
33							

Рис. 4.9. Пример представления результатов процедуры Анализ Фурье из примера

Пример. Провести анализ результатов, полученных в примере показанных на рис.4.9, построить зависимости изменения коэффициентов a_k и b_k от частоты, записать разложение в ряд Фурье в форме (4.10)

Решение.

1. Занесем в разные диапазоны значения коэффициентов a_k и b_k. Для этого в ячейки C36 и D36 вводим названия a_k, и b_k соответственно.

2. Устанавливаем табличный курсор в ячейку C37 и с панели инструментов Стандартная, нажатием на кнопку Вставка функции, вызываем Мастер функций. В появившемся диалоговом

окне выбираем Категория — Инженерные, Функция — МНИМ.ВЕЩ. Нажимаем кнопку ОК. В рабочее поле Компл_число диалогового окна МНИМ.ВЕЩ щелчком мыши на ячейке С2 вносим ссылку на первую ячейку коэффициентов — *С2.* Нажимаем кнопку ОК. Затем протягиванием из ячейки С37 копируем в диапазон С38:С68 значения коэффициентов a_k.

3. Устанавливаем табличный курсор в ячейку D37 и повторяем все операции аналогично п. 2 с использованием функции МНИМ.ЧАСТЬ. В диапазоне 037:068 получаем значения коэффициентов b_k.

4. Вычислим значения частот. Будем рассматривать только положительные частоты. В ячейку А68 и вносим первое значение *k — 1,* в ячейку А67 — второе значение — *2.* Выделяем блок А68:А67 и автозаполнением заносим значения *k* в диапазон А68:А54. Далее в ячейку В68 вводим формулу для вычисления первой частоты *=(2*ПИ())/(32*0,4)*А68.* Нажимаем клавишу Enter. Протягиванием копируем эту формулу в диапазон В67:В54.

5. Строим диаграмму зависимости коэффициентов a_k и b_k от частоты. Щелчком указателя мыши на кнопке на панели инструментов Стандартная вызываем Мастер диаграмм. В появившемся диалоговом окне выбираем тип диаграммы Точечная, вид — левый столбец, средний ряд (точечная диаграмма со значениями, соединенными сглаживающими линиями). После нажатия кнопки Далее (с помощью мыши) указываем диапазон данных — *B54:D68.* Проверяем положение переключателя Ряды в: столбцах. Щелкаем на вкладке Ряд и в рабочем поле Ряд выбираем *Ряд 1,* после чего щелчком мыши активировав поле Имя вводим a_k. Затем в рабочем поле Ряд выбираем *Ряд 2* и, щелчком мыши активировав поле Имя, вводим b_k. Нажав кнопку Далее, вводим названия осей *X* и У: *Частота* и *Значение X(N/2)* соответственно. Нажимаем кнопку Готово. Появляется диаграмма вида рис. 4.10.

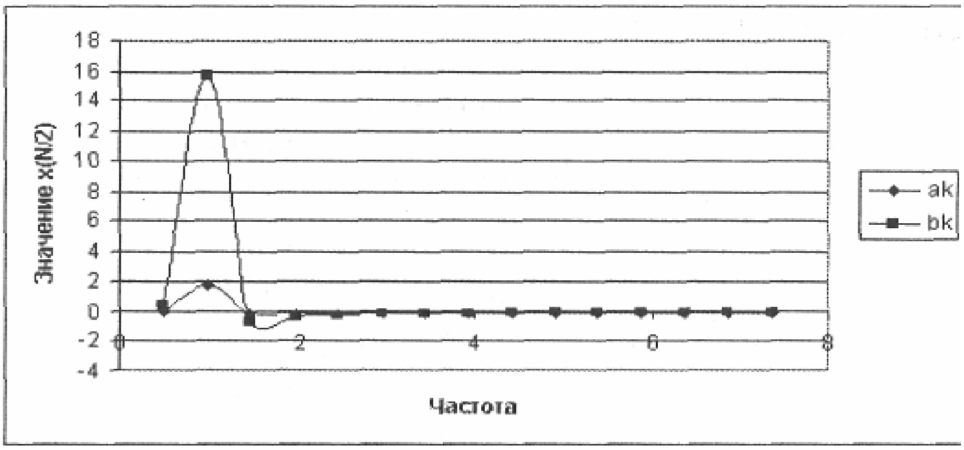


Рис. 4.10. Диаграммы коэффициентов a_k и b_k разложения в ряд Фурье из примера

Из диаграммы, изображенной на рис. 4.10, и найденных значений частот видно, что в ряду Фурье имеют значение только коэффициенты a_k и b_k для гармонических составляющих с частотой *0,98.* Таким образом, приближенно исходную функцию можно представить в виде:

$y \approx 1.67/(N/2)Cos0.98j\Delta x + 15.72/(N/2)Sin0.98j\Delta x$, то есть

$y \approx 0.11Cos\,0.98\,j\Delta x + 0.98\,Sin\,0.98\,j\Delta x$,

что примерно соответствует заданной функции y = sinjΔx. Отличия связаны с тем, что эта функция рассматривается на ограниченном интервале, содержащем только два периода.

Пример. По результатам разложения в ряд Фурье из примера 4.9 восстановить исходную функцию y = sinj Δ x

Решение.

1. Командой меню Сервис > Анализ данных вызываем диалоговое окно Пакета анализа и выбираем процедуру Анализ Фурье.

2. Задаем параметры диалогового окна Анализ Фурье: Входной интервал — мышью вводим ссылку на диапазон коэффициентов a_k и b_k, полученных в примере 4.9 — *С2:С33.* В поле Инверсия — устанавливаем флажок. В поле Параметры вывода переключатель ставим в положение Выходной интервал, щелчком мыши активируем рабочее поле справа и вводим ссылку на левую верхнюю ячейку выходного диапазона — *Н2.* Нажимаем кнопку ОК.

3. В диапазоне H2:H33 получаем значения с точностью до четырех знаков после запятой совпадающие со значениями исходной функции из диапазона B2:B33.

УПРАЖНЕНИЯ.

1. Получить коэффициенты a_k и b_k (k = 0,1,2,...) разложения в ряд Фурье отрезка функции $y = \sin j\Delta x$, j = 0, 1, 2,..., 31, с шагом $\Delta x := 0{,}2$.
2. По результатам разложения в ряд Фурье из рассмотренного примера восстановить исходную функцию $y = \sin j\Delta x$

Комплексные числа

Комплексные числа используются во многих приложениях математики. Теория функций комплексной переменной является мощным инструментом при применении математических методов в различных областях научной и инженерной деятельности.

Представление комплексных чисел

Комплексным числом называется выражение вида:

$$z = x+iy, \tag{4.12}$$

где x и y — действительные числа, i — мнимая единица.

Число x называется *действительной частью числа* z и обозначается $Re(z)$, а число y — *мнимой частью числа* z и обозначается $Im(z)$, то есть $x = Re(z)$, $y = Im(z)$.

Представление комплексного числа в виде (4.12) является алгебраической формой комплексного числа. Существуют также тригонометрическая и показательная формы.

В тригонометрической форме комплексное число может быть представлено как:

$$z = r(Cos\, j + iSin\, j),$$

а *в показательной:*

$$z = re^{ij},$$

Здесь r — модуль комплексного числа z:

$$r = |z| = \sqrt{x^2 + y^2},$$

а j — аргумент комплексного числа z:

$$j = Argz = arctg y / x .$$

Числа $z = x + iy$ *и* $z = x - iy$ *называются сопряженными.*

Два комплексных числа $z_1 = x_1 + iy_1$ и $z_2 = x_2 + iy_2$ называются *равными*, если равны их действительные и мнимые части.

В MS Excel комплексные числа вводятся в ячейки в алгебраическом формате *x+yi*. Если вещественная часть отрицательная, то перед числом ставится апостроф, а если значение мнимой части равно 1, то она все равно должна вводиться, например, *'-3 + 1i*. Для получения различных элементов комплексных чисел и их преобразований существует ряд функций: КОМПЛЕКСН, МНИМ.ABS, МНИМ.АРГУМЕНТ, МНИМ.ВЕЩ, МНИМ.СОПРЯЖ, МНИМ.ЧАСТЬ.

- Функции МНИМ.ВЕЩ(компл_число) и МНИМ.ЧАСТЬ(компл_число) определяют, соответственно, вещественную и мнимую части комплексного числа *компл_число*, представленного в алгебраической форме и записанного в одну ячейку в формате *x + yi*;
- функции МНИМ.ABS(компл_число) и МНИМ.АРГУМЕНТ(компл_число) вычисляют, соответственно, значения модуля и аргумента комплексного числа, представленного в алгебраической форме в формате *x + yi*;
- функция МНИМ.СОПРЯЖ(компл_число) вычисляет сопряженное комплексное число для комплексного числа, представленного в алгебраической форме в формате *x +yi*;
- функция КОМПЛЕКСН *(действительная_часть; мнимая_часть; мнимая_единица)* преобразует коэффициенты при вещественной и мнимой частях комплексного числа в комплексное число в форме *x + yi*. Здесь параметр *действительная _часть* — это действительная часть комплексного числа.

Мнимая _часть — это мнимая часть комплексного' числа.

Мнимая _единица — это обозначение мнимой единицы в комплексом числе. Если аргумент *мнимая _единица* опущен, то предполагается, что он равен *i*.

Пример. Представить комплексное число $z = 2 + i3$ в тригонометрической форме.

Решение.

1. Вводим число z в ячейку A1 в форме *2 + 3i.*

2. Вычисляем модуль комплексного числа. Для этого устанавливаем табличный курсор в ячейку A2 и с панели инструментов Стандартная нажатием на кнопку Вставка функции вызываем Мастер функций. В появившемся диалоговом окне выбираем Категория — Инженерные, Функция — МНИМ.ABS. Нажимаем на кнопку ОК. В рабочее поле Компл_число диалогового окна МНИМ.ABS щелчком мыши на ячейке A1 вводим преобразуемое комплексное число (рис. 4.11). Нажимаем на кнопку ОК. В ячейке A2 получаем значение модуля — *3,605551.*

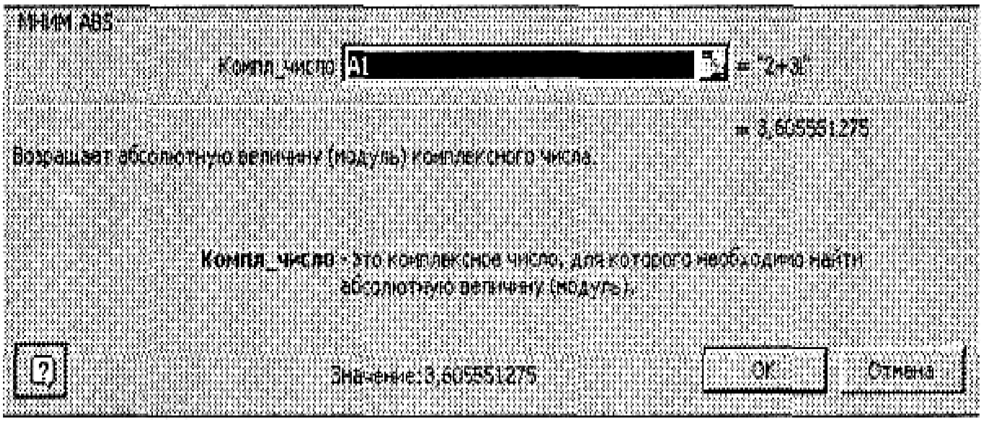


Рис. 4.11. Пример заполнения диалогового окна функции МНИМ.ABS

3. Вычисляем аргумент комплексного числа. Для этого устанавливаем табличный курсор в ячейку A3 и с панели инструментов Стандартная нажатием на кнопку Вставка функции вызываем Мастер функций. В появившемся диалоговом окне выбираем Категория — Инженерные, Функция — МНИМ.АРГУМЕНТ. Нажимаем на кнопку ОК. В рабочее поле Компл_число диалогового окна МНИМ.АРГУМЕНТ щелчком мыши на ячейке A1 вводим преобразуемое комплексное число. Нажимаем кнопку ОК. В ячейке A3 получаем значение аргумента — *0,982794.*

4. Таким образом, в тригонометрической форме комплексное число равно:

z = 3,60555 l(Cos0,982794 + iSin0,982794).

УПРАЖНЕНИЯ

1. Найти сопряженное комплексное число числу 2 - i5.

2. Вещественная часть комплексного числа равна 6, мнимая часть равна -2. Преобразовать в комплексное число в алгебраической форме.

3. Выделить вещественную и мнимую части комплексного числа -3 + i8.

4. Представить комплексное число z = -5 + i4 в тригонометрической форме.

Арифметические операции

На множестве комплексных чисел определены следующие арифметические операции.

1. Сумма (разность) комплексных чисел:

$$z_1 \pm z_2 = x_1 \pm x_2 + i(y_1 \pm y_2) \; ,$$

2. Произведение комплексных чисел:

$$z_1 z_2 = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) \, ,$$

3. Деление двух комплексных чисел:

$$\frac{z_1}{z_2} = \frac{(x_1 x_2 + y_1 y_2) + i(x_2 y_1 - x_1 y_2)}{x_2^2 + y_2^2} \qquad (z_2 \neq 0) \, .$$

4. Возведение комплексного числа в натуральную степень *n:*

$$z^n = r^n Cosn\, j \ + ir^n Sinn\, j \ .$$

5. Извлечение корня из комплексного числа:

$$\sqrt[n]{x} = \sqrt[n]{r} Cos \frac{j \ + 2pr}{n} + i \sqrt[n]{r} Sin \frac{j \ + 2pk}{n} \, ,$$

где k = 0, 1,2, ..., n - 1.

В MS Excel для выполнения арифметических операций с комплексными числами предназначены функции МНИМ.СУММ, МНИМ.РАЗН, МНИМ.ПРОИЗВЕД, МНИМ.ДЕЛ, МНИМ.СТЕПЕНЬ и МНИМ.КОРЕНЬ.

- функции *МНИМ.СУММкомпл_число1;компл_число2;...)*и МНИМ.ПРОИЗВЕД(компл-*_число1;компл_число2;...)* предназначены для вычисления суммы и произведения, соответственно, до 29 комплексных чисел *(компл_число1;компл_число2;...),* представленных в алгебраической форме;

- функции *МНИМ.РАЗН(компл_число1;компл_число2)* и МНИМ.ДЕЛ(*компл_число1;компл_число2)* предназначены, соответственно, для вычисления разности и частного от деления двух комплексных чисел *(компл_число1;компл_число2),* представленных в алгебраической форме;

- функции МНИМ.СТЕПЕНЬ(*компл_число1*; *число*) и МНИМ.КОРЕНЬ(*компл_число*) вычисляют целую или дробную степень *(число)* комплексного числа *(компл_число)* и квадратный корень из комплексного числа *(компл_число),* представленного в алгебраической форме.

Пример. Найти $(4+i6)^{-3}$.

Решение.

1. Вводим число *z* в ячейку А1 в форме *4 + 6i.*
2. Вычисляем степень комплексного числа. Для этого устанавливаем табличный курсор в ячейку А2 и с панели инструментов Стандартная нажатием на кнопку Вставка функции вызываем Мастер функций. В появившемся диалоговом окне выбираем Категория — Инженерные, Функция — МНИМ.СТЕПЕНЬ. Нажимаем на кнопку ОК. В рабочем поле Компл_число диалогового окна МНИМ.СТЕПЕНЬ щелчком мыши на ячейке А1 указываем возводимое в степень комплексное число. В рабочее поле Число с клавиатуры вводим степень --*3* (рис. 4.12). Нажимаем на кнопку ОК. В ячейке А2 получаем возведенное в степень комплексное число--*2,617E-003–5,121E-004i* или в другой записи - *0,002617-0,0005121i.*

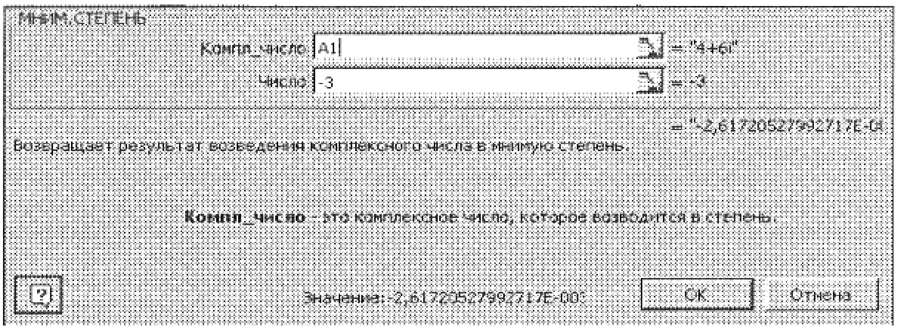


Рис. 4.12. Пример заполнения рабочих полей диалогового окна функции МНИМ.СТЕПЕНЬ

УПРАЖНЕНИЯ.

1. Разделить число 2 – i3 на -2 - i.
2. Сложить комплексные числа 3 + i4, 5 - i2 и -6 + i.
3. Извлечь корень из комплексного числа 4 + i5.
4. Вычислить $\dfrac{(2+i4)(-3-i2)}{(1-i2)(-2+i4)}$

Функции комплексной переменной

Функция $w = f(z)$, где $z = x + iy$ и $w = u + iv$, определена, если известны две функции от двух действительных переменных:

$$u = u(x, y),\ v = v(x, y)$$

Из большого числа функций комплексной переменной в Excel реализованы несколько функций, наиболее часто используемых на практике. К ним относятся:

1. Экспоненциальная функция комплексной переменной

$$e^z = e^x e^{iy} = e^x(Cosy + iSiny),$$

реализованная в специальной функции Excel МНИМ.EXP.

2. Функция натурального логарифма

$$\ln z = \ln\sqrt{x^2 + y^2} + iarctg\frac{y}{x},$$

которой соответствует функция МННМ.LN.

3. Функция десятичного логарифма

$$\lg z = \lg e \ln z$$

ее осуществляет функция МНИМ.LOG10.

4.Функция логарифма по основанию 2

$$\log_2 z = \log_2 e \ln z$$

которой соответствует функция МННМ.LOG2.

5. Функция синуса

$$\sin z = \sin x\, chy + i \cos x\, shy,$$

которой соответствует МНИМ.SIN.

6. Функция косинуса

$$\cos z = \cos x\, chy - i \sin x\, shy,$$

которой соответствует МНИМ.COS. Здесь *shy* и *chy* — гиперболические функции синуса и косинуса соответственно:

$$shy = \frac{e^y - e^{-y}}{2}$$

$$chy = \frac{e^y + e^{-y}}{2}$$

Все специальные функции Excel, реализующие названные функции комплексной переменной: МНИМ.EXP, МННМ.LN, МНИМ. LOG10, МНИМ.LOG2, МНИМ.SIN и МНИМ.COS имеют единственный параметр — *компл_число*. *Компл_число* — это комплексное число в алгебраической форме, которое является переменной z для соответствующей функции. Результатом вычислений также является комплексное число в алгебраической форме.

Пример. Найти значение e^{3-i2}.

Решение.

1. Вводим число **z** в ячейку A1 в форме 3 - 2i.

2. Вычисляем экспоненту комплексного числа. Для этого устанавливаем табличный курсор в ячейку А 2 и с панели инструментов Стандартная нажатием на кнопку Вставка функции вызываем Мастер функций. В появившемся диалоговом окне выбираем Категория — Инженерные, Функция — МНИМ.EXP. Нажимаем на кнопку ОК. В рабочее поле Компл_число диалогового окна МНИМ.EXP щелчком мыши на ячейке A1 вносим ссылку на комплексное число, являющееся степенью *e* (рис. 4.13). Нажимаем на кнопку ОК. В ячейке А2 получаем экспоненту комплексного числа - *-8,35853265093537-18,2637270406668i*

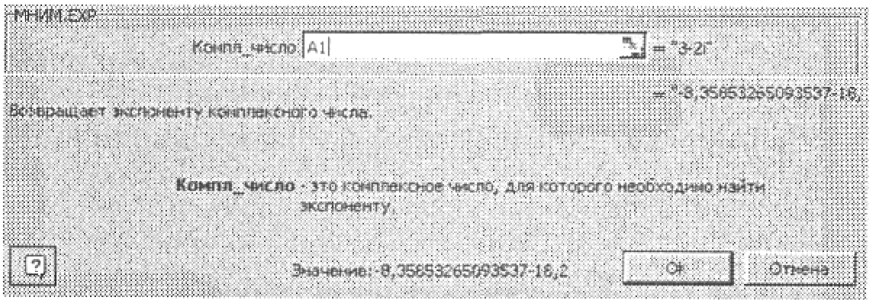


Рис. 4.13. Пример заполнения диалогового окна функции МНИМ.EXP

УПРАЖНЕНИЯ

1. Вычислить десятичный логарифм комплексного числа 4- *i*.
2. Найти ln(2 - i2)/sin(3 + i2).
3. Вычислить sin(l + i2)cos(2 – i3).

5. Теория вероятностей

Основные понятия теории вероятности.

Предметом теории вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий.

Понятие случайного события.

Наблюдение какого-либо явления, происходящего при определенных условиях, в теории вероятностей принято называть испытанием (случайным испытанием). Результат испытания является событием. Например, выполнение одиннадцатиметрового штрафного удара в футболе есть испытание, а попадание (или непопадание) в ворота – событие.

Операции над событиями дают возможность описывать новые более сложные события и их структуру. Рассмотрим основные операции над событиями:

Сумма событий. Сумма или объединение событий А и В это событие С, содержащие все исходы, входящие в А и В, и в оба события вместе.

Обозначаем как $C = A + B$ или $C = A \cup B$.

Например, объединяем события А, состоящего в том, что в результате бросания кости выпадает меньше 4 очков, и события В, состоящего в том, что в результате бросания кости выпадает 3 или 6 очков, будет событие А + В, состоящего в том, что в результате бросания кости выпадет 1, 2 ,3 или 6 очков.

Произведение событий. Произведением или пересечением событий А и В называется событие С, состоящее из всех исходов, общих для А и В.

Обозначается как $C = AB$ или $C = A \cap B$

Например, пересечением АВ событий А и В из предыдущего примера будет событие, состоящее в том, что в результате бросания кости выпадет 3 очка.

События, которые не могут произойти одновременно, то есть произведение которых равно нулю ($AB = 0$), называется *несовместимыми*.

Отрицание. Событие $\bar{A}$ является отрицанием или называется противоположным событию А, если оно содержит все исходы пространства событий, не вошедшие в А.

Возвращаясь к рассмотренному примеру, получим, что отрицание события А (событие $\bar{A}$) состоит в том, что в результате бросания кости выпадет 4, 5 или 6 очков.

Разность. Разностью называется событие С, содержащее все исходы А, не входящие в В (С = А - В).

В условиях предыдущего примера разностью А – В является событие, состоящее в том, в результате бросания кости выпадет 1 или 2 очка.

Вероятность события.

Вероятностью события А называется отношение числа благоприятных элементарных исходов к числу всех возможных исключающих друг друга исходов.

$P(A) = \frac{m}{n}$, где А – событие;

n – число всех возможных исходов;

m – число благоприятных исходов;

P(A) – вероятность события А.

События, вероятности которых одинаковы, называются равновероятными. Для равновероятных событий вероятности вычисляются по формуле $P(A_i)= 1/k$, где k – число элементарных событий пространства.

Пример.

Какова вероятность того, что наугад вырванный из календаря листок соответствует 30 числу, если в году 365 дней?

Решение. Событие А – на листке календаря число 30.

Количество всех возможных исходов n = 365. Количество благоприятных исходов m = 11.

Устанавливаем табличный курсор в ячейке А1 и вводим формулу: =11/365. Нажимаем Enter. Получаем, что вероятность события А:

$$P(A) = \frac{m}{n} = \frac{11}{365} = 0{,}030137$$

Ответ: вероятность того, что вырванный наугад из календаря листок соответствует 30 числу, равна P(A) = 0,03

Условная вероятность.

Если на событие А влияет исход некоторого события В (в зависимости от того, произошло ли событие В и с каким именно исходом, наступает или не наступает определенный исход, наступает ил не наступает определенный исход А), речь идет об условной вероятности А при условии В.

Условная вероятность Р(А/В) события А при условии, что событие В уже наступило, определяется по формуле:

$P(A/B) = P(AB) / P(B)$

Основные понятия комбинаторики.

В ряде случаев сравнительно нетрудно подсчитать как общее число элементарных исходов, так и число исходов, благоприятных для данного события. Однако в большинстве случаев именно этот подсчет и представляет наибольшую трудность при решении более сложных задач на классическую вероятность. Для того чтобы владеть некоторыми стандартными приемами при расчетах по схеме классической вероятности, рассмотрим понятие перестановки, размещения и сочетания и приведем основные правила комбинаторики.

Перестановки.

Перестановкой n элементов множества называется их комбинация, отличающаяся только порядком расположения.

В комбинаторике принято следующее обозначение: если в множестве имеется n элементов, то $П_n$ – число перестановок элементов этого множества.

Вычислить число перестановок можно по следующей формуле:

$$П_n = n * П_{n-1} \qquad (1.1)$$

Формула (1.1) называется рекуррентной, так как позволяет вычислить значение очередной величины через предыдущее.

Для нахождения числа перестановок в Excel используется специальная функция – ФАКТР.

Пример. Сколькими способами можно расставить шесть различных книг на полке?

Решение.

1. Устанавливаем табличный курсор в свободную ячейку, например A1. Здесь должно оказаться значение числа перестановок.

2. Для получения значения числа перестановок воспользуемся специальной функцией: нажимаем на панели инструментов кнопку Вставка функции (f_x).

3. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Математические. Справа в поле Функция выбираем функцию ФАКТР. Нажимаем на кнопку ОК.

4. Появляется диалоговое окно ФАКТР. В рабочее поле Число вводим с клавиатуры число переставляемых объектов (в примере 6). Нажимаем на кнопку ОК.

5. В ячейке A1 появляется искомое число перестановок – 720. следовательно $П_6 = 6! = 1 * 2 * 3 * 4 * 5 * 6 = 720$

УПРАЖНЕНИЯ.

1. Сколькими способами можно рассадить за столом 7 человек гостей?

2. Сколько различных восьмизначных чисел можно составить из цифр 1, 2, 3, 4, 5, 6, 7, 8?

3. Сколько различных комбинаций букв можно составить из всех букв слова "бухгалтер"?

4. Сколько различных слов можно составить из всех букв слова:

а) колобок;

б) пудель (если принять, что с "ь" слова не начинаются).

5. Вычислить:

а) $\dfrac{12!+131}{П_{11}}$

б) $\dfrac{9!-7!}{8!}$

в) $\dfrac{15!+17!}{16!}$

Сочетания

Сочетаниями называются конечные подмножества, составленные из элементов данного множества. Если в множестве элементов – n, а в подмножестве m, то общее количество всех сочетаний обозначается и читается как число сочетаний из n элементов по m:

Очевидно, что n ≥ m. Приведем формулу для вычисления числа сочетаний:

$$C_n^m = \frac{n!}{m!*(n-m)!}$$

Для нахождения числа сочетаний в Excel используется специальная функция – ЧИСЛКОМБ. В функции ЧИСЛКОМБ (число; число_выбранных) должны быть заданы следующие параметры: число – это число объектов n; число_выбранных – это число объектов в каждой комбинации m.

Пример. Сколькими способами можно выбрать 3 книги из 5 книг, имеющихся в наличии?

Решение.

1. Устанавливаем табличный курсор в свободную ячейку, например в А1. Здесь должно оказаться значение числа сочетаний.
2. Для получения значения числа сочетаний воспользуемся специальной функцией: нажимаем на панели: нажимаем на панели инструментов кнопку Вставка функции (f_x).
3. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Математические. Справа в поле Функция выбираем функцию ЧИСЛКОМБ. Нажимаем на кнопку ОК.
4. Появляется диалоговое окно ЧИСЛКОМБ. В рабочее поле Число вводим с клавиатуры общее число объектов n (в примере -
5. В рабочее поле Выбранное число вводим с клавиатуры число объектов, которые необходимо выбрать, m (в примере - 3). Нажимаем на кнопку ОК.
6. В ячейке А1 появляется искомое число $C_5^3 = 10$. Таким образом, 3 книги из 5 имеющихся можно выбрать десятью способами.

УПРАЖНЕНИЯ.

1. Сколько различных списков дежурных из 5 человек можно составить в группе из 25 человек?
2. Сколько различных букетов из 9 цветков можно составить из 15 полевых ромашек?

Размещения.

Размещениями называются конечные упорядоченные подмножества из элементов данного множества. Общее количество размещений обозначается, как: A_n^m и читается: число размещений из n по m (n ≥ m), чем отличаются подмножества сочетания и подмножества размещения? Во – втором случае важен порядок расположения элементов.

Приведем формулу для вычисления числа размещений:

$$A_n^m = C_n^m * m! = \frac{n!}{m!*(n-m)!} * m! = \frac{n!}{(n-m)!}$$

Для нахождения числа размещений на компьютере можно воспользоваться формулой $A_n^m = C_n^m * m!$ или специальной функцией – ПЕРЕСТ. В функции ПЕРЕСТ (число; число_выбранных) должны быть заданы следующие параметры: число – число объектов n; число_выбранных – число объектов в каждой комбинации m.

$$A_n^m = \frac{n!}{(n-m)!}, \text{ n} \geq \text{m}$$

Рассмотрим оба способа нахождения числа размещений.

Пример.

Сколькими способами можно расставить на полке 3 выбранных книги из 5 книг, имеющихся в наличии?

Решение 1.

1. Устанавливаем табличный курсор в свободную ячейку, например в A1.
2. Находим число сочетаний. Нажимаем на панели инструментов кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Математические. Справа в по-

ле Функция выбираем функцию ЧИСЛКОМБ. Нажимаем на кнопку ОК. в диалоговом окне ЧИСЛКОМБ в рабочем поле Число вводим с клавиатуры общее число объектов n (в примере - 5). В рабочее поле Выбранное число вводим с клавиатуры число объектов, которые необходимо выбрать, m (в примере - 3). Нажимаем на кнопку ОК. В ячейке А1 появляется искомое число сочетаний = 10.

3. Указателем мыши щелкаем в Строке формул после последней скобки формулы =ЧИСЛКОМБ (5;3) и вводим с клавиатуры знак умножения - *.

4. Для получения значения m! воспользуемся специальной функцией: нажимаем на панели инструментов кнопку Вставка функции (f_x).

5. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Математические. Справа в поле Функция выбираем функцию ФАКТР. Нажимаем кнопку ОК.

6. Появляется диалоговое окно ФАКТР. В рабочее поле Число вводим с клавиатуры число переставляемым объектом m (в примере - 3). Нажимаем на кнопку ОК.

7. В ячейке А1 появляется искомое число размещений $A_5^3 = 60$

Решение 2.

1. Устанавливаем табличный курсор в свободную ячейку, например в А2.

2. Для получения значения числа размещений воспользуемся специальной функцией ПЕРЕСТ – нажимаем на панели инструментов кнопку Вставка функции (f_x).

3. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функции. Выбираем Статистические. Справа в поле Функция выбираем функцию ПЕРЕСТ. Нажимаем на кнопку ОК.

4. Появляется диалоговое окно ПЕРЕСТ. В рабочее поле Число вводим с клавиатуры общее число объектов n (в примере - 5). В рабочее поле Выбранное число вводим с клавиатуры число объектов, которые необходимо выбрать и переставить, m (в примере - 3). Нажимаем на кнопку ОК.

5. В ячейке А2 появляется искомое число размещений $A_5^3 = 60$

Таким образом, 3 книги из 5 имеющихся можно выбрать и расставить на полке шестьюдесятью способами.

УПРАЖНЕНИЯ.

1. Сколько можно составить трехзначных чисел из цифр 7, 9, 6, 5, 4? Что здесь нужно считать: перестановки, сочетания или размещения? Почему?
2. Сколько трехзначных чисел, не начинающихся с 0, можно составить из всех цифр?

Основные правила комбинаторики.

Если некоторый выбор А можно осуществить m способами, а выбор В – n способами, то выбрать либо А либо В можно m+nспособами. Это правило называется правилом суммы.

Если некоторый выбор А можно осуществить m различными способами, а для каждого из этих способов некоторый другой выбор В можно осуществить n способами, то выбор А и В можно осуществить m * n способами. Это правило произведения.

Оба приведенные выше правила могут быть распространены на произвольное число совместно осуществляемых выборов.

Решая комбинаторную задачу, прежде всего надо ответить на вопрос – с каким из основных понятий в данной ситуации мы имеем дело?

А для этого отвечаем на два вопроса:

1. Все элементы множества используются или нет? Если используются все элементы, то это перестановка.
2. Важен порядок расположения элементов или нет? Если порядок важен, то это размещение, в противном случае – сочетание.

Бином Ньютона.

Бином Ньютона служит для возведения в степень суммы двух слагаемых (бином – это двучлен).

Формула общего члена может быть представлена выражением:

$$T_{л+1} = C_n^k * a^{n-k} * b^k$$

В сокращенном виде формула бинома будет выглядеть так:

$$(a+b)^n = \sum_{k=0}^{n} C_n^k * a^{n-k} * b^k$$

УПРАЖНЕНИЯ.

1. Сколькими способами из 9 человек можно выбрать комиссию из 5 человек?
2. Сколько различных пятибуквенных слов можно составить из букв слова "треугольник"? (Слово может начинаться с любой буквы и представляет собой любую комбинацию букв).
3. В группе из 20 юношей и 10 девушек формируется волейбольная команда. Какова вероятность того, что команда будет состоять из 4 юношей и 2 девушек?
4. С какой вероятностью можно угадать три номера в тираже спортлото "5 из 36"?
5. Составляется букет из 11 ромашек и 8 васильков. Если в букете 9 цветов, какова вероятность того, что он будет состоять из 5 ромашек и 4 васильков?
6. Какова вероятность того, что патруль будет состоять из одного офицера и двух солдат, если в подразделении 60 солдат и 5 офицеров?
7. Какова вероятность того, что, составляя слово из букв И, М, А, З, получится слово "зима"? Из букв А, М, А, М, получится "мама"?
8. В меню указано три первых блюда, пять вторых, четыре третьих. Сколько различных комплексных обедов можно составить?
9. Из Санкт-Петербурга в Москву можно добираться самолетом, поездом, на автомобиле. Из Москвы в Владивосток можно долететь, доплыть, доехать поездом. Сколькими способами можно добраться из Санкт-Петербурга до Владивостока через Москву?
10. Сколькими способами можно выбрать с полки 3 книги, если их там 10 штук?
11. Сколькими способами можно выбрать для сотрудничества 2 турфирмы, если их 25?

12. Сколько различных маршрутов можно составить для знакомства с городами Париж, Вена, Берлин, Мадрид, Рим?

13. Сколько возможно маршрутов, если клиент пожелал первым посетить Рим?

14. Сколько возможно маршрутов, если путешествие должно завершиться в Париже?

15. Сколько возможно маршрутов, если маршрут должен начаться в Риме, а закончиться в Вене?

16. Вы заключили контракт с турфирмами 9 европейских городов. В маршрут хотите включить 3. Сколько различных маршрутов можно составить?

Случайные величины.

Зная распределения вероятностей интересующихся нас случайных величин, можно делать выводы о событиях, в которых участвуют эти величины. Правда, эти выводы будут также носить случайный характер.

Большинство применяемых на практике распределений являются дискретными или непрерывными.

Случайная величина с дискретным распределением.

Случайная величина называется *дискретной*, если значения, которые она может принять, можно пронумеровать. Число этих значений и неограниченным, нужно лишь, чтобы мог быть указан метод нумерации, при котором не будет пропущено ни одного возможного значения случайной величины. Иначе говоря, дискретной случайной величиной называется такая случайная величина, которая может принимать значения, образующие счетные множества.

Примером случайной величины является число лепестков в цветке сирени. Как известно, вероятность того, что у случайно выбранного цветка сирени имеется четыре лепестка, близка к единице, вероятности наличия трех и пяти лепестков малы, а вероятности наблюдения других значений числа лепестков ничтожно малы.

Непрерывной называется случайная величина, возможные значения которой непрерывно заполняют какой-либо интервал (например, температура воздуха на улице).

В MS EXCEL для вычисления некоторых числовых характеристик дискретных распределений вероятностей могут быть использованы специальные функции СРЗНАЧ, ДИСПР, СТАНДОТКЛОНП, КВАНТИЛЬ И ПЕРСЕНТИЛЬ.

- СРЗНАЧ – вычисляет математическое ожидание для всей совокупности значений дискретной случайной величины. Например, если ячейки А1: А7 содержат числа 10, 14, 5, 6, 10, 12 и 13, являющиеся дискретным распределением, то математическим ожиданием СРЗНАЧ (А1:А7) является 10;
- ДИСПР – позволяет оценить дисперсию дискретного распределения. Например, ДИПСР (10, 14, 5, 6, 10, 12, 13) равняется 10;
- СТАНДОТКЛОНП – вычисляет стандартное отклонение дискретного распределения. Например, СТАНДОТКЛОНП (10, 14, 5, 6, 10, 12, 13) равняется 3,162278;
- КВАНТИЛЬ – позволяет определить кваНтили распределения. Функция КВАНТИЛЬ имеет формат КВАНТИЛЬ (массив; значение). Массив представляет собой интервал ячеек, содержащих все значения дискретной случайной величины, или сами значения, а значение определяет, какая квартиль должна быть найдена (0- минимальное значение распределения, 1- нижний квартиль, 2- медиана, 3- верхний квартиль, 4- максимальное значение распределения). Например, КВАНТИЛЬ ({10, 14, 5, 6, 10, 12, 13}; 3) равна 12,5. это значение, что четверть всех значений распределения больше чем 12,5.
- ПЕРСЕНТИЛЬ позволяет получить р-квантили заданного распределения. Например, если ячейки А1:А7 содержат числа 10, 14, 5, 6, 10, 12 и 13, то квантилью со значением 0,1 является ПЕРСИНТИЛЬ (А1:А7; 0,1), равна 5,6.

Пример.

Дано дискретное распределение со значениями 1, 2, 3, 4, 10, 2, 3, 1, 3, 3, 5. найти 90%- ную р-квантиль.

Решение.

Для нахождения p-квантили прежде всего необходимо ввести данные в рабочую таблицу. Открываем новую рабочую таблицу. Вводим в ячейки C1:C11 значения случайной величины.

Затем необходимо установить табличный курсор в свободную ячейку (C12). На панели инструментов Стандартная нажать кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выбрать категория Статистические и функцию ПЕРСЕНТИЛЬ, после чего нажать на кнопку ОК. Появившееся диалоговое окно ПЕРСЕНТИЛЬ за серое поле мышью отодвинуть вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши в рабочее поле Массив ввести диапазон данных распределения (C1:C11). В рабочее поле k с клавиатуры ввести требуемое значение – 0,9. нажать на кнопку ОК.

В ячейке C12 появится значение 90% квантили распределения – 5. это означает, что 90% всех значений распределения меньше чем 5.

УПРАЖНЕНИЯ.

1. Найти математическое ожидание и стандартное отклонение дискретного распределения 0,2; 0,5; 2; 3; 5,1; 8; 2; 3.
2. Найти 25%-ную квартиль для распределения из упражнения 1.
3. Найти дисперсию для распределения из упражнения 1.

Законы распределения вероятностей.

Среди дискретных распределений наиболее важные – биноминальное и пуассоновское, среди непрерывных – нормальное, показательное и распределения, связанные с нормальным: Стьюдента, хи - квадрат и F – распределение Фишера. Итак рассмотрим некоторые из наиболее важных распределений вероятностей.

Биноминальное распределение.

Биноминальное распределение – это одно из дискретных распределений, оно служит вероятностью моделью для многих явлений. Оно возникает в тех случаях, когда нас интересует, сколько раз происходило некоторое событие в серии из определенного числа независимых наблюдений (опытов), выполняемых в одинаковых условиях. При этом распределении разброс вариант (в простейшем случае есть событие или нет события) является следствием влияния ряда независимых и случайно сочетающихся факторов.

В MS EXCEL для вычисления вероятности отдельного значения биноминального распределения или значения случайной величины по заданной вероятности используются функции БИНОМРАСП и КРИТБИНОМ.

Функция БИНОМРАСП применяется для вычисления вероятности в задачах с фиксированным числом тестов и испытаний, когда результатом любого испытания может быть только успех или неудача, испытания независимы, и вероятность успеха постоянна на протяжении всего эксперимента.

Функция использует следующие параметры – БИНОМРАСП (*число_успехов; число_испытаний; вероятность_успеха; интегральная*); Здесь:

- *число_успехов* – это количество успешных испытаний;
- *число_испытаний* – это число независимых испытаний. При этом *число_успехов и число_испытаний* являются целыми числами;
- *вероятность_успеха* – это вероятность успеха каждого испытания;
- *интегральная* – это логическое значение, определяющее форму функции. Если аргумент интегральная имеет значение ИСТИНА (1), то функция БИНОРАСП возвращает интегральную функцию распределения, то есть вероятность того, что число успешных испытаний не менее значения аргумента число_успехов; если этот аргумент имеет значение ЛОЖЬ (0), то вычисляется значение функции плотности распределения, то есть вероятность того, что число успеешных испытаний в точности равно значению аргумента число_успехов.

Интегральное биноминальное распределение имеет следующий вид:

$$P(k) = \sum_{m=0}^{k} p_n(m) = \sum_{m=0}^{k} \frac{n!}{m!(n-m)!} p^m q^{n-m}$$

Функция КРИТБИНОМ вычисляет наименьшее значение числа успешных исходов случайной величины, для которого интегральное биноминальное распределение больше или равно заданной величине (критерию). Эта функция часто используется в приложениях, связанных с контролем качества.

Функция использует параметры: КРИТБИНОМ (*число_испытаний; вероятность_успеха; альфа*). Здесь:

- *число_испытаний* – это число независимых двух альтернативных испытаний;
- *вероятность_успеха* – это вероятность успеха в каждом испытании;
- *альфа* – это значение критерия, которое фактически является уровнем значимости.

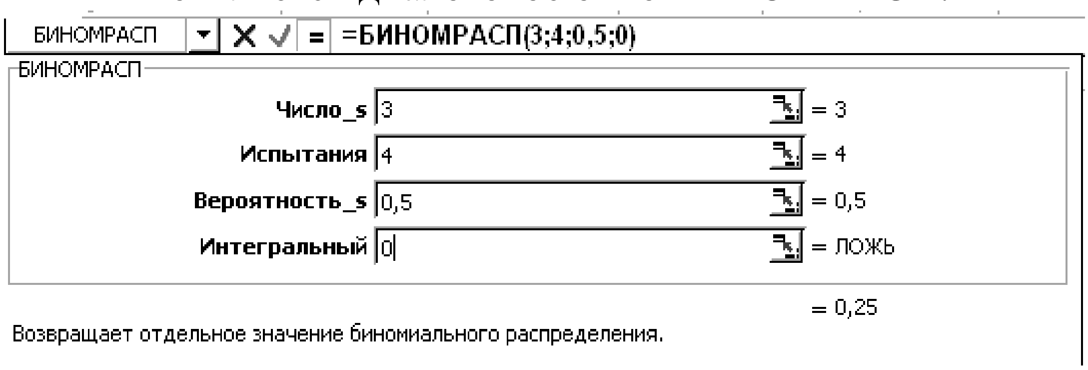
Пример.

Какова вероятность того, что трое из четырех следующих новорожденных будут мальчиками?

Решение.

1. Устанавливаем табличный курсор в свободную ячейку, например в A1. здесь должно оказаться значение искомой вероятности.
2. Для получения значения вероятности воспользуемся специальной функцией: нажимаем на панели инструментов кнопку Вставка функции (f_x).
3. В появившемся диалоговом окне Мастер функций шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Статистическая. Справа в поле Функция выбираем функцию БИНОМРАСП. Нажимаем ОК.

Появляется диалоговое окно БИНОМРАСП.



В рабочее поле Число_s вводим с клавиатуры количество успешных испытаний m (в примере - 3). В рабочее поле Испытания вводим с клавиатуры общее количество испытаний n (в примере - 4). В рабочее поле Вероятность_s вводим с клавиатуры вероятность успеха в отдельном испытании р 9в примере – 0,50. в рабочее поле Интегральный вводим с клавиатуры вид функции распределения – интегральная или весовая (в примере - 0). Нажимаем на кнопку ОК. в ячейке А1 появляется искомое значение вероятности р = 0,25.

Таким образом, ровно 3 мальчика из 4 новорожденных могут появиться с вероятностью р = 0,25.

Если изменить формулировку условия задачи и выяснить вероятность того, что появится не более 3 мальчиков, то в этом случае в п.4 в рабочее поле Интегральный вводим 1 (вид функции распределения интегральный) тогда вероятность этого события будет равна р = 0,9375.

Пример.

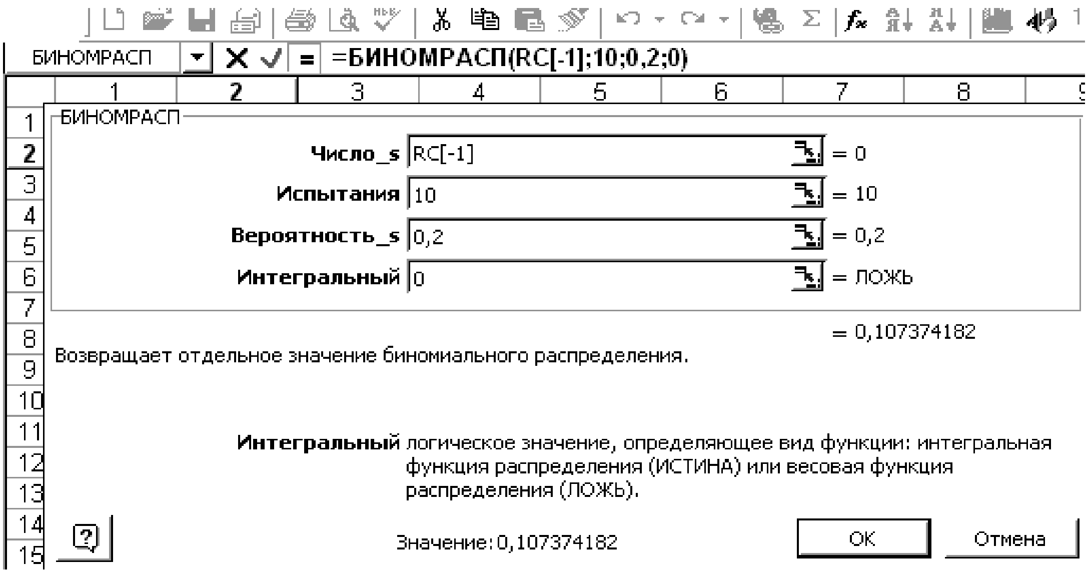
Построим диаграмму биномиальной функции плотности вероятности Р(А = m) при n = 10 и р = 0,2.

Решение.

1. В ячейку А1 вводим символ количества успешных исходов m, а в ячейку В1 – символ вероятности – р.
2. Заполняем диапазон А2:А12 возможными значениями исходов: 0, 1, 2, …,10.
3. Устанавливаем табличный курсор в ячейку В2 и для получения значения вероятности воспользуемся специальной функцией – нажимаем на панели инструментов кнопку Вставка функции (f_x).
4. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева Категория указаны виды функций. Выбираем Статистические. Справа в поле Функция выбираем функцию БИНОМРАСП. Нажимаем на кнопку ОК.
5. Появляется диалоговое окно БИНОМРАСП. В рабочее поле Число_s вводим количество успешных испытаний m (в примере адрес ячейки А2 щелчком мыши на этой ячейке). В рабочее поле Испытания вводим с клавиатуры общее количество испытаний n (в примере - 10). В рабочее поле Вероятность_s вводим с клавиатуры вероятность успеха в отдельном испытании р (в при-

мере – 0,2). В рабочее поле Интегральный вводим с клавиатуры вид функции распределения – интегральная или весовая (в примере - 0). Нажимаем на кнопку ОК.

В ячейке В2 появляется значение вероятности р = 0,107374



Указателем мыши за правый нижний угол табличного курсора протягиваем (при нажатой левой кнопке мыши) из ячейке В2 до В12 копируем функцию БИНОМРАСП в диапазоне В3:В12.

По полученным данным строим искомую диаграмму биноминальной функции распределения. Щелчком указателя мыши на кнопку на панели инструментов вызываем Мастер диаграмм. В появившемся диалоговом окне выбираем тип диаграммы Гистограмма, вид – левый верхний. После нажатия кнопки Далее указываем диапазон данных – В2:В12 (с помощью мыши). Проверяем положение переключателя Ряды в: в столбцах. Выбираем вкладку Ряд с помощью мыши вводим диапазон подписей оси Х: А2:А12. нажав на кнопку Далее, вводим название осей X и Y: m и р, соответственно. Нажимаем на кнопку Готово.

Полученная диаграмма биноминальной функции плотности распределения приведена выше.

Пример.

В условиях предыдущего примера (биноминальная функция плотности вероятности P(A= m) при n = 10 и p = 0,2) найти значение числа m, для которого вероятность интегрального распределения равна или больше $P \geq 0{,}3$.

Решение.

1. Устанавливаем табличный курсор в свободную ячейку, например в A1. здесь должно оказаться значение искомого числа m.
2. Для получения значения вероятности воспользуемся специальной функцией КРИТБИНОМ: нажимаем на панели инструментов кнопку Вставка функции (f_x).
3. В появившемся диалоговом окне Мастер функций – шаг 1 из 3 слева в поле Категория указаны виды функций. Выбираем Статистические. Справа в поле Функция выбираем функцию КРИТБИНОМ. Нажимаем ОК.
4. Появляется диалоговое окно КРИТБИНОМ. В рабочее поле Испытания вводим с клавиатуры количество испытаний n (в примере - 10). В рабочее поле Вероятность_s вводим с клавиатуры вероятность успеха в отдельном испытании p (в примере – 0,2). В рабочее поле Альфа вводим с клавиатуры критическое значение вероятности интегрального распределения P (в примере – 0,3). Нажимаем на кнопку ОК.

5. В ячейке А1 появляется искомое значение числа успешных событий m = 1.

Таким образом, при вероятности интегрального распределения Р ≥0,3, появится не менее одного (m ≥0,3) успешного события. Действительно, вычислив БИНОМРАСП (1; 10; 0,2; 1) получим 0,37581, а БИНОМРАСП (0; 10; 0,2; 1) – 0,107374.

УПРАЖНЕНИЯ.

1. При бросании монеты может выпасть орел или решка. Вероятность того, что при очередном бросании выпадет орел, равна 0,5. найти вероятность того, что орел выпадет в точности 6 раз из 10.
2. Построить диаграмму биноминальной функции плотности вероятности P(A=m) при n = 10 и p=0,5.
3. Построить диаграмму биноминальной интегральной функции распределения P(A ≤ m) при n = 10 и p=0,2.
4. Выборочный контроль продукции проводят так: из партии в 100 изделий выбирается 20 и при обнаружении в этой выборке хотя бы одного дефектного изделия вся партия бракуется. В партии имеется 10 дефектных изделий. Какова вероятность того, что хотя бы одно дефектное изделие попадет в выборку?
5. Найти количество успешных испытаний для критического значения интегральной функции распределения, равного 0,75, если общее количество испытаний равно 6, а вероятность успеха в испытании – 0,5.

Нормальный закон распределения.

Нормальное распределение получило широкое распространение для приближенного описания случайных явлений, в которых на результат воздействует большое количество независимых случайных факторов, среди которых нет сильно выделяющихся. Например, рассеяние снарядов при стрельбе.

В Excel для вычисления значений нормального распределения используются функции: НОРМРАСП, НОРМСТРАСП, НОРМОБР, НОРМСТОБР и НОРМАЛИЗАЦИЯ.

Функция НОРМРАСП вычисляет значения вероятности нормальной функции распределения для указанного среднего и стандартного отклонения. Эта функция имеет очень широкий круг приложений в статистике, включая проверку гипотез.

Функция имеет параметры – НОРМРАСП (*x; среднее; стандартное_откл; интегральная*). Здесь:

- *x* – значение, для которого строится распределение;
- *среднее* – среднее арифметическое распределение;
- *стандартное_откл* – стандартное отклонение распределения;
- *интегральная* – логическое значение, определяющее форму функции. Если интегральная имеет значение Истина, то функция НОРМРАСП возвращает интегральную функцию распределения; если это аргумент имеет значение ЛОЖЬ, то вычисляет значение функция плотности распределения.

Если среднее = 0 и стандратное_откл = 1, то функция НОРМРАСП возвращает стандартное нормальное распределение, то есть НОРМСТРАСП.

Функция НОРМСТРАСП используется для вычисления стандартного нормального интегрального распределения. Это распределение имеет среднее равное нулю и стандартное отклонение равное единице. Эта функция может использоваться вместо таблицы для стандартной нормальной кривой.

Функция имеет параметры – НОРМСТРАСП(z) , где z - значение случайной величины, для которого строится распределение.

Функция НОРМОБР вычисляет значения квантилей для указанного среднего и стандартного отклонения (решения уравнения F(x) = p).

Функция имеет параметры – НОРМОБР (*вероятность; среднее; стандартное_откл*). Здесь:

- *вероятность* – вероятность, соответствующая нормальному распределению;
- *среднее* – среднее арифметическое распределения;
- *стандартное_откл* – это стандартное отклонение распределения.

Функция НОРМСТОБР аналогична функции НОРМОБР в случае, если среднее = 0 и стандартное_отклонение = 1. при этом используется стандартное нормальное распределение. Единственным параметром функции НОРМСТОБР является вероятность:

Функция НОРМАЛИЗАЦИЯ позволяет по значению х и параметрам распределения найти нормализованное значение, соответствующее заданному х. Формат функции – НОРМАЛИЗАЦИЯ (х; среднее; стандартное_отклонение).

Пример.

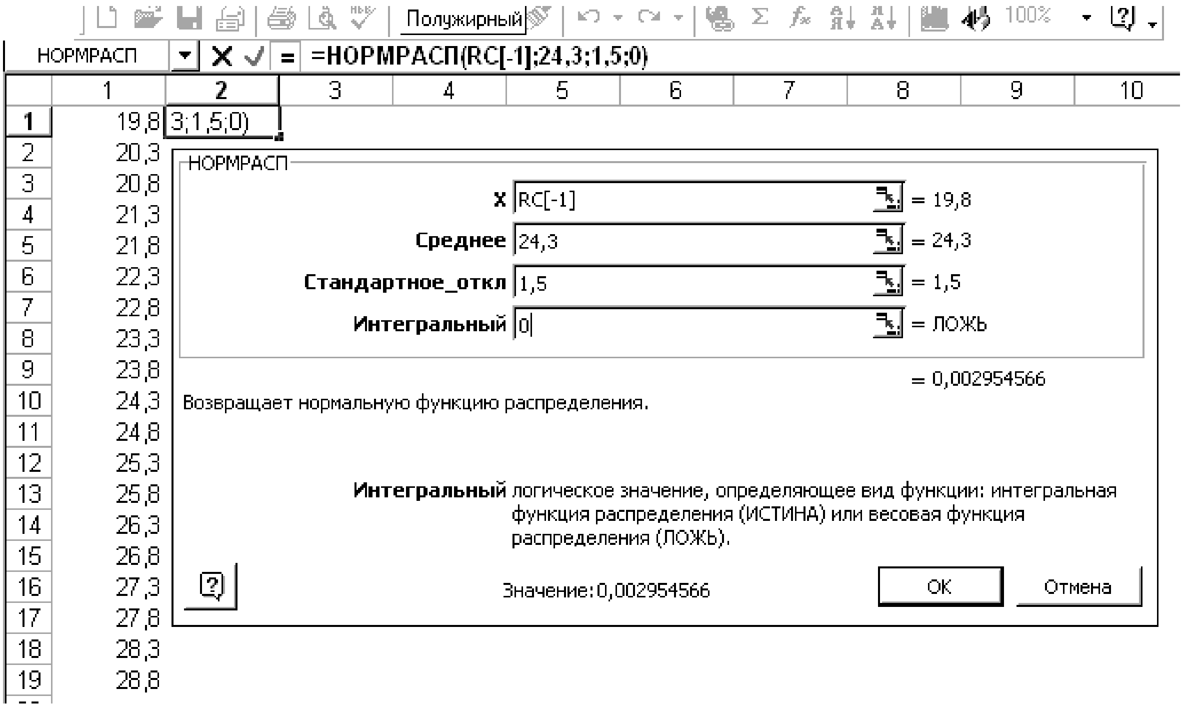
Построить диаграмму нормальной функции плотности вероятности f(x) при M = 24,3 и σ =1,5.

Решение.

1. В ячейку A1 вводим символ случайной величины x, а в ячейку B1 – символ функции плотности вероятности – f(x).
2. Вычисляем диапазон M $\pm$ 3σ - от 19,8 до 28,8. вводим в диапазон A2:A21 значения x от 19,8 до 28,8 с шагом 0,5. Для этого в ячейку A2 вводим левую границу диапазон (19,8), а затем Правка – Заполнить – Прогрессия – расположение по столбцам , шаг – 0,5, предельное значение – 28,8.
3. Устанавливаем табличный курсор в ячейку B2 и для получения значения вероятности воспользуемся специальной функцией – нажимаем на панели инструментов кнопку Вставка функции (f_x).

В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Статистическая. Справа в поле Функция выбираем функцию НОРМРАСП. Нажимаем на кнопку ОК.

Появляется диалоговое окно НОРМРАСП. В рабочее поле x вводим значение x, для которого строится распределение (в примере адрес ячейки A2 щелчком мыши на этой ячейке). В рабочее поле Среднее вводим с клавиатуры значение математического ожидания M (в примере – 24,3). В рабочее поле Стандартное_откл вводим с клавиатуры значение среднеквадратического отклонения σ (в примере –1,5). В рабочее поле Интегральный вводим с клавиатуры вид функции распределения – интегральная или весовая (в примере - 0). Нажимаем на кнопку ОК.

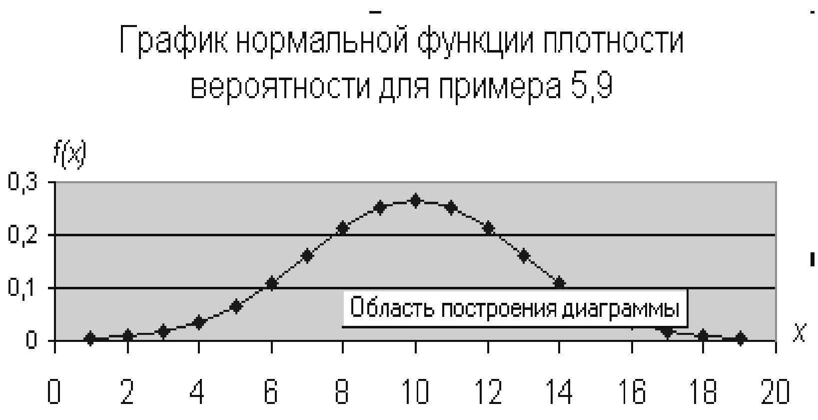


В ячейке В2 появляется вероятность р=0,002955.

Указателем мыши а правый нижний угол табличного курсора протягиваем (при нажатой левой кнопке мыши) яз ячейки В2 до В21 копируем функцию НОРМРАСП в диапазон В3:В21.

По полученным данным строим диаграмму нормальной функции распределения. Щелчком указателя мыши на кнопке на панели инструментов вызываем Мастер диаграмм. В появившемся диалоговом окне выбираем тип диаграммы График, вид – левый верхний. После нажатия кнопки Далее указываем диапазон данных – В1:В21 (с помощью мыши). Проверяем положение переключателя Ряды : в столбцах. Выбираем вкладку Ряд и с помощью мыши вводим диапазон подписей оси Х: А2:А21. нажав на кнопку Далее, вводим названия осей Х и Y: х и f(x), соответственно. Нажав на кнопку Готово.

Получен приближенный график нормальной функции плотности распределения:

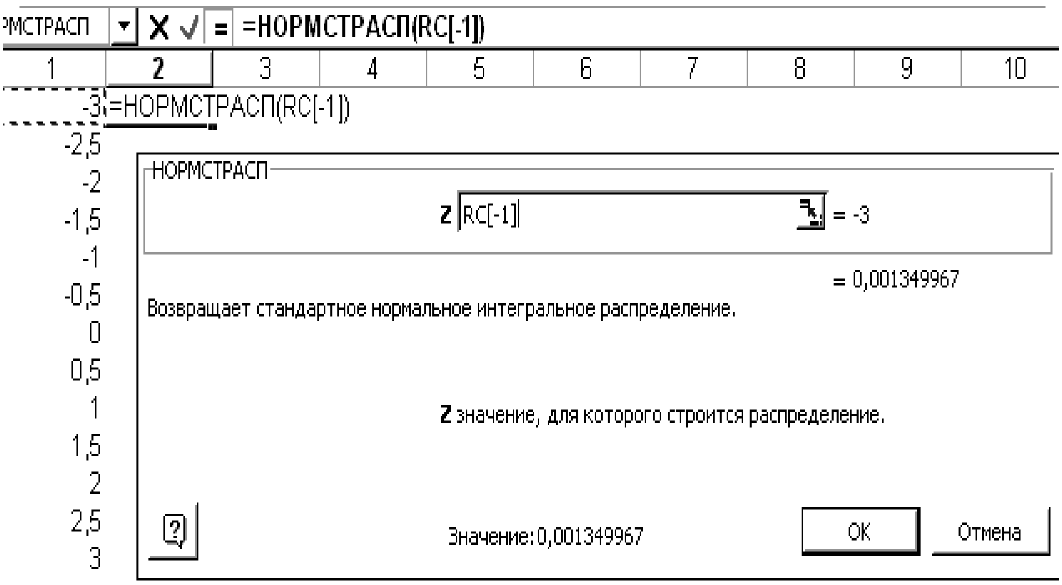


Пример.

Построить диаграмму стандартного нормального интегрального распределения.

Решение.

1. В ячейку А1 вводим символ случайной величины x, а в ячейку В1 – символ функции стандартного нормального распределения вероятности – Ф(x).
2. Вводим в диапазон А2:А14 значения x от –3 до 3 с шагом 0,5. для этого вводим в ячейку А2 вводим начальное значение –3. Затем Правка – Заполнить – Прогрессия – Расположение по столбцам – Шаг – 0,5 – Предельное значение 3 – Прогрессия арифметическая.
3. Устанавливаем табличный курсор в ячейку В2 и для получения значения вероятности воспользуемся специальной функцией – нажимаем на панели инструментов кнопку Вставка функции f(x).
4. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Статистическая. Справа в поле Функция выбираем функцию НОРМСТРАСП. Нажимаем на кнопку ОК.
5. Появляется диалоговое окно НОРМСТРАСП. В рабочее поле z вводим значение x для которого строится распределение (в примере адрес ячейки А2 щелчком мыши на этой ячейке). Нажимаем на кнопку ОК.



В ячейке В2 появляется вероятности р = 0,00135.

Указателем мыши за правый нижний угол табличного курсора протягиваем (при нажатой левой кнопке мыши) из ячейки В2 до В14 копируем функцию НОРМСТРАСП в диапазон В3:В14.

По полученным данным строим искомую диаграмму нормальной функции распределения. Щелчком указателя мыши на кнопке на панели инструментов вызываем Мастер диаграмм. В появившемся диалоговом окне выбираем тип диаграммы График, вид – левый верхний. После нажатия кнопки Далее указываем диапазон данных – В1:В14 (с помощью мыши). Проверяем положение переключателя Ряды: в столбцах. Выбираем вкладку Ряд и с помощью мыши вводим диапазон подписей оси Х: А2:А14. нажав на кнопку Далее, вводим названия осей Х и Y: х и Ф(х), соответственно. Нажимаем на кнопку Готово.

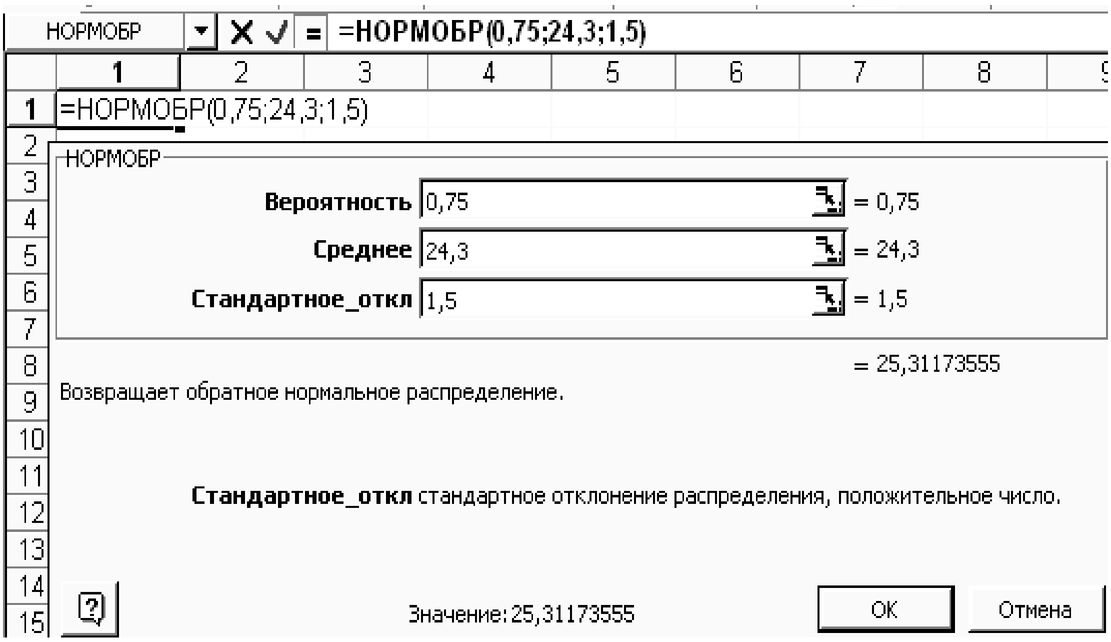
Получен график стандартизованной нормальной функции распределения.

Пример.

Найти верхнюю и нижнюю квартили для нормальной функции плотности вероятности f(x) при M = 24,3 и σ = 1,5.

Решение.

1. Устанавливаем табличный курсор в ячейку А1 и для получения значения верхней квартили воспользуемся специальной функцией: НОРМОБР – нажимаем на панели инструментов кнопку Вставка функции f(x).
2. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева Категория указаны виды функций. Выбираем Статистическая. Справа в поле Функция выбираем функцию НОРМОБР. Нажимаем на ОК.
3. Появляется диалоговое окно НОРМОБР. В рабочее поле Вероятность вводим значение вероятности верхней квартили – 0,75 (с клавиатуры). В рабочее поле Среднее вводим с клавиатуры значение математического ожидания M (в примере – 24,3). В рабочее поле Стандартное_откл вводим с клавиатуры значение среднеквадратического отклонения σ (в примере – 1,5). Нажимаем на ОК.
4. В ячейке А1 появляется значение верхней квартили 25,31174.



Устанавливаем табличный курсор в ячейку А2 и для получения значения нижней квартили повторяем пункты 1-4. За исключением того, что в п.3 в рабочее поле Вероятность вводим значение вероятности нижней квартили – 0,25.

В ячейке А2 появляется значение нижней квартили 23,28826.

Таким образом, верхняя и нижняя квартили для нормальной функции плотности вероятности f(x) при М = 24,3 и σ = 1,5 соответственно равны 25,31174 и 23,28826.

УПРАЖНЕНИЯ.

1. Найти вероятность того, что появится случайная величина x ≤ 42 при нормальном законе распределения вероятностей с М = 40 и σ - 1,5.
2. Построить диаграмму нормальной функции плотности вероятности f(x) при М = 10 и σ = 2.
3. Построить диаграмму интегральной нормальной функции распределения вероятности f(x) при М = 30 и σ = 3.
4. Найти квантиль для p = 0,908789 и нормального распределения с М = 40 и σ = 1,5.
5. Найти нормализованное значение x, если x = 21, М = 20, σ =1,5.

Другие законы распределения.

Распределение Пуассона – это дискретное распределение. Оно проявляется в ситуациях, когда в течение определенного отрезка времени или на определенном пространстве происходит случайное число каких-либо событий (число радиоактивных распадов, выпадение частиц аэрозоля, случаи заболеваний и т.п.).

В Excel для вычисления значений пуассоновского распределения используется функция ПУАССОН (*х; среднее; интегральный*). Она позволяет вычислить вероятность заданного числа появлений событий х при заданном значении среднего.

Показательное распределение.

Показательное распределение (экспоненциальное) – это непрерывное распределение. Обычно используется в задачах массового обслуживания и в задачах, связанных с "временем жизни". Так, например, распределены интервалы между обращениями граждан во многие организации.

В Excel для вычисления значений показательного распределения используется функция ЭКСПРАСП (х; лямбда; интегральный). Она позволяет вычислить вероятность заданного значения функции х при заданном значении параметра лямбда.

Распределение хи – квадрат.

Это непрерывное распределение, связанное с нормальным.

В Excel для вычисления значений распределения хи-квадрат используются функции ХИ2РАСП *(х; степени_свободы*) и ХИ2ОБР (*вероятность; степени_свободы*), которые при заданном числе степеней свободы вычисляют вероятность (по заданному значению c_n^2, для которого требуется вычислить распределение) и значение случайной величины c_n^2 (по заданной вероятности), соответственно.

Распределение Стьюдента.

Это также непрерывное распределение, связанное с нормальным.

В Excel вычисления значений распределения Стьюдента используются функции СТЬЮДРАСП (*х; степени свободы; хвосты*) и СТЬЮДРАСПОБР (*вероятность; степени_свободы*), которые при заданном числе степеней свободы вычисляют вероятность (по заданному численному значению x, для которого требуется вычислить распределение) и t – значение распределения Стьюдента (по заданной вероятности), соответственно.

F-распределение (Фишера).

В Ecxel для вычисления значений F – распределения используются функции FРАСП и FРАСПОБР (*x; степени_свободы1; степени_свободы2*), которые при заданном числе степеней свободы m и n вычисляют вероятность (по заданному численному значению $F_{m,\ n}$, для которого требуется вычислить распределение) и численное значение $F_{m,\ n}$ (по заданной вероятности), соответственно.

Другие распределения в Excel.

Excel позволяет также рассчитывать параметры еще нескольких распределений вероятностей. К ним относятся бета-распределение (функции БЕТАРАСП и БЕТАОБР), распределение Вейбулла (функция ВЕЙБУЛЛ), гамма-распределение (функции ГАММАРАСП и ГАММАОБР), гипергеометрическое распределение (функция ГИПЕРГЕОМЕТР), распределение Паскаля (функция ОТРБИНОМРАСП) и логнормальное распределение (функции ЛОГНОРМРАСП и ЛОГНОРМОБР).

Пример.

Найти вероятности появления значений меньших и равных 11,98 в экспоненциальном распределении со среднем значением 4.

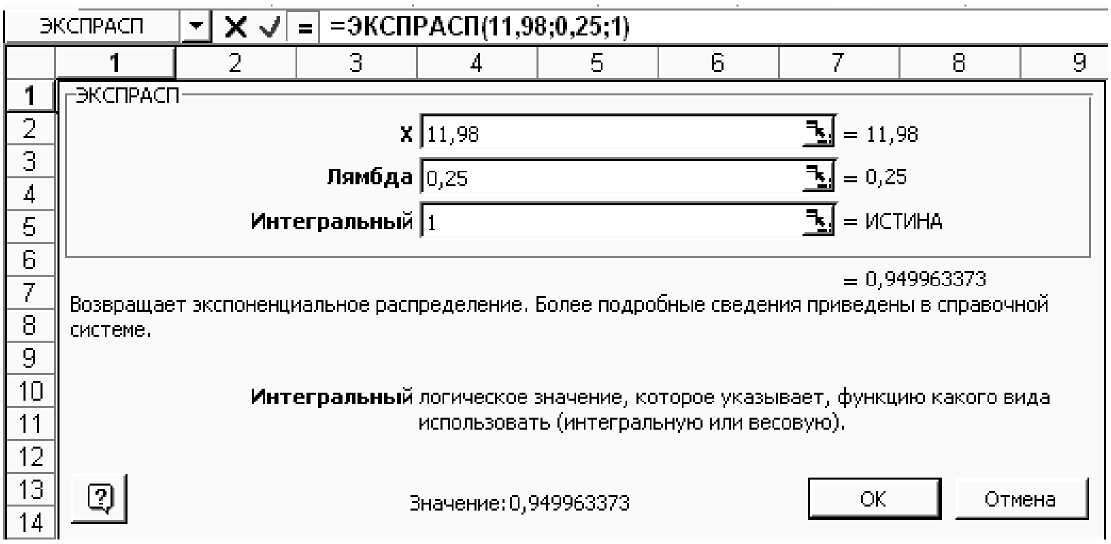
Решение.

1. Устанавливаем табличный курсор в ячейку А1 и для получения значений вероятностей экспоненциального распределения воспользуемся специальной функцией: ЭКСПРАСП – нажимаем на панели инструментов кнопку вставка функции (f_x).

2. В появившемся диалоговом окне Мастер функций – шаг 1 из 2 слева в поле Категория указаны виды функций. Выбираем Статистическая. Справа в поле Функция выбираем функцию ЭКСПРАСП. Нажимаем на кнопку ОК.

3. Появляется диалоговое окно ЭКСПРАСП. В рабочее поле x вводим численное значение функции – 11,98 (с клавиатуры). В рабочее поле Лямбда вводим с клавиатуры величину. Обратную

значению математического ожидания М (в примере – 0,25). В рабочее поле Интегральный вводим с клавиатуры значение ИСТИНА (в примере - 1). Нажимаем на кнопку ОК.



4. В ячейке А1 появляется значение вероятности – 0,949963.

Таким образом, вероятность того, что в экспоненциальном распределении со средним значением 4 появится число, меньше или равное 11,98 равно 0,95.

УПРАЖНЕНИЯ.

1. Найти р-квантали для распределения хи-квадрат с 14 степенями свободы для р=0,95; 0,975; 0,99.
2. Найти значение интегрального экспоненциального распределения с параметрами: случайная величина х=0,2 и математическое ожидание МХ = 0,1.
3. Найти значение случайной величины $F_{6.4}$ в F – распределении, появляющееся с вероятностью р = 0,01.
4. Найти вероятность появления случайной величины t = 1,96, распределенной по двустороннему закону Стьюдента с 60 степенями свободы.

Генерация случайных чисел.

Бывают ситуации, когда необходимо получить последовательность случайных чисел. Это требуется для моделирования объектов, имеющих случайную природу, по известному распределению вероятностей.

Можно воспользоваться процедурой Генерация случайных величин, которая является одним из инструментов пакета анализа. Эта процедура используется для заполнения диапазона случайными числами, извлеченными из одного или нескольких распределений.

Можно также воспользоваться функциями из категории Математические Мастера функций: СЛЧИС() (в результате ее выполнения на листе вычислений будет получено равномерно распределенное случайное число, большее или равное 0 и меньшее 1) и СЛУЧМЕЖДУ() (в результате будет получено случайное число, лежащее между произвольными заданными значениями).

В случае использования процедуры Генерация случайных чисел необходимо заполнить рабочие поля диалогового окна Генерация случайных величин.

В поле Число переменных вводится число столбцов значений, которые необходимо разместить в выходном диапазоне. Если это число не введено, то все столбцы в выходном диапазоне будут заполнены.

В поле Число случайных чисел вводится число случайных значений, которое необходимо вывести для каждой переменной. Каждое случайное значение будет помещено в строке выходного диапазона. Если число случайных чисел не будет введено, все строки выходного диапазона будут заполнены.

В поле Распределение необходимо выбрать тип распределения, которое следует использовать для генерации случайных переменных.

В их число входят:

- *равномерное* – характеризуется верхней и нижней границами. Переменные извлекаются с одной и той же вероятностью для всех значений интервала. Обычно приложения используют равномерное распределение в интервале 0 …1;
- *нормальное* – характеризуется средним значением и стандартным отклонением. Обычно приложения для этого распределения используют средне значение 0 и стандартное отклонение 1;
- *биноминальное* – характеризуется вероятностью успеха (величина p) для некоторого числа попыток. Например, можно сгенерировать случайные двухальтернативные переменные по

числу попыток, сумма которых будет биноминальной случайной переменой;

- *дискретное* - характеризуется значением и соответствующим ему интервалом вероятности. Диапазон должен состоять из двух столбцов: левого, содержащего значения, и правого, содержащего вероятности, связанные со значением в данной строке. Сумма вероятностей должна быть равна 1;
- *распределения Бернулли, Пуассона и Модельное.*

В рабочее поле Параметры вводятся параметры выбранного распределения.

В поле Случайное рассеивание вводится произвольное значение, для которого необходимо генерировать случайные числа.

В поле Выходной диапазон вводится ссылка на левую верхнюю ячейку выходного диапазона. Размер выходного диапазона будет определен автоматически, и на экран будет выведено сообщение в случае возможного наложения выходного диапазона на исходные данные.

Если необходимо получить случайные числа на новом листе или в новой книге – в полях Новый лист и Новая книга устанавливаются соответствующие переключатели.

Пример.

Повар столовой может готовить 7 различных первых блюд (уха, харчо, щи, борщ, бульон, грибной и гороховый супы). Необходимо составить меню на месяц, так чтобы первые блюда чередовались в случайном порядке.

Решение.

1. Пронумеруем первые блюда по порядку: 1 – уха; 2 – харчо; 3 – щи; 4 – борщ; 5 – бульон; 6 – грибной и 7 – гороховый. Введем числа с 1-7 в диапазон A1:A7 рабочей таблицы.
2. Укажем желаемую вероятность появления каждого первого блюда. Пусть все блюда будут равновероятными (p=1/7). Вводим число 1/7 в диапазон B1:B7. для этого в ячейку B1 вводим формулу =1/7 и затем копируем эту формулу в диапазон B2:B7 (автозаполнением).
3. В меню Сервис выбираем пункт Анализ данных и далее указываем строку Генерация случайных чисел. В появившемся диалоговом окне указываем Число переменных – 1, Число случайных чисел – 30 (количество дней в месяце). В поле Входной интервал значений и вероятностей вводим (мышью) диапазон, содержащий номера супов и их вероятности – диапазон A1:B7.

указываем выходной диапазон. Для этого устанавливаем флажок в левое поле Выходной диапазон и в правое поле ввода Выходной диапазон вводим, например, С1. нажимаем на кнопку ОК.

В столбце С появляются случайные числа, например: 3, 1, 5, 7, 7, 7, 1, 3, …и т д.

Полученные результаты означают, что в первый день следует готовить щи, во второй – уху и т.д.

Пример.

Пусть спортсмену необходимо составить график тренировок на 10 дней, так чтобы дистанция, пробегаемая каждый день, случайным образом менялась от 5 до 10 км.

Решение.

1. В меню Сервис выбираем пункт Анализ данных и далее указываем строку Генерация случайных чисел. В появившемся диалоговом окне указываем Число переменных – 1, Число случайных чисел – 10 (требуемое количество дней). В поле Распределение указываем *Равномерное* (действительные числа, появляющиеся с равной вероятностью). В поле Параметры *Между* вводим (с клавиатуры) значение – 5, в поле *и* –10. в поле Случайное рассеивание – любое число. Указываем выходной диапазон. Для этого устанавливаем флажок в левое поле Выходной диапазон и в правое поле ввода Выходной диапазон вводим, например, А1. нажимаем на кнопку ОК.
2. В столбце А появляются 10 случайных чисел, например: 7,759331; 6,807917; 9,960326; 6,042055; 9,380474; 7,511521; 7,486496; 5,538499; 6,706595; 6,963561.

Эти числа обозначают расстояние в километрах, которое нужно пробегать спортсмену в соответствующий день.

УПРАЖНЕНИЯ.

1. Создать последовательность, состоящую из 10 действительных случайных чисел, равномерно распределенных в диапазоне от 3 до 7.
2. Создать последовательность из 10 целых случайных чисел из диапазона 2, 3, 4, 5, с равной вероятностью появления.
3. Составить расписание на месяц для случайной демонстрации на телевидении одного из четырех рекламных роликов турфирмы. Причем вероятность появления рекламного ролика № 1 должна быть в 2 раза выше, чем остальных рекламных роликов.

6. Статистика

Дадим определение понятия математической статистики, методов статистического исследования, а также рассмотрим выборочные функции распределения и основные выборочные характеристики.

Понятие математической статистики

Раздел математики, посвященный методам сбора, анализа и обработки статистических данных для научных и практических целей, называется математической статистикой.

Основные понятия и определения

Математическая статистика имеет дело с массовыми явлениями. Она тесно связана с теорией вероятностей и базируется на ее математическом аппарате.

Целью статистического исследования является обнаружение и исследование соотношений между статистическими данными и их использование для изучения, прогнозирования и принятия решений.

Статистические данные представляют собой данные, полученные в результате обследования большого числа объектов или явлений.

Математическая статистика подразделяется на две основные области: *описательную и аналитическую статистику*. Описательная статистика охватывает методы описания статистических данных, представления их в форме таблиц, распределений и т. п.

Аналитическая статистика или теория статистических выводов ориентирована на обработку данных, полученных в ходе эксперимента, с целью формулировки выводов, имеющих прикладное значение для самых различных областей человеческой деятельности.

Выборочный метод

Множество всех единиц наблюдения, охватываемых таким сплошным наблюдением, называется *генеральной совокупностью*.

Основным методом не сплошного наблюдения является *выборочный метод*.

Выборка — это группа элементов, выбранная для исследования из всей совокупности элементов. Задача выборочного метода состоит в том, чтобы сделать правильные выводы относительно всего собрания объектов, их совокупности. Например, пробуя пищу, повар по одной ложке делает заключения о качестве приготавливаемого во всей кастрюле.

Конечной целью изучения выборочной совокупности всегда является получение информации о генеральной совокупности.

Выборочная функция распределения

Рассмотренные в разделе «Числовые характеристики распределения вероятностей» характеристики случайной величины опираются на знание закона ее распределения $F(x)$.. Здесь закон распределения обычно неизвестен, или известен с точностью до некоторых неизвестных параметров. В частности, невозможно рассчитать точное значение соответствующих вероятностей, так как нельзя определить количество общих и благоприятных исходов. Поэтому вводится статистическое определение вероятности. По этому определению вероятность равна отношению числа испытаний (m), в которых событие появилось, к общему количеству произведенных испытаний (n). Такая вероятность называется *статистической частотой*.

В результате на практике сведения о законе распределения случайной величины получают *независимыми многократными повторениями опыта*, в котором измеряются значения интересующей исследователей случайной величины (варианты). На основе информации из полученной выборки можно построить приблизительные значения для функции распределения и других характеристик случайной величины.

Выборочной (*эмпирической*) функцией распределения случайной величины ζ, построенной по выборке $x_1, x_2, ..., x_n$ называется функция $F_n(x)$, равная доле таких значений x_i, что $x_i \le x, i = 1,..., n$.

Другими словами, $F_n(x)$ есть частота события $x_i < x$ в ряду $x_1, x_2, ..., x_n$

Связь между эмпирической функцией распределения и функцией распределения (теоретической функцией распределения) такая же, как связь между частотой события и его вероятностью: функция $F_n(x)$ —> $F(x)$ при n —> ∞.

Для построения выборочной функции распределения весь диапазон изменения случайной величины *X* разбивают на ряд интервалов одинаковой ширины. Число интервалов обычно выбирают не менее 5 и не более 15. Затем определяют число значений случайной величины *X*, попавших в каждый интервал. Поделив эти числа на общее количество наблюдений *n*, находят относительную частоту попадания случайной величины *X* в заданные интервалы. По найденным относительным частотам строят гистограммы выборочных функций распределения. Если соответствующие точки относительных частот соединить ломаной линией, то полученная диаграмма будет называться *полигоном частот*. Кумулятивная кривая будет получена, если по оси абсцисс откладывать интервалы, а по оси ординат — число или доли элементов совокупности, имеющих значение, меньшее или равное заданному.

При увеличении до бесконечности размера выборки выборочные функции распределения превращаются в теоретические: гистограмма превращается в график плотности распределения, а кумулятивная кривая — в график функции распределения.

В Excel для построения выборочных функций распределения используются специальная функция ЧАСТОТА и процедура пакета анализа Гистограмма.

-- Функция ЧАСТОТА вычисляет частоты появления случайной величины в интервалах значений и выводит их как массив цифр. Функция задается в качестве формулы массива *ЧАСТОТА(массив_данных;массив_карманов)*. Здесь:

- *массив_данных* — это массив или ссылка на множество данных, для которых вычисляются частоты.
- *массив_карманов* — это массив или ссылка на-множество интервалов, в которые группируются значения аргумента *массив__данных*.

Отметим, что количество элементов в возвращаемом массиве на единицу больше числа элементов в *массив_карманов*. Дополнительный элемент в возвращаемом массиве содержит количество значений, больших, чем максимальное значение в интервалах.

-- Процедура Гистограмма используется для вычисления выборочных и интегральных частот попадания данных в указанные интервалы значений. Процедура .выводит результаты в виде таблицы и гистограммы.

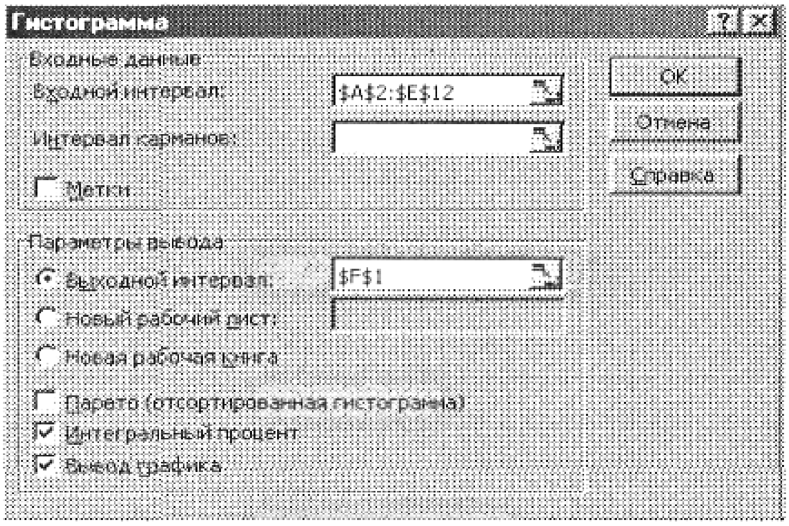


Рис. 6.1. Пример заполнения диалогового окна Гистограмма

Параметры диалогового окна Гистограмма представлены на рис. 6.1:

- во Входной диапазон вводится диапазон исследуемых данных;
- в поле Интервал карманов (необязательный параметр) может вводиться диапазон ячеек или необязательный набор граничных значений, определяющих выбранные интервалы (карманы). Эти значения должны быть введены в возрастающем порядке. В MS Excel вычисляется число попаданий данных между началом интервала и соседним большим по порядку. При этом включаются значения на нижней границе интервала и не включаются значения на верхней границе. Если диапазон карманов не был введен, то набор интервалов, равномерно распределенных между минимальным и максимальным значениями данных, будет создан автоматически;
- рабочее поле Выходной диапазон предназначено для ввода ссылки на левую верхнюю ячейку выходного диапазона. Размер выходного диапазона будет определен автоматически;

- переключатель Интегральный процент позволяет установить режим генерации интегральных процентных отношений и включения в гистограмму графика интегральных процентов;
- переключатель Вывод графика позволяет установить режим автоматического создания встроенной диаграммы на листе, содержащем выходной диапазон.

Пример. Построить эмпирическое распределение веса студентов в килограммах для следующей выборки: 64,57,63,62,58,61,63, 60, 60, 61,65, 62, 62, 60,64, 61, 59, 59, 63, 61, 62, 58,58, 63,61, 59, 62,60, 60, 58, 61, 60, 63, 63, 58,60, 59, 60, 59,61, 62, 62, 63, 57, 61, 58, 60, 64, 60, 59, 61, 64, 62, 59, 65.

Решение.

1. В ячейку A1 введите слово *Наблюдения,* а в диапазон A2:E12 — значения веса студентов.

2. Выберите ширину интервала 1 кг. Тогда при крайних значениях веса 57 кг и 65 кг получится 9 интервалов. В ячейки G1 и G2 введите названия интервалов *Вес* и *кг,* соответственно. В диапазон G4:G12 введите граничные значения интервалов ***(57,58,59, 60, 61, 62, 63, 64, 65).***

3. Введите заголовки создаваемой таблицы: в ячейки H1:H2 — *Абсолютные частоты,* в ячейки I1:I2 — *Относительные частоты,* в ячейки J1:J2 — *Накопленные частоты.*

4. Заполните столбец абсолютных частот. Для этого выделите для них блок ячеек H4:H12 (используемая функция ЧАСТОТА задается в виде формулы массива). С панели инструментов Стандартная вызовите Мастер функций (кнопка fx). В появившемся диалоговом окне Мастер функций Выберите категорию Статистические и функцию ЧАСТОТА, после чего нажмите кнопку ОК. Появившееся диалоговое окно ЧАСТОТА необходимо за серое поле мышью отодвинуть вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши в рабочее поле Массив_ данных введите диапазон данных наблюдений *(A2:E12).* В рабочее поле Двоичный массив мышью введите диапазон интервалов (G4:G12). Последовательно нажмите комбинацию клавиш Ctrl+Shift+ Enter. В столбце H4:H12 появится массив абсолютных частот.

5. В ячейке H13 найдите общее количество наблюдений. Табличный курсор установите в ячейку H13. На панели инструментов Стандартная нажмите кнопку Автосумма. Убедитесь, что диапазон суммирования указан правильно *(H4:H12)*, и нажмите клавишу Enter. В ячейке H13 появится число *55*.

6. Заполните столбец относительных частот. В ячейку I4 введите формулу для вычисления относительной частоты: *=H4/H$13*. Нажмите клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) скопируйте введенную формулу в диапазон 15:112. Получим массив относительных частот.

7. Заполните столбец накопленных частот. В ячейку J4 скопируйте значение относительной частоты из ячейки I4 *(0,036364)*. В ячейку J5 введите формулу: *=J4 + J5*. Нажмите клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) скопируйте введенную формулу в диапазон J6:J12, Получим массив накопленных частот.

8. В результате после форматирования получим таблицу, представленную на рис. 6.2.

	A	B	C	D	E	F	G	H	I	J
1	Наблюдения						Вес	Абсолютные	Относи-	Накопленные
2	64	62	58	63	61		кг	частоты	тельные	частоты
3	57	62	63	58	58				частоты	
4	63	60	61	60	60		57	2	0,036	0,036
5	62	64	59	59	64		58	6	0,109	0,145
6	58	61	62	60	60		59	7	0,127	0,273
7	61	59	60	59	59		60	10	0,182	0,455
8	63	59	60	61	61		61	9	0,164	0,618
9	60	63	58	62	64		62	8	0,145	0,764
10	60	61	61	62	62		63	7	0,127	0,891
11	61	62	60	63	59		64	4	0,073	0,964
12	65	58	63	57	65		65	2	0,036	1,000
13								55		

Рис. 6.2. Результат вычислений относительных и накопленных частот из примера

9. Постройте диаграмму относительных и накопленных частот. Щелчком указателя мыши по кнопке на панели инструментов вызовите Мастер диаграмм. В появившемся диалоговом окне выберите вкладку Нестандартные и тип диаграммы График/гистограмма2. После нажатия кнопки Далее укажите диапазон данных — I1:J12 (с помощью мыши). Проверьте положение переключателя Ряды в: столбцах. Выберите вкладку Ряд и с по-

мощью мыши введите в рабочее поле Подписи оси X диапазон подписей оси X: *G4:G12.* Нажав кнопку Далее, введите названия осей X и У: в рабочее поле Ось X (категорий) — *Вес;* Ось Y (значений) — *Относ.частота;* Вторая ось Y (значений) — *Накоплен.частота.* Нажмите кнопку Готово.

После минимального редактирования диаграмма будет иметь такой вид, как на рис. 6.3.

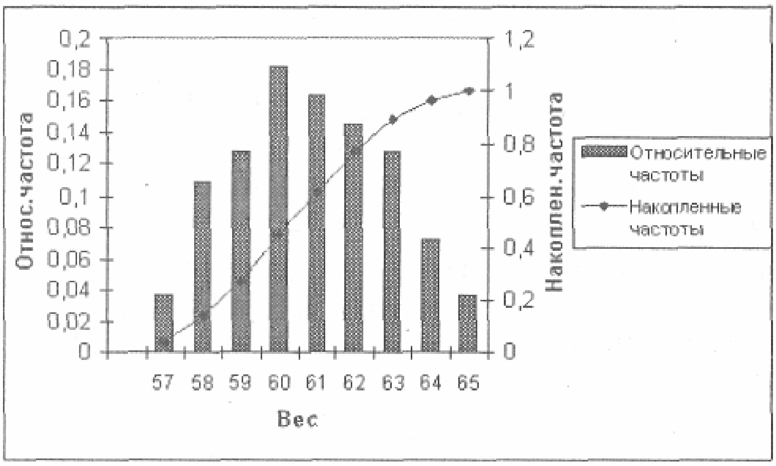


Рис. 6.3. Диаграмма относительных и накопленных частот из примера

Пример. Для данных из предыдущего примера построить эмпирические распределения, воспользовавшись процедурой Гистограмма.

Решение.

1. В ячейку A1 введите слово *Наблюдения,* а в диапазон A2:E12 — значения веса студентов.
2. Для вызова процедуры Гистограмма выберите из меню Сервис подпункт Анализ данных и в открывшемся окне в поле Инструменты анализа укажите процедуру Гистограмма.
3. В появившемся окне Гистограмма заполните рабочие поля (см. рис. 6.1):

- во Входной диапазон введите диапазон исследуемых данных *(A2:E12);*
- в Выходной диапазон — ссылку на левую верхнюю ячейку выходного диапазона *(F1).* Установите переключатели в положение Интегральный процент и

Вывод графика;

После этого нажмите кнопку ОК.]

В результате появляется таблица и диаграмма (рис. 6.4).

Карман	Частота	Интегральный %
57	2	3,64%
59,66667	13	27,27%
62,33333	27	76,36%
Еще	13	100,00%

Рис. 6.4. Таблица и диаграмма из примера

Как видно, диаграмма на рис. 6.4. несколько отличается от диаграммы на рис. 6.3. Это объясняется тем, что диапазон карманов не был введен. Количество и границы интервалов определялись в процедуре ГИСТОГРАММА автоматически. Если бы в рабочее поле Интервал карманов был бы введен диапазон ячеек, определяющих выбранные интервалы, как в примере 6.1 *(57,58,59,..., 65)*, то полученная диаграмма была бы идентична предыдущей.

УПРАЖНЕНИЯ.

1. Постройте эмпирические функции распределения (относительные и накопленные частоты) для роста (в см) группы из 20 мужчин: 181,169,178, 178,171,179, 172, 181, 179, 168, 174, 167, 169, 171, 179, 181, 181, 183, 172, 176.

2. Найдите распределение по абсолютным частотам для следующих результатов тестирования в баллах: 79, 85, 78, 85, 83, 81, 95, 88 и 97 (используйте границы интервалов 70, 79, 89).

3. Постройте эмпирические функции распределения (абсолютные и накопленные частоты) успеваемости в группе из 20 студентов: 4, 4, 5, 3, 4, 5,4, 5, 3, 5, 3, 3, 5,4, 5,4,3,5,3,5.

Выборочные характеристики

Замена теоретической функции распределения $F(x)$ на ее выборочный аналог $F_n(x)$ в определении математического ожидания, дисперсии, стандартного отклонения и т. п. приводят к выборочному среднему, выборочной дисперсии, выборочному стандартному отклонению и т.д. Выборочные характеристики являются оценками соответствующих характеристик генеральной совокупности. Эти оценки должны удовлетворять определенным требованиям. В соответствии с важнейшими требованиями, оценки должны быть:

- *несмещенными*, то есть стремиться к истинному значению характеристики генеральной совокупности при неограниченном увеличении количества испытаний;

- *состоятельными*, то есть с ростом размера выборки оценка должна стремиться к значению соответствующего параметра генеральной совокупности с вероятностью, приближающейся к 1;

- *эффективными*, то есть для выборок равного объема используемая оценка должна иметь минимальную дисперсию.

Среди выборочных характеристик выделяют показатели, относящиеся к центру распределения (меры положения), показатели рассеяния вариант (меры рассеяния) и меры формы распределения. К показателям, характеризующим центр распределения, относят различные виды средних (арифметическое, геометрическое и т. п.), а также моду и медиану.

Простейшим показателем, характеризующим центр выборки, является мода.

Мода — это элемент выборки с наиболее часто встречающимся значением (наиболее вероятная величина).

Средним значением выборки, или выборочным аналогом математического ожидания, называется величина

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

Иначе говоря, среднее значение — это центр выборки, вокруг которого группируются элементы выборки. При увеличении числа наблюдений среднее приближается к математическому ожиданию. Среднее значение обозначается также буквой *М*.

Выборочная медиана — это число, которое является серединой выборки, то есть половина чисел имеет значения большие, чем медиана, а половина чисел имеет значения меньшие, чем медиана. Для нахождения медианы обычно выборку ранжируют — располагают элементы в порядке возрастания. Если количество членов ранжированного ряда нечетное, медианой является значение ряда, которое расположено посередине, то есть элемент с номером $(n + 1)/2$. Если число членов ряда четное, то медиана равна среднему членов ряда с номерами $n/2$ и и/2 + 1 .

Основными показатели рассеяния вариант являются интервал, дисперсия выборки, стандартное отклонение и стандартная ошибка.

Интервал (амплитуда, вариационный размах) — это разница между максимальным и минимальным значениями элементов выборки. Интервал является простейшей и наименее надежной мерой вариации или рассеяния элементов в выборке.

Более точно отражают рассеяние показатели, учитывающие не только крайние, но и все значения элементов выборки.

Дисперсией выборки, или выборочным аналогом дисперсии, называется величина

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

Дисперсия выборки — это параметр, характеризующий степень разброса элементов выборки относительно среднего значения. Чем больше дисперсия, тем дальше отклоняются значения элементов выборки от среднего значения.

Выборочным стандартным отклонением (среднее квадратичное отклонение) называется величина

$$s = \sqrt{s^2}$$

Это параметр, также характеризующий степень разброса элементов выборки относительно среднего значения. Чем больше среднее квадратичное отклонение, тем дальше отклоняются значения элементов выборки от среднего значения. Параметр аналогичен дисперсии и используется в тех случаях, когда необходимо, чтобы показатель разброса случайной величины выражался в тех же единицах, что и среднее значение этой случайной величины. Часто выборочное стандартное отклонение обозначают буквой а (сигма).

Стандартная ошибка или ошибка среднего находится из выражения

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

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

Выборочной квантилью называется решение уравнения

$$F_n(x) = p$$

В частности, выборочная медиана есть решение уравнения

$$F_n(x) = 0{,}5$$

Показателями, характеризующими форму распределения, являются выборочные эксцесс и асимметрия.

Эксцесс — это степень выраженности «хвостов» распределения, то есть частоты появления удаленных от среднего значений.

Асимметрия — величина, характеризующая несимметричность распределения элементов выборки относительно среднего значения. Принимает значения от -1 до 1. В случае симметричного распределения асимметрия равна 0.

Часто значения асимметрии и эксцесса используют для проверки гипотезы о том, что данные (выборка) принадлежат к определенному теоретическому распределению, в частности, нормальному распределению.

Определение основных статистических характеристик

Для нормального распределения асимметрия равна нулю, а эксцесс — трем.

В результате наблюдений или эксперимента получаются наборы данных, называемые выборками. Для проведения их анализа данные подвергаются статистической обработке. Первое, что всегда делается при обработке данных, это вычисление элементарных статистических характеристик выборок (как минимум: среднего, среднеквадратичного отклонения, ошибки среднего) по каждому параметру и по каждой группе. Полезно также вычислить эти характеристики для объединения родственных групп и суммарно по всем данным.

Использование специальных функций

В мастере функций Excel имеется ряд специальных функций, предназначенных для вычисления выборочных характеристик – это функции, характеризующие центр распределения.

- Функция СРЗНАЧ вычисляет среднее арифметическое из нескольких массивов (аргументов) чисел. Аргументы *число1, число2, ...* — это от 1 до 30 массивов, для которых вычисляется среднее. Например, если ячейки A1:A7 содержат числа 10, 14, 5, 6, 10, 12 и 13, то средним арифметическим СРЗНАЧ(A1:A7) является 10 (рис. 6.5.).

- Функция СРГАРМ позволяет получить среднее гармоническое множества данных. Среднее гармоническое — это величина, обратная к среднему арифметическому обратных величин. Например, СРГАРМ *(10;14;5;6;10;12;13)* равняется 8,317.

- Функция СРГЕОМ вычисляет среднее геометрическое значений массива положительных чисел. Функцию СРГЕОМ можно использовать для вычисления средних показателей динамического ряда. Например, СРГЕОН *(10;14;5;6;10;12;13)* равняется 9,414.

- Функция МЕДИАНА позволяет получать медиану заданной выборки. Медиана — это элемент выборки, число элементов выборки со значениями больше которого и меньше которого равно. Например, МЕДИАНА *(10;14;5;6;10;12;13)* равняется 10.

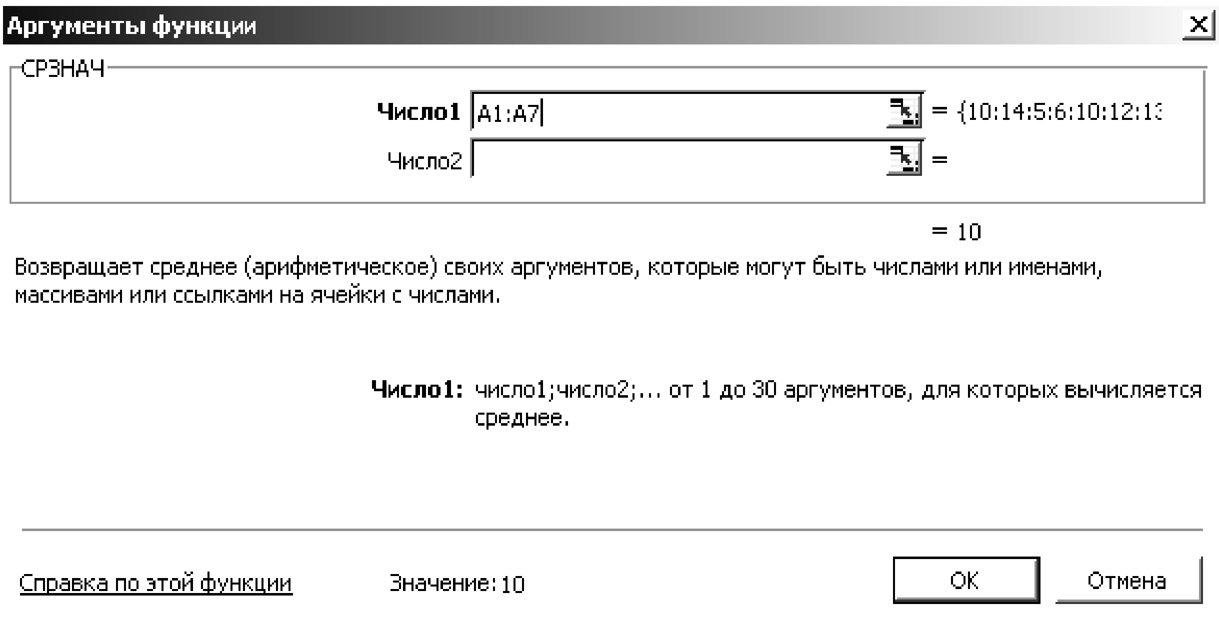


Рис. 6.5. Диалоговое окно функции СРЗНАЧ

- Функция МОДА вычисляет наиболее часто встречающееся значение в выборке.

Например, МОДА *(10;14;5;6;10;12;13)* равняется 10.

К специальным функциям, вычисляющим выборочные характеристики, характеризующие рассеяние вариант, относятся ДИСП, СТАНДОТКЛОН, ПЕРСЕНТИЛЬ.

- Функция ДИСП позволяет оценить дисперсию по выборочным данным. Например, ДИСП *(10;14;5;6;10;12;13)* равняется 11,667.

- Функция СТАНДОТКЛОН вычисляет стандартное отклонение. Например, СТАНДОТКЛОН*(10;14;5;6;10;12;13)* равняется 3,416.

- Функция ПЕРСЕНТИЛЬ позволяет получить квантили заданной выборки. Например, если ячейки А1:А7 содержат числа 10, 14, 5, 6, 10, 12 и 13, то квантилью со значением 0,1 является ПЕРСЕНТИЛЬ (А1:А7;0,1), равная 5,6.

Форму эмпирического распределения позволяют оценить специальные функции ЭКСЦЕСС и СКОС.

- Функция ЭКСЦЕСС вычисляет оценку эксцесса по выборочным данным. Например, ЭКСЦЕСС (*10;14;5;6;10;12;13)* равняется -1,169.

- Функция СКОС позволяет оценить асимметрию выборочного распределения. Например, *СКОС (10;14;5;6;10;12;13)* равняется -0,527.

Пример. Рассматриваются ежемесячные количества реализованных турфирмой путевок за периоды до и после начала активной рекламной компании. Ниже приведены количества реализованных путевок по месяцам.

С рекламой	Без рекламы
162	135
156	126
144	115
137	140
125	121
145	112
151	130

Требуется найти средние значения и стандартные отклонения этих данных.

Решение.

1. Для проведения статистического анализа прежде всего необходимо ввести данные в рабочую таблицу. Откройте новую рабочую таблицу. Введите в ячейку А1 слово *Реклама,* затем в ячейки А2:А8 — соответствующие значения числа реализованных путевок. В ячейку В1 введите *слова Без рекламы,* а в В2:В8 — значения числа реализованных путевок до начала рекламной компании. Отметим, что рассматриваемые группы данных со статистической точки зрения являются выборками.

2. При статистическом анализе прежде всего необходимо определить характеристики выборки, и важнейшей характеристикой является среднее значение. Для определения среднего значения в контрольной группе необходимо установить табличный курсор в свободную ячейку (А9). На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию СРЗНАЧ, после чего нажмите кнопку ОК. Появившееся диалоговое окно СРЗНАЧ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных контрольной группы для определения среднего значения (А2:А8). Нажмите кнопку ОК. В ячейке А9 появится среднее значение выборки — *145,714.*

В качестве упражнения определите в ячейке В9 среднее значение числа реализованных путевок без активной рекламы. Для этого табличный курсор установите в ячейку В9. На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне выберите категорию Статистические и функцию СРЗНАЧ, после чего нажмите кнопку ОК. Появившееся диалоговое окно СРЗНАЧ за серое поле мышью отодвиньте вправо на 1 - 2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных для определения среднего значения *(В2:В8).* Нажмите кнопку ОК. В ячейке В9 появится среднее значение выборки — *125,571*.

3. Следующей по важности характеристикой выборки является мера разброса элементов выборки от среднего значения. Такой мерой является среднее квадратичное или стандартное отклонение. Для определения стандартного отклонения в контрольной группе необходимо установить табличный курсор в сво-

бодную ячейку (АЮ). На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию СТАНДОТКЛОН, после чего нажмите кнопку ОК. Появившееся диалоговое окно СТАНДОТКЛОН за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных контрольной группы для определения стандартного отклонения *(A2:A8).* Нажмите кнопку ОК. В ячейке АЮ появится стандартное отклонение выборки — 12,298. Существует правило, согласно которому при отсутствии артефактов данные должны лежать в диапазоне М ± 3σ (в примере 145,7±36,9).

В качестве упражнения требуется в ячейке В10 определить стандартное отклонение числа проданных путевок до начала рекламной компании. Для этого установите табличный курсор в ячейку В10. На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне выберите категорию Статистические и функцию СТАНДОТКЛОН, после чего нажмите кнопку ОК. Появившееся диалоговое окно СТАНДОТКЛОН за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных для определения стандартного отклонения *(B2:B8).* Нажмите кнопку ОК. В ячейке В10 появится стандартное отклонение выборки — *10,277.*

УПРАЖНЕНИЯ.

1. Найдите среднее значение и стандартное отклонение результатов бега на дистанцию 100 м у группы студентов: 12,8; 13,2; 13,0; 12,9; 13,5; 13,1.
2. Найдите выборочные среднее, медиану, моду, дисперсию и стандартное отклонение для следующей выборки 26, 35, 29, 27, 33, 35, 30, 33, 31, 29.
3. Определите верхнюю (0,75) и нижнюю (0,25) квартили для выборки результатов измерений роста группы студенток: 164,160,157,166,162,160,161,159,160, 163, 170, 171.
4. Определите выборочные асимметрию и эксцесс для данных измерений роста из упражнения 6.

Использование инструментов Пакета анализа

В пакете Excel помимо мастера функций имеется набор более мощных инструментов для работы с несколькими выборками и углубленного анализа данных, называемый Пакет анализа, который может быть использован для решения задач статистической обработки выборочных данных.

Для установки раздела Анализ данных в пакете Excel сделайте следующее:

- в меню Сервис выберите команду Надстройки;
- в появившемся списке установите флажок Пакет анализа.

Ввод данных. Исследуемые данные следует представить в виде таблицы, где столбцами являются соответствующие показатели. При создании таблицы Excel информация вводится в отдельные ячейки. Совокупность ячеек, содержащих анализируемые данные, называется входным диапазоном.

Последовательность обработки данных. Для использования статистического пакета анализа данных необходимо:

- указать курсором мыши на пункт меню Сервис и щелкнуть левой кнопкой мыши;
- в раскрывающемся списке выбрать команду Анализ данных (если команда Анализ данных отсутствует в меню Сервис, то необходимо установить в Excel пакет анализа данных);
- выбрать необходимую строку в появившемся списке Инструменты анализа;
- ввести входной и выходной диапазоны и выбрать необходимые параметры.

Нахождение основных выборочных характеристик. Для определения характеристик выборки используется процедура Описательная статистика. Процедура позволяет получить статистический отчет, содержащий информацию о центральной тенденции и изменчивости входных данных. Для выполнения процедуры необходимо:

- выполнить команду Сервис > Анализ данных;
- в появившемся списке Инструменты анализа выбрать строку Описательная статистика и нажать кнопку ОК (рис. 6.6);

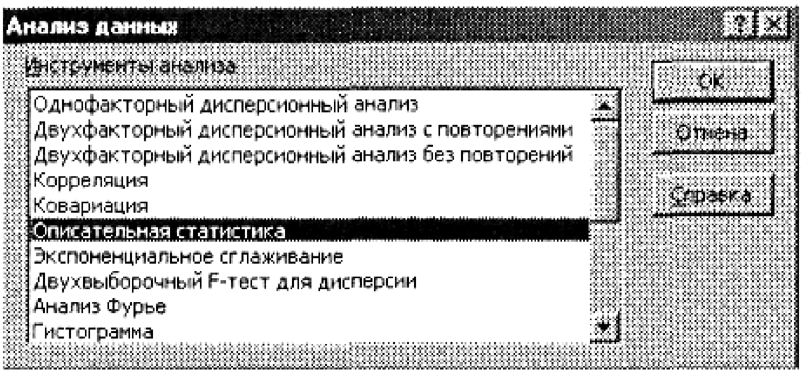


Рис. 6.6. Окно выбора метода обработки данных

- в появившемся диалоговом окне указать входной диапазон, то есть ввести ссылку на ячейки, содержащие анализируемые данные. Для этого следует навести указатель мыши на левую верхнюю ячейку данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к правой нижней ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши;
- указать выходной диапазон, то есть ввести ссылку на ячейки, в которые будут выведены результаты анализа. Для этого следует поставить переключатель в положение Выходной диапазон (навести указатель мыши и щелкнуть левой клавишей), далее навести указатель мыши в поле ввода Выходной диапазон и щелкнуть левой кнопкой мыши, затем указатель мыши навести на левую верхнюю ячейку выходного диапазона и щелкнуть левой кнопкой мыши;
- в разделе Группировка переключатель установить в положение по столбцам; о установить флажок в поле Итоговая статистика;
- нажать кнопку ОК.

В результате анализа в указанном выходном диапазоне для каждого столбца данных выводятся следующие статистические характеристики: среднее, стандартная ошибка (среднего), медиана, мода, стандартное отклонение, дисперсия выборки, эксцесс, асимметричность, интервал, минимум, максимум, сумма, счет, наибольшее, наименьшее, уровень надежности.

Пример. Рассматривается зарплата основных групп работников гостиницы: администрации, обслуживающего персонала и работников ресторана. Были получены следующие данные:

Администрация	Персонал	Ресторан
4500	2100	3200
4000	2100	3000
3700	2000	2500
3000	2000	2000
2500	2000	1900
	1900	1800
	1800	
	1800	

Необходимо определить основные статистические характеристики в группах данных.

Решение.

1. Для использования инструментов анализа исследуемые данные следует представить в виде таблицы, где столбцами являются соответствующие показатели. Значения зарплат сотрудников администрации введите в диапазон A1:A5, обслуживающего персонала — в диапазон B1:B8 и т. д. В результате получится таблица, представленная на рис. *6.7*.

	A	B	C
1	4500	2100	3200
2	4000	2100	3000
3	3700	2000	2500
4	3000	2000	2000
5	2500	2000	1900
6		1900	1800
7		1800	
8		1800	

Рис. 6.7. Таблица из примера

2. Далее необходимо провести элементарную статистическую обработку. Для этого, указав курсором мыши на пункт меню Сервис, выберите команду Анализ данных. Затем в появившемся списке Инструменты анализа выберите строку Описательная статистика.

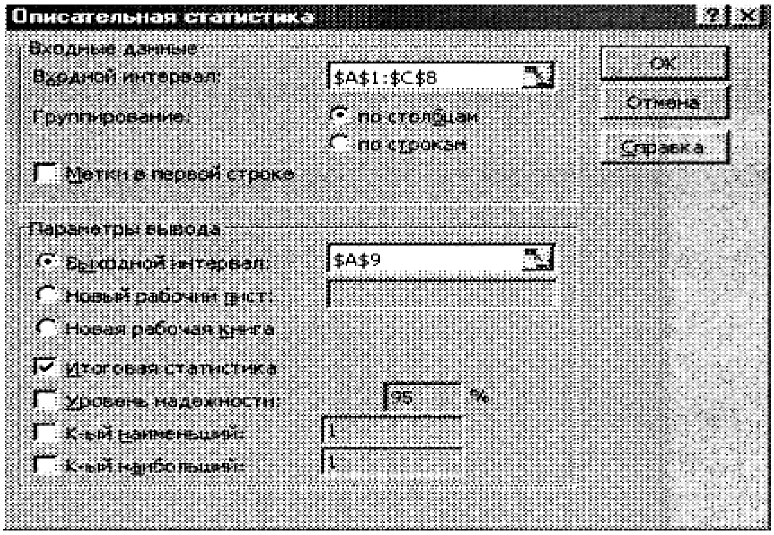


Рис. 6.8. Пример заполнения диалогового окна Описательная статистика

3. В появившемся диалоговом окне (рис. 6.8) в рабочем поле Входной интервал укажите входной диапазон —*A1:C8.* Активировав переключателем рабочее поле Выходной интервал, укажите выходной диапазон — ячейку *A9.* В разделе Группировка переключатель установите в положение по столбцам. Установите флажок в поле Итоговая статистика и нажмите кнопку ОК.

В результате анализа (рис. 6.9) в указанном выходном диапазоне для каждого столбца данных получим соответствующие результаты.

	A	B	C	D	E	F
9	*Столбец1*		*Столбец2*		*Столбец3*	
10						
11	Среднее	3540	Среднее	1962,5	Среднее	2400
12	Стандартная ошибка	355,8089	Стандартная ошибка	41,99277	Стандартная ошибка	243,5843
13	Медиана	3700	Медиана	2000	Медиана	2250
14	Мода	#Н/Д	Мода	2000	Мода	#Н/Д
15	Стандартное отклонение	795,613	Стандартное отклонение	118,7735	Стандартное отклонение	596,6574
16	Дисперсия выборки	633000	Дисперсия выборки	14107,14	Дисперсия выборки	356000
17	Эксцесс	-1,29385	Эксцесс	-1,22929	Эксцесс	-2,06887
18	Асимметричность	-0,24502	Асимметричность	-0,39433	Асимметричность	0,457606
19	Интервал	2000	Интервал	300	Интервал	1400
20	Минимум	2500	Минимум	1800	Минимум	1800
21	Максимум	4500	Максимум	2100	Максимум	3200
22	Сумма	17700	Сумма	15700	Сумма	14400
23	Счет	5	Счет	8	Счет	6

Рис. 6.9. Результаты работы инструмента Описательная статистика

Все полученные характеристики были рассмотрены ранее в разделе «Выборочные характеристики» данной главы, за исключением последних четырех:

- минимум — значение минимального элемента выборки;
- максимум — значение максимального элемента выборки;
- сумма — сумма значений всех элементов выборки;
- счет — количество элементов в выборке.

Среди этих характеристик наиболее важными являются показатели Среднее, Стандартная ошибка (среднего) и Стандартное отклонение.

УПРАЖНЕНИЯ.

1. Найдите наиболее популярный туристический маршрут из четырех реализуемых фирмой (моду), если за неделю последовательно были реализованы следующие маршруты (приводятся номера маршрутов): 1,3, 3, 2,1,1,4,4, 2,4, 1, 3, 2, 4,1,4,4,3,1,2,3,4,1,1,3.
2. В рабочей зоне производились замеры концентрации вредного вещества. Получен ряд значений (в мг/м^3): 12, 16, 15, 14, 10, 20, 16, 14, 18, 14, 15, 17, 23, 16. Необходимо определить основные выборочные характеристики.

Проверка статистических гипотез

Здесь решается вопрос, отражают ли наблюдаемые данные объективно существующую реальность. Указанный вопрос решается проверкой соответствующих статистических гипотез. При этом могут выявляться достоверности различий между выборками, взаимосвязи между выборками, влияющие факторы и т. п.

Принятие статистических решений

Статистическая гипотеза — это предположение о виде или отдельных параметрах распределения вероятностей, которое подлежит проверке на имеющихся данных.

Проверка статистических гипотез — это процесс формирования решения о возможности принять или отвергнуть утверждение (гипотезу), основанный на информации, полученной из анализа выборки. Методы проверки гипотез называются критериями.

В большинстве случаев рассматривают так называемую нулевую гипотезу (нуль-гипотезу H_0), состоящую в том, что все события произошли случайно, естественным образом. Альтернативная гипотеза (H_1) состоит в том, что события случайным образом произойти не могли, и имело место воздействие некого фактора.

Обычно нулевая гипотеза формулируется таким образом, чтобы на основании эксперимента или наблюдений ее можно было отвергнуть с заранее заданной вероятностью ошибки а. Эта, заранее заданная вероятность ошибки, называется уровнем значимости.

Уровень значимости — максимальное значение вероятности появления события, при котором событие считается практически невозможным. В статистике наибольшее распространение получил уровень значимости, равный ά = 0,05. Поэтому если вероятность, с которой интересующее событие может произойти случайным образом p < 0,05, то принято считать это событие маловероятным, и если оно все же произошло, то это не было случайным. В наиболее ответственных случаях, когда требуется особая уверенность в достоверности полученных результатов, надежности выводов, уровень значимости принимают равным ά = 0,01 или даже ά = 0,001.

Величину P, равную 1 - ά, называют доверительной вероятностью (уровнем надежности), то есть вероятностью, признанной достаточной для того, чтобы уверенно судить о принятом статистическом решении. Соответственно, в качестве доверительных вероятностей выбирают значения 0,95, 0,99 и 0,999.

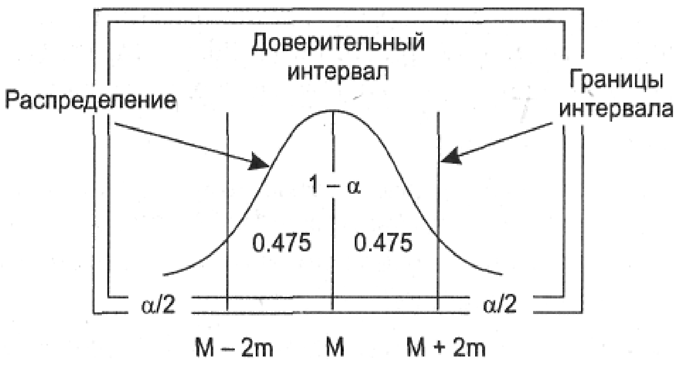


Рис. 6.10. 95%-ный доверительный интервал для среднего значения

Интервал, в котором с заданной доверительной вероятностью P = 1 - ά находится оцениваемый параметр, называется доверительным интервалом. В соответствии с доверительными вероятностями на практике используются 95%-, 99%-, 99,9%-ные доверительные интервалы. Граничные точки доверительного интервала называют доверительными пределами (рис. 6.10).

Выбор того или иного уровня значимости, выше которого результаты отвергаются как статистически не подтвержденные, или, соответственно, доверительной вероятности, в общем случае является произвольным. Окончательное решение зависит от исследователя, традиций и накопленного практического опыта в данной области исследований.

Анализ одной выборки

Анализ однородности выборки. Одним из важных вопросов, возникающих при анализе выборки, является вопрос: относится та или иная варианта к данной статистической совокупности? Решение вопроса не представляет сложности, если распределение в этой совокупности является нормальным. Для этого достаточно использовать правило трех сигм. Согласно этому правилу, в пределах $M \pm 3a$ находится 99,7% всех вариант. Поэтому, если варианта попадает в этот интервал, то она считается принадлежащей к данной совокупности. Если не попадает, то она может быть отброшена. Хотя этот метод и предполагает нормальность исходного распределения, на практике он успешно работает и может быть использован в большинстве других случаев.

При числе элементов в выборке $n < 30$ способ более точного определения границ доверительного интервала по формуле

$$\left[M - t_{n.p.}s; M + t_{n.p.}s\right] \qquad (6.1)$$

будет показан ниже в примере 6.5. В формуле (6. 1) M — среднее значение, s — стандартное отклонение, $t_{n,p.}$ — табличное значение распределения Стьюдента с числом степеней свободы n и доверительной вероятностью p.

Построение доверительных интервалов для среднего. Еще одной важной задачей, возникающей при анализе одной выборки, является сравнение выборочного среднего арифметического со средним значением генеральной совокупности. Эта задача решается с помощью статистических критериев. При этом выясняется, значимо ли отличие выборочного среднего значения от среднего значения генеральной совокупности, из которой предположительно взята выборка, или наблюдаемое различие является случайным.

Действительно, средние значения, получаемые по выборочным данным, обычно не совпадают с генеральным средним (математическим ожиданием). В связи с этим возникает вопрос: можно ли по результатам выборочной оценки судить о свойствах всей генеральной совокупности?

Поскольку каждую оценку, полученную в отдельной выборке, можно рассматривать как случайную величину, то при увеличении числа выборок распределение отдельных оценок будет принимать характер нормального распределения. Это значит, что в случае средних арифметических значения выборочных средних относительно генерального среднего распределяются по нормальному закону. То есть так же, как относительные отклонения нормально распределенных вариант от среднего арифметического выборки.

Отсюда, в частности, следует, что 68,3% всех выборочных средних находятся в пределах $\Delta = M \pm m$, где Δ — предельная ошибка выборки, M — среднее выборочное, m — стандартное отклонение среднего значения. Иными словами, имеется вероятность 0,683, что выборочное среднее отличается от генерального не более, чем на $\pm m$. Здесь 0,683 — доверительная вероятность, 1 - 0,683 = 0,317 — уровень значимости а, А = $M \pm m$ — 68% доверительный интервал.

Для принятой в большинстве исследований доверительной вероятности 0,95, доверительный интервал для средних при достаточно большом числе наблюдений (и > 30) примерно равен $\pm 2m$ (см. рис. 6.8). При доверительной вероятности 0,99, доверительный интервал составит примерно $\pm 3m$. Для более точного определения границ доверительного интервала можно воспользоваться формулой

$$\left[M - t_{n.p.} \frac{s}{\sqrt{n}} ; M + t_{n.p.} \frac{s}{\sqrt{n}} \right]$$

где M — среднее значение, s — стандартное отклонение, t_{nf} — табличное значение распределения Стьюдента с числом степеней свободы n и доверительной вероятностью p, n — количество элементов в выборке.

В MS Excel для более точного вычисления границ доверительного интервала и при числе элементов в выборке $n < 30$ можно воспользоваться функцией ДОВЕРИТ или процедурой Описательная статистика.

Функция *ДОВЕРИТ (альфа;станд_откл;размер)* определяет полуширину доверительного интервала и содержит следующие параметры:

- *Альфа* — уровень значимости, используемый для вычисления доверительной вероятности. Доверительная вероятность равняется 100*(1 - *альфа)%* процентам, или, другими словами, *альфа,* равное 0,05, означает 95%-ный уровень доверительной вероятности;
- *Станд_откл* — стандартное отклонение генеральной совокупности для интервала данных, предполагается известным;
- *Размер* — это размер выборки.

Пример. Найти границы 95%-ного доверительного интервала для среднего значения, если у 25 телефонных аккумуляторов среднее время разряда в режиме ожидания составило 140 часов, *а* стандартное отклонение — 2,5 часа.

Решение.

1. Откройте новую рабочую таблицу. Установите табличный курсор в ячейку А1.
2. Для определения границ доверительного интервала необходимо на панели инструментов Стандартная нажать кнопку Вставка функции (fx). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ДОВЕРИТ, после чего нажмите кнопку ОК.
3. В рабочие поля появившегося диалогового окна ДОВЕРИТ с клавиатуры введите условия задачи: Альфа — *0,05',* Станд_откл — *2,5',* Размер —*25* (рис. 6.11). Нажмите кнопку ОК.

В ячейке А1 появится полуширина 95%-ного доверительного интервала для среднего значения выборки — 0,979981. Другими словами, с 95%-ным уровнем надежности можно утверждать, что средняя продолжительность разряда аккумулятора составляет 140 + 0,979981 часа или от 139,02 до 140,98 часа.

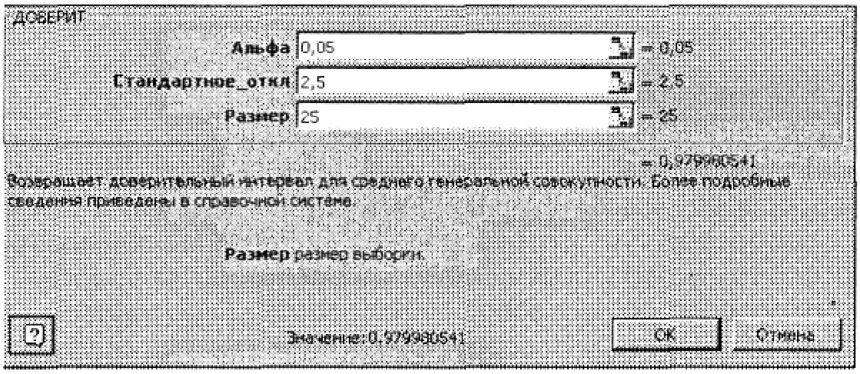


Рис. 6.11. Пример заполнения диалогового окна ДОВЕРИТ

Пример. Пусть имеется выборка, содержащая числовые значения: 13, 15, 17, 19, 22,25,19. Необходимо определить границы 95%-ного доверительного интервала для среднего значения и для нахождения «выскакивающей» варианты.

Решение.

1. В диапазон А1:А7 введите исходный ряд чисел.

2. Далее вызовите процедуру Описательная статистика. Для этого, указав курсором мыши на пункт меню Сервис, выберите команду Анализ данных. Затем в появившемся списке Инструменты анализа выберите строку Описательная статистика.

3. В появившемся диалоговом окне в рабочем поле Входной и интервал: укажите входной диапазон —*A1:A7*. Переключателем активизируйте Выходной интервал и укажите выходной диапазон — ячейку *B1*. В разделе Группировка переключатель установите в положение по столбцам. Установите флажок в левое поле Уровень надежности: и в правом поле (%) — *95*. Затем нажмите кнопку ОК.

4. В результате анализа в указанном выходном диапазоне для доверительной вероятности 0,95 получаем значения доверительного интервала (рис. 6.12).

	A	B	C
1	13	*Столбец1*	
2	15		
3	17	Уровень надежности(95,0%)	3,770272
4	19		
5	22		
6	25		
7	19		

Рис. 6.12. Исходная выборка (А1:А7) и результат вычислений (С3) из примера

Уровень надежности — это половина доверительного интервала для генерального среднего арифметического. Из полученного результата следует, что с вероятностью 0,95 среднее арифметического для генеральной совокупности находится в интервале 18,571 ± 3,77. Здесь 18,571 — выборочное среднее M для рассматриваемого примера, которое находится обычно процедурой Описательная статистика одновременно с доверительным интервалом.

5. Для нахождения доверительных границ для «выскакивающей» варианты необходимо полученный выше доверительный интервал умножить на $\sqrt{n}$ (в примере — $\sqrt{7}$, то есть 3,77* $\sqrt{7}$ = 9,975). В Excel это можно выполнить следующим образом. Табличный курсор установите в свободную ячейку C4; введите с клавиатуры знак =; мышью укажите на ячейку C3 (в которой находится результат вычислений); введите с клавиатуры знак *; с панели инструментов Стандартная вызовите Мастер функций (кнопка fx); выберите категорию Математические, тип функции Корень; нажмите ОК; введите с клавиатуры число *n*= *7* и нажмите ОК. В результате получим в ячейке C 4 значение доверительного интервала — *9,975.*

Таким образом, варианта, попадающая в интервал 18,571 ± 9,975, считается принадлежащей данной совокупности с вероятностью 0,95. Выходящая за эти границы может быть отброшена с уровнем значимости ά = 0,05.

Проверка соответствия теоретическому распределению.

Следующей задачей, возникающей при анализе одной выборки, является оценка меры соответствия (расхождения) полученных эмпирических данных и каких-либо теоретических распределений. Это связано с тем, что в большинстве случаев при решении реальных задач закон распределения и его параметры неизвестны. В то же время применяемые статистические методы в качестве предпосылок часто требуют определенного закона распределения.

Наиболее часто проверяется предположение о нормальном распределении генеральной совокупности, поскольку большинство статистических процедур ориентировано на выборки, полученные из нормально распределенной генеральной совокупности.

Для оценки соответствия имеющихся экспериментальных данных нормальному закону распределения обычно используют графический метод, выборочные параметры формы распределения и критерии согласия.

Графический метод позволяет давать ориентировочную оценку расхождения или совпадений распределений (рис. 6.13).

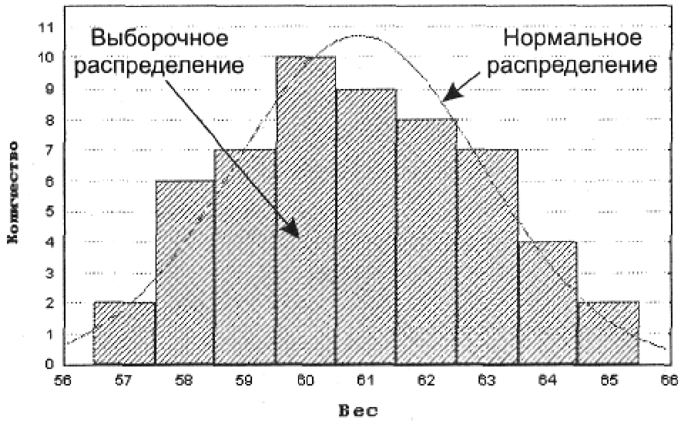


Рис. 6.13. Сопоставление выборочного распределения веса студенток и кривой нормального распределения

При большом числе наблюдений (n > 100) неплохие результаты дает вычисление выборочных параметров формы распределения: эксцесса и асимметрии. Предположение о нормальности распределения не противоречит имеющимся данным, если асимметрия близка к нулю, то есть лежит в диапазоне от -0,2 до 0,2, а эксцесс — от 2 до 4.

Наиболее убедительные результаты дает использование критериев согласия. *Критериями согласия* называют статистические критерии, предназначенные для проверки согласия опытных данных и теоретической модели. Здесь нулевая гипотеза H_∂ представляет собой утверждение о том, что распределение генеральной совокупности, из которой получена выборка, не отличается от нормального. Среди критериев согласия большое распространение получил непараметрический критерий χ^2 (хи-квадрат). Он основан на сравнении эмпирических частот интервалов группировки с теоретическими (ожидаемыми) частотами, рассчитанными по формулам нормального распределения.

Уверенно о нормальности закона распределения можно судить, если имеется не менее 50 результатов наблюдений. В случаях меньшего числа данных можно говорить только о том, что данные не противоречат нормальному закону, и в этом случае обычно используют графические методы оценки соответствия. При большем числе наблюдений целесообразно совместное использование графических и статистических (например, тест хи-квадрат или аналогичные) методов оценки, естественно дополняющих друг друга.

Использование критерия согласия хи-квадрат. Для применения критерия желательно, чтобы объем выборки $n \geq 40$, выборочные данные были сгруппированы в интервальный ряд с числом интервалов не менее 7, а в каждом интервале находилось не менее 5 наблюдений (частот).

Сравниваться должны именно абсолютные частоты, а не относительные (частости). Критерий хи-квадрат не доказывает справедливость нулевой гипотезы (соответствие эмпирического распределения нормальному), а лишь может позволить ее отвергнуть с определенной вероятностью (уровнем значимости).

В MS Excel критерий хи-квадрат реализован в функции ХИ2ТЕСТ. Функция ХИ2ТЕСТ вычисляет вероятность совпадения наблюдаемых (фактических) значений и теоретических (гипотетических) значений. Если вычисленная вероятность ниже уровня значимости (0,05), то нулевая гипотеза отвергается и утверждается, что наблюдаемые значения не соответствуют нормальному закону распределения. Если вычисленная вероятность близка к 1, то можно говорить о высокой степени соответствия экспериментальных данных нормальному закону распределения.

Функция имеет следующие параметры: *ХИ2ТЕСТ (фактический_интервал;ожидаемый_интервал)*. Здесь:

- *фактический_интервал* — это интервал данных, которые содержат наблюдения, подлежащие сравнению с ожидаемыми значениями;
- *ожидаемый_интервал* — это интервал данных, который содержит теоретические (ожидаемые) значения для соответствующих наблюдаемых.

Пример. Проверить соответствие выборочных данных из предыдущего примера.
(64,57,63,62,58,61,63,60,60,61,65,62,62,60,64,61,59,59,63,61,62,58,58,63,61,59,62,60,60,58,61,60,63,63,58,60,59,60,59,61,62,62,63,57,61,58,60,64,60,59,61,64,62, 59, 65) нормальному закону распределения.

Решение.

1. Повторите пункты 1-7 решения предыдущего примера. В результате получится таблица (см. рис. 6.2).

2. Найдите теоретические частоты нормального распределения. Для этого предварительно необходимо найти среднее значение и стандартное отклонение выборки.

В ячейке I13 с помощью функции СРЗНАЧ найдите среднее значение для данных из диапазона A2:E12 (60,855). В ячейке J13 с помощью функции СТАНДОТКЛОН найдите стандартное отклонение для этих же данных (2,05). В ячейки K1 и K2 введите название столбца — *Теоретические частости*. Затем с помощью функции НОРМРАСП найдите теоретические частости. Установите курсор в ячейку K4, вызовите указанную функцию и заполните ее рабочие поля: *x* — *G4;* Среднее — *I13;* Стандартное_откл — *J13;* Интегральный — *0.* Получим в ячейке K4 *0,033.* Далее протягиванием скопируйте содержимое ячейки K4 в диапазон ячеек K5:K12. Затем в ячейки L1 и L2 введите название нового столбца — *Теоретические частоты.* Установите курсор в ячейку L4 и введите формулу *=H$13*K4.* Далее протягиванием скопируйте содержимое ячейки L4 в диапазон ячеек L5:L12. Результаты вычислений представлены на рис. 6.14.

	G	H	I	J	K	L
1	Вес	Абсолютные	Относительны	Накопленные	Теоретические	Теоретические
2	кг	частоты	частоты	частоты	частости	частоты
3						
4	57	2	0,036	0,036	0,033205828	1,82632055
5	58	6	0,109	0,145	0,073795567	4,058756212
6	59	7	0,127	0,273	0,129258576	7,109221655
7	60	10	0,182	0,455	0,178443849	9,814411704
8	61	9	0,164	0,618	0,194158732	10,67873029
9	62	8	0,145	0,764	0,16650428	9,157735407
10	63	7	0,127	0,891	0,112540024	6,189701326
11	64	4	0,073	0,964	0,059951732	3,297345259
12	65	2	0,036	1,000	0,025171529	1,384434082

Рис. 6.14. Результаты вычисления теоретических частостей и частот из примера

3. С помощью функции ХИ2ТЕСТ определите соответствие данных нормальному закону распределения. Для этого установите табличный курсор в свободную ячейку L13. На панели инструментов Стандартная нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ХИ2ТЕСТ, после чего на-

жмите кнопку ОК. Появившееся диалоговое окно ХИ2ТЕСТ отодвиньте вправо на 1-2 см от данных. Указателем мыши в рабочие поля введите фактический Н4:Н12 и ожидаемые L4:L12 диапазоны частот (рис. 6.15). Нажмите кнопку ОК. В ячейке L13 появится значение вероятности того, что выборочные данные соответствуют нормальному закону распределения — 0,9842.

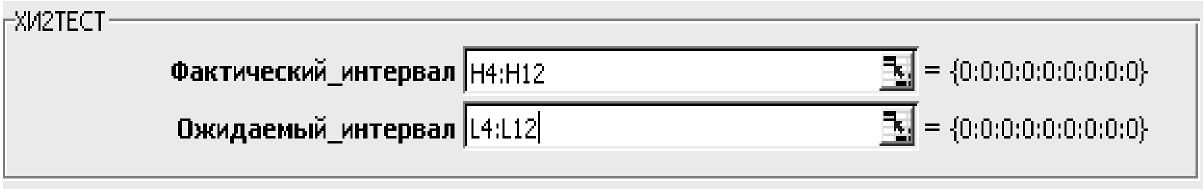


Рис. 6.15. Пример заполнения рабочих полей функции ХИ2ТЕСТ

4. Поскольку полученная вероятность соответствия экспериментальных данных p = 0,98 много больше, чем уровень значимости ά = 0,05, то можно утверждать, что нулевая гипотеза не может быть отвергнута и, следовательно, данные не противоречат нормальному закону распределения. Более того, поскольку полученная вероятность p = 0,98 близка к 1, можно говорить о высокой степени вероятности того, что экспериментальные данные соответствуют нормальному закону.

УПРАЖНЕНИЯ.

1. Определите, лежит ли значение 19 внутри границ 95%-ного доверительного интервала выборки 2, 3, 5, 7, 4, 9, 6, 4, 9, 10,4, 7, 19.
2. Определите с уровнем значимости а = 0,05 максимальное отклонение среднего значения генеральной совокупности от среднего выборки 3,4, 4, 2, 5, 3, 4, 3, 5,4,3,5,6.
3. Найдите соответствие экспериментальных данных нормальному закону распределения для следующей выборки весов детей (кг):
21,21,22,22,22,22,22,22,22,22,22,23,23,23,23,23,23,24,24,24,24,24, 24,24,24,24,24,24,25,25,25, 25,25,25,25,25,25,25,25,25,25,25,25,25,25,26,26,26,26,26,26,26,26, 26,26,26, 26, 26, 26, 27, 27.

Анализ двух выборок

После определения основных выборочных характеристик и анализа одной выборки, следующей задачей статистического анализа, является совместный анализ нескольких выборок. При анализе двух выборок, важнейшим является вопрос о наличии различий между этими выборками. Обычно для этого проводят проверку статистических гипотез о принадлежности обеих выборок одной генеральной совокупности или о равенстве генеральных средних. В рассмотренном ранее примере такие различия выявляются путем сравнения данных реализации турфирмой путевок за периоды до и после начала активной рекламной компании. Если сопоставить средние значения числа реализованных за месяц путевок до (125,6) и после (145,7) начала рекламной компании, видно, что они различаются. Можно ли по этим данным сделать вывод об эффективности рекламной компании?

Для решения таких задач используются *критерии различия*. Для проверки одной и той же гипотезы могут быть использованы разные статистические критерии.

Статистические критерии различия подразделяются на *параметрические и непараметрические критерии*. Параметрические критерии служат для проверки гипотез о параметрах определенных распределений генеральной совокупности (чаще всего нормального распределения). Непараметрические критерии для проверки гипотез не используют предположений о законе распределения генеральной совокупности и не требуют знания параметров распределения.

Параметрические критерии.

Параметрические критерии служат для проверки гипотез о положении и рассеивании. Из параметрических критериев наибольшей популярностью при проверке гипотез о равенстве генеральных средних (математических ожиданий) пользуется *t*-критерий Стьюдента (t-критерий различия).

Критерий Стьюдента *(t)* наиболее часто используется для проверки гипотезы: «Средние двух выборок относятся к одной и той же совокупности». Критерий позволяет найти вероятность того, что оба средних относятся к одной и той же совокупности. Если эта вероятность p ниже уровня значимости (p < 0,05), то принято считать, что выборки относятся к двум разным совокупностям.

При использовании *t*-критерия можно выделить два случая. В первом случае его применяют для проверки гипотезы о равенстве генеральных средних двух независимых, несвязанных выборок (так называемый двухвыборочный *t*-критерий). В этом случае есть контрольная группа и опытная группа, состоящие из разных пациентов, количество которых в группах может быть различно. Во втором случае, когда одна и та же группа объектов порождает числовой материал для проверки гипотез о средних, используется так называемый парный t-критерий. Выборки при этом называют зависимыми, связанными. Например, измеряется содержание лейкоцитов у здоровых животных, а затем у тех же самых животных после облучения определенной дозой излучения.

В обоих случаях должно выполняться требование нормальности распределения исследуемого признака в каждой из сравниваемых групп и равенства дисперсий в сравниваемых совокупностях. Однако на практике по большому счету корректное применение *t*-критерия Стьюдента для двух групп часто бывает затруднительно.

Для оценки достоверности отличий по критерию Стьюдента принимается нулевая гипотеза, что средние выборок равны между собой. Затем вычисляется значение вероятности того, что изучаемые события (например, количества реализованных путевок в обеих выборках) произошли случайным образом. В MS Excel для оценки достоверности отличий по критерию Стьюдента используются специальная функция ТТЕСТ и процедуры пакета анализа. Все перечисленные инструменты вычисляют вероятность, соответствующую критерию Стьюдента, и используются, чтобы определить, насколько вероятно, что две выборки взяты из генеральных совокупностей, которые имеют одно и то же среднее.

Функция ТТЕСТ использует следующие параметры: *ТТЕСТ(массив1;массив2;хвосты;-тип)*. Здесь:

- *массив* 1 — это первое множество данных;
- *массив* 2 — это второе множество данных;
- *хвосты* — число хвостов распределения. Обычно число хвостов равно 2;

- *тип* — это вид исполняемого £-теста. Возможны 3 варианта выбора: 1 — парный тест, 2 — двухвыборочный тест с равными дисперсиями, 3 — двухвыборочный тест с неравными дисперсиями.

Пример. Выявить, достоверны ли отличия при сравнении данных реализации турфирмой путевок за периоды до и после начала активной рекламной компании (см. пример 6.3).

Решение.

1. Введите данные (как в пункте 1 примера 6.3).
2. Для выявления достоверности отличий табличный курсор установите в свободную ячейку (All). На панели инструментов необходимо нажать кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ТТЕСТ, после чего нажмите кнопку ОК. Появившееся диалоговое окно ТТЕСТ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных контрольной группы в поле Массив 1 *(А2:А8).* В поле Массив 2 введите диапазон данных исследуемой группы *(В2:В8).* В поле Хвосты всегда вводится с клавиатуры цифра *2* (без кавычек), а в поле Тип с клавиатуры введите цифру *3.* Нажмите кнопку ОК. В ячейке All появится значение вероятности - ***0,006295.***
3. Поскольку величина вероятности случайного появления анализируемых выборок (0,006295) меньше уровня значимости (а = 0,05), то нулевая гипотеза отвергается. Следовательно, различия между выборками не случайные и средние выборок считаются достоверно отличающимися друг от друга. Поэтому на основании применения критерия Стьюдента можно сделать вывод о большей эффективности реализации путевок после начала рекламной компании ($p < 0{,}05$).

Как указывалось выше, при использовании i-критерия выделяют два основных случая. В первом случае его применяют для проверки гипотезы о равенстве генеральных средних двух независимых, несвязанных выборок (так называемый двухвыборочный t-критерий). В этом случае есть две различных выборки, количество элементов в которых может быть также различно. При заполнении диалогового окна ТТЕСТ при этом указывается Тип 3.

Во втором случае, когда одна и та же группа объектов порождает числовой материал для проверки гипотез о средних, используется так называемый парный t-кри-терий. Выборки при этом называют зависимыми, связанными (при заполнении диалогового окна ТТЕСТ указывается Тип *1)*. Например, сравнивается реализация путевок двумя фирмами в соответствующие месяцы. В качестве упражнения рассмотрим пример.

Пример. Сравнивается количество наличных денег у двух групп студентов (в рублях):

30	10
30	20
40	30
50	40
60	50

Необходимо определить достоверность различия между группами при двух вариантах постановки задачи:

- группы состоят из различных студентов (тип 3);
- группы состоят из одних и тех же студентов, но первая — до посещения буфета, а вторая — после (тип 1).

Решение. В ячейки С1:С5 введите количество денег у студентов первой группы. В ячейки D1:D5 введите количество денег у студентов второй группы.

1. Табличный курсор установите в свободную ячейку (С6). На панели инструментов необходимо нажать кнопку Вставка функции (f_x). В появившемся диалоговом'окне Мастер функций выберите категорию Статистические и функцию ТТЕСТ, после чего нажмите кнопку ОК. Появившееся диалоговое окно ТТЕСТ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши ввести диапазон данных первой группы в поле Массив 1 *(C1:C5)*. В поле Массив 2 введите диапазон данных второй группы *(D1:D5)*. В поле Хвосты всегда вводится цифра *2* (без кавычек), а в поле Тип введите цифру *3*. Нажмите кнопку ОК. В ячейке Сб появится значение вероятности - *0,228053*.

Поскольку величина вероятности случайного появления анализируемых выборок (0,228053) больше уровня значимости (а = 0,05), то нулевая гипотеза не может быть отвергнута (принимается). Следовательно, различия между выборками могут быть случайными и средние выборок не считаются достоверно отличающимися друг от друга. Поэтому на основании применения критерия Стьюдента нельзя сделать вывод о достоверности отличий двух групп студентов по количеству карманных денег, имеющихся у них ($p > 0{,}05$).

2. Табличный курсор установите в свободную ячейку (D6). На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ТТЕСТ, после чего нажмите кнопку ОК. Появившееся диалоговое окно ТТЕСТ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных первой группы в поле Массив 1 *(C1:C5)*. В поле Массив 2 введите диапазон данных второй группы *(D1:D5)*. В поле Хвосты всегда вводится цифра 2 (без кавычек), а в поле Тип введите цифру *1*. Нажмите кнопку ОК. В ячейке Об появится значение вероятности — *0,003883*.

Поскольку величина вероятности случайного появления анализируемых выборок (0,003883) меньше уровня значимости (а = 0,05), то нулевая гипотеза отвергается. Следовательно, различия между выборками не могут быть случайными и средние выборок считаются достоверно отличающимися друг от друга. Поэтому на основании применения критерия Стьюдента можно сделать вывод о том, что в двух группах студентов выявлены достоверные отличия по количеству карманных денег ($p < 0{,}05$), что явилось результатом посещения буфета.

Применение различных типов критерия Стьюдента может приводить к различным результатам на основании одних и тех же исходных данных. Можно предложить следующий приблизительный способ выбора типа критерия: если не ясно, какой тип критерия выбирать, выбирается тип 3; если очевидно, что выборки зависимы, связаны (например, это одни и те же студенты), то следует выбирать тип 1.

Критерий Фишера.

Критерий Фишера используют для проверки гипотезы о принадлежности двух дисперсий одной генеральной совокупности и, следовательно, их равенстве. При этом предполагается, что данные независимы и распределены по нормальному закону. Гипотеза о равенстве дисперсий принимается, если отношение большей дисперсии к меньшей меньше критического значения распределения Фишера.

$$F = \frac{s_1^2}{s_2^2} \qquad F \langle F_{крит}$$

где $F_{крит}$ зависит от уровня значимости и числа степеней свободы для дисперсий в числителе и знаменателе.

В MS Excel для расчета уровня вероятности выполнения гипотезы о равенстве дисперсий могут быть использованы функция ФТЕСТ(*массив1*;*массив2)* и процедура пакета анализа Двухвыборочный F-тест для дисперсий.

Непараметрические критерии.

Непараметрические критерии используются в тех случаях, когда закон распределения данных отличается от нормального или неизвестен. Из большого числа непараметрических критериев рассмотрим критерий хи-квадрат.

Критерий согласия x^2- Бывают ситуации, когда необходимо сравнить две относительные или выраженные в процентах величины (доли). Примером может служить случай проверки успешности трудоустройства молодых специалистов, когда известен процент трудоустроившихся выпускников двух институтов. Для проверки достоверности различий здесь критерий Стьюдента применить не удастся. В таких задачах обычно используют критерий x^2 (хи-квадрат). Критерий хи-квадрат относится к непараметрическим критериям.

Здесь, как и в случае с критерием Стьюдента, принимается нулевая гипотеза о том, что выборки принадлежат к одной генеральной совокупности. Кроме того, определяется ожидаемое значение результата. Обычно это среднее значение между выборками рассматриваемого показателя. Затем оценивается вероятность того, что ожидаемые значения и наблюдаемые принадлежат к одной генеральной совокупности.

В MS Excel критерий хи-квадрат реализован в функции ХИ2ТЕСТ. Функция ХИ2ТЕСТ вычисляет вероятность совпадения наблюдаемых (фактических) значений и теоретических (гипотетических) значений. Если вычисленная вероятность ниже уровня значимости (0,05), то нулевая гипотеза отвергается и утверждается, что наблюдаемые значения не соответствуют теоретическим (ожидаемым) значениям.

Функция имеет следующие параметры: *КУ\21ЕС\(фактический_интервал;ожидаемый_интервал).* Здесь:

- *фактический_интервал* — это интервал данных, которые содержат наблюдения, подлежащие сравнению с ожидаемыми значениями;
- *ожидаемый_интервал* — это интервал данных, который содержит теоретические (ожидаемые) значения для соответствующих наблюдаемых.

Пример. Пусть после окончания двух институтов экономического профиля трудоустроилось по специальности из первого института 90 человек, а из второго 60 (обе группы молодых специалистов включали по 100 человек).

Решение.

1. Принимается нулевая гипотеза, что выборки принадлежат к одной генеральной совокупности.
2. Определяется ожидаемое значение результата (среднее значение между выборками): (60 + 90)/2 = 75, то есть мы ожидали, что разницы между группами нет, и в обоих случаях должно было трудоустроиться по 75 человек.
3. Затем вычисляется значение вероятности того, что изучаемые события (трудоустройство в обеих выборках) произошли случайным образом. Для этого введите данные в рабочую таблицу: *90* — в ячейку E1, *60* — в F1, 75 — в E2,F2. Табличный курсор установите в свободную ячейку (ЕЗ). На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию ХИ2ТЕСТ, после чего нажмите кнопку ОК. Появившееся диалоговое окно ХИ2ТЕСТ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой кнопке). Указателем мыши введите диапазон данных наблюдавшегося количества трудоустроившихся в поле Фактический интервал (E1:F1). В поле Ожидаемый интервал введите диапазон данных предполагаемого количества трудоустроившихся *(E2:F2).* На-

жмите кнопку OK. В ячейке E3 появится значение вероятности — *0,014306*.

4. Поскольку величина вероятности случайного появления анализируемых выборок (0,0143) меньше уровня значимости (a = 0,05), то нулевая гипотеза отвергается. Следовательно, различия между выборками не могут быть случайными и выборки считаются достоверно отличающимися друг от друга. Поэтому на основании применения критерия хи-квадрат можно сделать вывод о том, что в двух группах выпускников выявлены достоверные отличия по успешности трудоустройства ($p < 0,05$), что, по-видимому, явилось результатом более высокой репутации выпускников первого института.

УПРАЖНЕНИЯ.

1. Даны результаты бега на дистанции 100 м в секундах в двух группах студентов. Студенты первой группы в течение года посещали факультативные занятия по физкультуре. Определите, достоверны ли отличия по результатам бега в этих группах.

Посещавшие факультатив	Не посещавшие
12,6	12,8
12,3	13,2
11,9	13,0
12,2	12,9
13,0	13,5
12,4	13,1

2. В ходе социологического опроса на вопрос о перенесенном в детстве заболевании ответы распределились следующим образом:

	Да	Нет	Не помню
Мужчины	58	11	10
Женщины	35	25	23

Есть ли достоверные отличия в ответах женщин и мужчин?

3. Приведены данные ежемесячной результативности (количество голов) футбольной команды в двух сезонах:

Месяц	3	4	5	6	7	8	9	10	11
2000 г.	3	4	5	8	9	1	2	4	5
2001 г.	6	19	3	2	14	4	5	17	1

Определите, есть ли статистические различия в ежемесячной результативности команды в рассматриваемых сезонах?

4. Определите, имеют ли выборки {6; 7; 9; 15; 21} и (20; 28; 31; 38; 40} различные уровни разнородности (отличаются ли дисперсии)?

Использование инструмента Пакет анализа для выявления различий между выборками

Для анализа двух выборок с помощью f-теста Стьюдента могут быть использованы следующие процедуры: Парный двухвыборочный t-тест для средних; Двухвыборочный t-тест с одинаковыми дисперсиями и Двухвыборочный t-тест с различными дисперсиями. Как указывалось в разделе «Анализ двух выборок», в общем случае необходимо воспользоваться процедурой Двухвыборочный t-тест с различными дисперсиями, так как процедуры Парный двухвыборочный t-тест для средних и Двухвыборочный t-тест с одинаковыми дисперсиями относятся к частным, специальным случаям.

Для выполнения процедуры анализа необходимо:

- выполнить команду Сервис > Анализ данных;
- в появившемся списке Инструменты анализа выбрать строку Двухвыборочный t-тест с различными дисперсиями, щелкнуть левой кнопкой мыши и нажать кнопку ОК;
- в появившемся диалоговом окне указать Интервал переменной 1, то есть ввести ссылку на первый диапазон анализируемых данных, содержащий один столбец данных. Для этого следует навести указатель мыши на верхнюю ячейку первого столбца данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к нижней ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши;
- указать Интервал переменной 2, то есть ввести ссылку на второй диапазон анализируемых данных, содержащий один столбец данных. Для этого следует навести указатель мыши в поле ввода Интервал переменной 2 и щелкнуть левой кнопкой мыши,

затем навести указатель мыши на верхнюю ячейку второго столбца данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к нижней ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши;

- указать выходной диапазон, то есть ввести ссылку на ячейки, в которые будут выведены результаты анализа. Для этого следует поставить флажок в левое поле Выходной диапазон (навести указатель мыши и щелкнуть левой кнопкой), далее навести указатель мыши на правое поле ввода Выходной диапазон и щелкнуть левой кнопкой мыши, затем указатель мыши навести на левую верхнюю ячейку выходного диапазона и щелкнуть левой кнопкой мыши. Размер выходного диапазона будет определен автоматически, и на экран будет выведено сообщение в случае возможного наложения выходного диапазона на исходные данные;
- нажать кнопку ОК.

Результаты анализа. В выходной диапазон будут выведены: средняя, дисперсия и число наблюдений для каждой переменной, гипотетическая разность средних, *df* (число степеней свободы), значение t-статистики, *P(T<=t)* одностороннее, *t* критическое одностороннее, *P(T<= t)* двухстороннее, *t* критическое двухстороннее.

Интерпретация результатов. Если величина вероятности случайного появления анализируемых выборок *(P(T<=t)* двухстороннее) меньше уровня значимости (a = 0,05), принято считать, что различия между выборками не случайные, то есть различия достоверные.

Пример. Рассматривается заработная плата обслуживающего персонала и работников ресторана (из примера приведенного выше).

Персонал	Ресторан
2100	3200
2100	3000
2000	2500
2000	2000
2000	1900
1900	1800
1800	
1800	

Можно ли по этим данным сделать вывод о большей зарплате работников ресторана?

Решение. Для решения задач такого типа используются так называемые критерии различия, в частности, ^-критерий Стьюдента.

1. Введите данные: для персонала – в диапазон A1:A8; работников ресторана — в диапазон B1:Bб.

2. Выбор процедуры осуществляется из трех вариантов Z-теста. Поскольку данные не имеют попарного соответствия, число их различно и говорить о равенстве дисперсий затруднительно, выберите процедуру Двухвыборочный t-тест с различными дисперсиями.

Для реализации процедуры в пункте меню Сервис выберите строку Анализ данных и далее укажите курсором мыши на строку Двухвыборочный t-тест с различными дисперсиями.

3. В появившемся диалоговом окне задайте Интервал переменной 1. Для этого наведите указатель мыши на верхнюю ячейку столбца (A1), нажмите левую кнопку мыши и, не отпуская ее, протяните указатель мыши к нижней ячейке (A8), затем отпустите левую кнопку мыши.

4. Аналогично укажите Интервал переменной 2, то есть введите ссылку на диапазон второго столбца *B 1:B6*.

5. Далее укажите выходной диапазон. Для этого поставьте переключатель в положение Выходной диапазон (наведите указатель мыши и щелкните левой кнопкой), затем наведите указатель мыши на правое поле ввода Выходной диапазон и, щелкнув левой кнопкой мыши, указатель мыши наведите на левую верхнюю ячейку выходного диапазона (C1). Щелкните левой кнопкой мыши и нажмите кнопку ОК.

Результаты анализа. В выходном диапазоне C1:E13 появятся результаты процедуры Двухвыборочный t-тест с различными дисперсиями (рис. 6.16).

	A	B	C	D	E
1	2100	3200	Двухвыборочный t-тест с различными дисперсиями		
2	2100	3000			
3	2000	2500		Переменная 1	Переменная 2
4	2000	2000	Среднее	1962,5	2400
5	2000	1900	Дисперсия	14107,14286	356000
6	1900	1800	Наблюдения	8	6
7	1800		Гипотетическая разность средних	0	
8	1800		df	5	
9			t-статистика	-1,76998297	
10			P(T<=t) одностороннее	0,068475305	
11			t критическое одностороннее	2,015049176	
12			P(T<=t) двухстороннее	0,13695061	
13			t критическое двухстороннее	2,570577635	

Рис. 6.16. Исходные данные (A1:B8)
и результаты анализа (C1:E13) из примера

Интерпретация результатов. Средние значения заработной платы (1962 руб. для персонала и 2400 руб. для работников ресторана) довольно сильно отличаются. Тем не менее нулевая гипотеза о том, что разницы между группами нет (то есть средние выборок равны между собой), отвергнута быть не может. Это следует из того, что вероятность реализации нулевой гипотезы достаточно велика (p = 0,1389, что больше чем уровень значимости 0,05, то есть p > 0,05) и величина вероятности случайного появления анализируемых выборок ***(P(T<=t)*** двухстороннее) больше уровня значимости (а = 0,05). А это позволяет говорить, что различия между выборками могут быть случайными, то есть различия недостоверные.

Таким образом, из полученных результатов исследования вытекает, что на основании приведенных данных нельзя сделать вывод о достоверно большей зарплате работников ресторана.

УПРАЖНЕНИЯ.

1. Определите, достоверны ли различия в количестве приобретаемых туристских путевок семейными парами и отдельными туристами.

	Количество приобретаемых путевок					
Месяцы	1	2	3	4	5	6
Пары	67	75	58	89	96	94
Одиночки	43	56	78	87	85	90

2. В таблице приведены результаты группы студентов по скоростному чтению до и после специального курса по быстрому чтению.

Студент	1	2	3	4	5	6	7	8	9	10
До курса	86	83	86	70	66	90	70	85	77	86
После	82	79	91	77	68	86	81	90	85	94

Произошли ли статистически значимые изменения скорости чтения у студентов?

Дисперсионный анализ

Дисперсионный анализ предназначен для исследования задачи о действии на измеряемую случайную величину (отклик) одного или нескольких независимых факторов, имеющих несколько градаций. В однофакторном, двухфакторном и т. д. анализе влияющие на результат факторы считаются известными, и речь идет только о выяснении существенности или оценке этого влияния.

Применение дисперсионного анализа возможно, если можно предполагать соответствие выборочных групп генеральным совокупностям с нормальным распределением и независимость распределений наблюдений в группах.

Задача заключается в том, чтобы сравнить дисперсию, обусловленную случайными причинами, с дисперсией вызываемой наличием исследуемого фактора. Если они значимо различаются, то считают, что фактор оказывает статистически значимое влияние на исследуемую переменную. Значимость различий проверяется по критерию Фишера.

Влияние случайной составляющей характеризуют внутригрупповая дисперсия, а влияние изучаемого фактора — межгрупповая. Внутригрупповая дисперсия рассчитывается по формуле:

$$s_2^2 = \frac{1}{m(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - M_i)^2$$

межгрупповая:

$$s_1^2 = \frac{1}{m-1} \sum_{i=1}^{m} (M_i - M)^2$$

Здесь M — общее среднее, $M_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$, m — количество групп, n — количество элементов в группе.

В MS Excel для проведения однофакторного дисперсионного анализа используется процедура Однофакторный дисперсионный анализ.

Для проведения дисперсионного анализа необходимо:

- ввести данные в таблицу, так чтобы в каждом столбце оказались данные, соответствующие одному значению исследуемого фактора, а столбцы располагались в порядке возрастания (убывания) величины исследуемого фактора;
- выполнить команду Сервис > Анализ данных;
- в появившемся диалоговом окне Анализ данных в списке Инструменты анализа выбрать процедуру Однофакторный дисперсионный анализ, указав курсором мыши и щелкнув левой кнопкой мыши. Затем нажать кнопку ОК;
- в появившемся диалоговом окне задать Входной интервал, то есть ввести ссылку на диапазон анализируемых данных, содержащий все столбцы данных. Для этого следует навести указатель мыши на верхнюю левую ячейку диапазона данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к нижней правой ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши (рис. 6.17);
- в разделе Группировка переключатель установить в положение по столбцам;
- указать выходной диапазон, то есть ввести ссылку на ячейки, в которые будут выведены результаты анализа. Для этого следует поставить переключатель в положение Выходной интервал (навести указатель мыши и щелкнуть левой кнопкой), далее навести указатель мыши на правое поле ввода Выходной интервал и щелкнуть левой кнопкой мыши, затем указатель мыши навести на левую верхнюю ячейку выходного диапазона и щелкнуть левой кнопкой мыши. Размер выходного диапазона будет определен автоматически, и на экран будет выведено сообщение в случае возможного наложения выходного диапазона на исходные данные;
- нажать кнопку ОК.

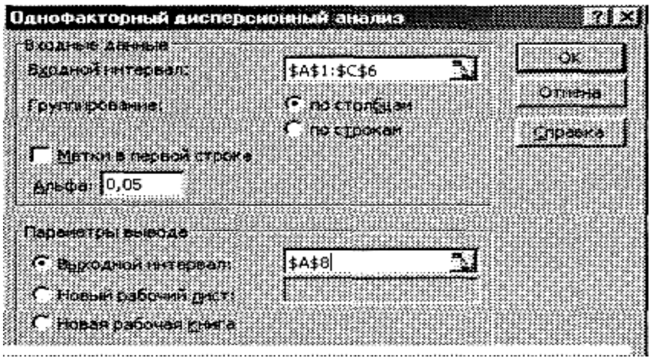


Рис. 6.17. Пример заполнения диалогового окна Однофакторный дисперсионный анализ

Результаты анализа. Выходной диапазон будет включать в себя результаты дисперсионного анализа: средние, дисперсии, критерий Фишера и другие показатели.

Интерпретация результатов. Влияние исследуемого фактора определяется по величине значимости критерия Фишера, которая находится в таблице Дисперсионный анализ на пересечении строки Между группами и столбца Р-Значение. В случаях, когда Р-Значение < 0,05, критерий Фишера значим и влияние исследуемого фактора можно считать доказанным.

Кроме рассмотренной процедуры однофакторного дисперсионного анализа, для проведения двухфакторного дисперсионного анализа в пакете анализа реализованы процедуры Двухфакторный дисперсионный анализ с повторениями и Двухфакторный дисперсионный анализ без повторений.

Пример. Необходимо выявить, влияет ли расстояние от центра города на степень заполняемости гостиниц. Пусть введены 3 уровня расстояний от центра города: 1) до 3 км, 2) от 3 до 5 км и 3) свыше 5 км. Данные заполняемости представлены в таблице.

Расстояние	Заполняемость, %					
до 3 км	92	98	89	97	90	94
от 3 до 5 км	90	86	84	91	83	82
свыше 5 км	87	79	74	85	73	77

Решение.

1. Исследуемые данные введите в рабочую таблицу Excel по столбцам: в столбец А — заполняемость гостиниц в центре города, в столбец В — гостиниц, находящихся на расстоянии от 3 до 5 км и т. д. (диапазон A1:Cб).

2. Выполните команду Сервис > Анализ данных. В появившемся диалоговом окне Анализ данных в списке Инструменты анализа щелчком мыши выберите процедуру Однофакторный дисперсионный анализ. Нажмите кнопку ОК.

3. В появившемся диалоговом окне Однофакторный дисперсионный анализ в поле Входной интервал задайте *A1:C6*. Для этого наведите указатель мыши на ячейку A1 и протяните его к ячейке Сб при нажатой левой кнопке мыши.

4. В разделе Группировка переключатель установите в положение по столбцам.

5. Далее необходимо указать выходной диапазон. Для этого поставьте переключатель в положение Выходной интервал (наведите указатель мыши и щелкните левой кнопкой), затем щелкните указателем мыши в правом поле ввода Выходной интервал, и щелчком мыши на ячейке A8 укажите расположение выходного диапазона (рис. 6.17). Нажмите кнопку ОК.

Результаты анализа. В результате будет получена таблица, показанная на рис. 6.18.

	A	B	C	D	E	F	G
8	Однофакторный дисперсионный анализ						
9							
10	ИТОГИ						
11	*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
12	Столбец 1	6	560	93,33333	13,46667		
13	Столбец 2	6	516	86	14		
14	Столбец 3	6	475	79,16667	32,96667		
15							
16							
17	Дисперсионный анализ						
18	*Источник вариации*	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
19	Между группами	602,3333	2	301,1667	14,95036	0,0002684	3,682316674
20	Внутри групп	302,1667	15	20,14444			
21							
22	Итого	904,5	17				

Рис. 6.18. Результат работы инструмента Однофакторный дисперсионный анализ

Интерпретация результатов. В таблице Дисперсионный анализ на пересечении строки Между группами и столбца Р-Значение находится величина 0,0002684. Величина Р-Зна-чение < 0,05, следовательно, критерий Фишера значим и влияние фактора расстояния от центра города на эффективность заполнения гостиниц доказано статистически.

УПРАЖНЕНИЯ.

1. Определите, влияет ли фактор образования на уровень зарплаты в гостинице на основании следующих данных:

Образование	Зарплата сотрудника					
высшее	3200	3000	2600	2000	1900	1900
среднее спец	2600	2000	2000	1900	1800	1700
среднее	2000	2000	1900	1800	1700	1700

Корреляционный анализ

Важным разделом статистического анализа является корреляционный анализ, служащий для выявления взаимосвязей между выборками.

Коэффициент корреляции

Выявление взаимосвязей. Одна из задач статистического исследования состоит в изучении связи между некоторыми наблюдаемыми переменными. Знание взаимозависимостей отдельных признаков дает возможность решать одну из кардинальных задач любого научного исследования: возможность предвидеть, прогнозировать развитие ситуации при изменении конкретных характеристик объекта исследования. Например, основное содержание любой экономической политики в конечном счете может быть сведено к регулированию экономических переменных, осуществляемому на базе выявленной тем или иным образом информации об их взаимовлиянии. Поэтому проблема изучения взаимосвязей показателей различного рода является одной из важнейших в статистическом анализе.

Обычно взаимосвязь между выборками носит не функциональный, а вероятностный (или стохастический) характер. В этом случае нет строгой, однозначной зависимости между величинами. При изучении стохастических зависимостей различают *корреляцию и регрессию*.

Регрессионный анализ устанавливает формы зависимости между случайной величиной У и значениями одной или нескольких переменных величин.

Корреляционный анализ состоит в определении степени связи между двумя случайными величинами *X* и *Y*. В качестве меры такой связи используется коэффициент корреляции. Коэффициент корреляции оценивается по выборке объема *n* связанных пар наблюдений (x_i, y_i)из совместной генеральной совокупности *X* и *Y*. Для оценки степени взаимосвязи наибольшее распространение получил *коэффициент линейной корреляции (Пирсона)*, предполагающий нормальный закон распределения наблюдений.

Коэффициент корреляции **(*R, r*)** — параметр, характеризующий степень линейной взаимосвязи между двумя выборками. Коэффициент корреляции изменяется от -1 (строгая обратная линейная зависимость) до 1 (строгая прямая пропорциональная зависимость). При значении 0 линейной зависимости между двумя выборками нет. Здесь под прямой зависимостью понимают зависимость, при которой увеличение или уменьшение значения одного признака ведет, соответственно, к увеличению или уменьшению второго. Например, при увеличении температуры возрастает давление газа, а при уменьшении — снижается (при постоянном объеме). При обратной зависимости увеличение одного признака приводит к уменьшению второго и наоборот. Примером обратной корреляционной зависимости может служить связь между температурой воздуха на улице и количеством топлива, расходуемого на обогрев помещения.

Выборочный коэффициент линейной корреляции между двумя случайными величинами *X* и У рассчитывается по формуле

$$r = \frac{\sum (x - M_x)(y - M_y)}{\sqrt{\sum (x - M_x)^2 (y - M_y)^2}}$$

Коэффициент корреляции является безразмерной величиной и его значение не зависит от единиц измерения случайных величин *X* и У.

На практике коэффициент корреляции принимает некоторые промежуточные значения между 1 и -1 (рис. 6.19). Для оценки степени взаимосвязи можно руководствоваться следующими эмпирическими правилами. Если коэффициент корреляции (г) по абсолютной величине (без учета знака) больше, чем 0,95, то принято считать, что между параметрами существует практически линейная зависимость (прямая — при положительном *r* и обратная — при отрицательном *r*). Если коэффициент корреляции $|r|$ лежит в диапазоне от 0,8 до 0,95, говорят о сильной степени линейной связи между параметрами. Если $0{,}6 < |r| < 0{,}8$, говорят о наличии линейной связи между параметрами. При $|r| < 0{,}4$ обычно считают, что линейную взаимосвязь между параметрами выявить не удалось.

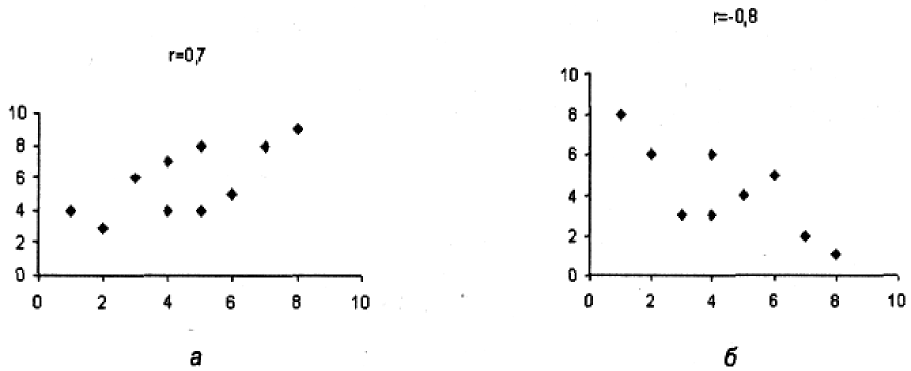


Рис. 6.19. Примеры прямой (r =0,7,а) и обратной (r =-0,8,б) корреляционной зависимости

- В MS Excel для вычисления парных коэффициентов линейной корреляции используется специальная функция КОРРЕЛ. Параметрами функции являются КОРРЕЛ(*массив1;массив2)*, где: *массив1* — это диапазон ячеек первой случайной величины;

- *массив2* — это второй интервал ячеек со значениями второй случайной величины.

Пример. Имеются результаты семимесячных наблюдений реализации путевок двух туристских маршрутов тура А и тура В.

Тур А	Тур В
120	20
121	15
105	18
92	16
113	19
90	16
80	15

Необходимо определить, имеется ли взаимосвязь между количеством продаж путевок обоих маршрутов.

Решение. Для выявления степени взаимосвязи прежде всего необходимо ввести данные в рабочую таблицу. Откройте новую рабочую таблицу. Введите в ячейку А1 слова *Тур А*. Затем в ячейки А2:А8 — соответствующие значения числа продаж. В ячейки В1:В8 введите название и значения для тура В. Затем вычисляется значение коэффициента корреляции между выборками. Для этого табличный курсор установите в свободную ячейку (А9). На панели инструментов нажмите кнопку Вставка функции (f_x). В появившемся диалоговом окне Мастер функций выберите категорию Статистические и функцию КОРРЕЛ, после чего нажмите кнопку ОК. Появившееся диалоговое окно КОРРЕЛ за серое поле мышью отодвиньте вправо на 1-2 см от данных (при нажатой левой клавише). Указателем мыши введите диапазон данных *Тур А* в поле Массив 1 *(А2:А8)*. В поле Массив 2 введите диапазон данных *Тур В (В2:В8)*. Нажмите кнопку ОК. В ячейке А9 появится значение коэффициента корреляции — *0,995493*. Значение коэффициента корреляции больше чем 0,95. Значит, можно говорить о том, что в течение периода наблюдения имелась высокая степень прямой линейной взаимосвязи между количествами проданных путевок обоих маршрутов (r = 0,995493).

Корреляционная матрица

При большом числе наблюдений, когда коэффициенты корреляции необходимо последовательно вычислять из нескольких рядов числовых данных, для удобства получаемые коэффициенты сводят в таблицы, называемые корреляционными матрицами.

Корреляционная матрица — это квадратная (или прямоугольная) таблица, в которой на пересечении соответствующих строки и столбца находится коэффициент корреляции между соответствующими параметрами.

В MS Excel для вычисления корреляционных матриц используется процедура Корреляция. Процедура позволяет получить корреляционную матрицу, содержащую коэффициенты корреляции между различными параметрами.

Для реализации процедуры необходимо:

- выполнить команду Сервис > Анализ данных;

- в появившемся списке Инструменты анализа выбрать строку Корреляция и нажать кнопку ОК;

- в появившемся диалоговом окне указать Входной интервал, то есть ввести ссылку на ячейки, содержащие анализируемые данные. Для этого следует навести указатель мыши на левую верхнюю ячейку данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к правой нижней ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши. Входной интервал должен содержать не менее двух столбцов;

- в разделе Группировка переключатель установить в соответствии с введенными данными;

- указать выходной диапазон, то есть ввести ссылку на ячейки, в которые будут выведены результаты анализа. Для этого следует поставить флажок в левое поле Выходной интервал (навести указатель мыши и щелкнуть левой кнопкой), далее навести указатель мыши на правое поле ввода Выходной интервал и щелкнуть левой кнопкой мыши, затем указатель мыши навести на левую верхнюю ячейку выходного диапазона и щелкнуть левой кнопкой мыши. Размер выходного диапазона будет определен автоматически, и на экран будет выведено сообщение в случае возможного наложения выходного диапазона на исходные данные (рис. 6.20).

- нажать кнопку ОК.

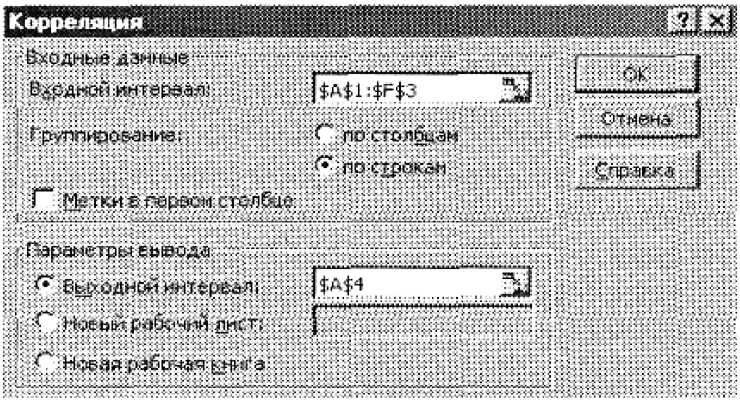


Рис. 6.20. Пример установки параметров корреляционного анализа

Результаты анализа. В выходной диапазон будет выведена корреляционная матрица, в которой на пересечении каждых строки и столбца находится коэффициент корреляции между соответствующими параметрами. Ячейки выходного диапазона, имеющие совпадающие координаты строк и столбцов, содержат значение 1, так как каждый столбец во входном диапазоне полностью коррелирует с самим собой.

Интерпретация результатов. Рассматривается отдельно каждый коэффициент корреляции между соответствующими параметрами. Его числовое значение оценивается по эмпирическим правилам, изложенным в разделе «Коэффициент корреляции». Отметим, что хотя в результате будет получена треугольная матрица, корреляционная матрица симметрична, и коэффициенты корреляции $r_{ij} = r_{ji}$

Пример. Имеются ежемесячные данные наблюдений за состоянием погоды и посещаемостью музеев и парков.

Число ясных дней	Количество посетителей музея	Количество посетителей парка
8	495	132
14	503	348
20	380	643
25	305	865
20	348	743
15	465	541

Необходимо определить, существует ли взаимосвязь между состоянием погоды и посещаемостью музеев и парков.

Решение. Для выполнения корреляционного анализа введите в диапазон A1:G3 исходные данные (рис. 6.21).

Затем в меню Сервис выберите пункт Анализ данных и далее укажите строку Корреляция. В появившемся диалоговом окне укажите Входной интервал *B1:G3*. Укажите, что данные рассматриваются по строкам. Укажите выходной диапазон. Для этого поставьте флажок в левое поле Выходной интервал и в правое поле ввода Выходной интервал введите *A4* (рис. 6.20). Нажмите кнопку ОК.

	A	B	C	D	E	F	G
1	Ясные дни	8	14	20	25	20	15
2	Посещаемость музея	495	503	380	305	348	465
3	Посещаемость парка	132	348	643	865	743	541

Рис. 6.21. Исходные данные из примера

	A	B	C	D
4		*Строка 1*	*Строка 2*	*Строка 3*
5	Строка 1	1		
6	Строка 2	-0,92185	1	
7	Строка 3	0,974576	-0,91938	1

Рис. 6.22. Результаты вычисления корреляционной матрицы из примера

Результаты анализа. В выходном диапазоне получаем корреляционную матрицу (рис. 6.22).

Интерпретация результатов. Из таблицы видно, что корреляция между состоянием погоды и посещаемостью музея равна -0,92, а между состоянием погоды и посещаемостью парка — 0,97, между посещаемостью парка и музея – r =-0,92.

Таким образом, в результате анализа выявлены зависимости: сильная степень обратной линейной взаимосвязи между посещаемостью музея и количеством солнечных дней (r =-0,92) и практически линейная (очень сильная прямая) связь между посещаемостью парка и состоянием погоды (r = 0,97). Между посещаемостью музея и парка имеется сильная обратная взаимосвязь (r = -0,92).

Подразумевается, что в пустых клетках в правой верхней половине таблицы находятся те же коэффициенты корреляции, что и в нижней левой (симметрично расположенные относительно диагонали).

УПРАЖНЕНИЯ.

1. Определите, имеется ли взаимосвязь между рождаемостью и смертностью (количество на 1000 человек) в Санкт-Петербурге:

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

2. Определите, имеется ли взаимосвязь между годовым уровнем инфляции (%), ставкой рефинансирования (%) и курсом доллара (руб./$), по следующим данным ежегодных наблюдений:

Уровень инфляции	Ставка рефинансирования	Курс $
84	85	6,3
45	55	14
56	65	20
34	40	28
23	28	29

Регрессионный анализ

При исследовании взаимосвязей между выборками помимо корреляции различают также и регрессию. Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных. Наряду с корреляционным анализом еще одним инструментом изучения стохастических зависимостей является регрессионный анализ.

Регрессионный анализ устанавливает формы зависимости между случайной величиной У (зависимой) и значениями одной или нескольких переменных величин (независимых), причем значения последних считаются точно заданными. Такая зависимость обычно определяется некоторой математической моделью (уравнением регрессии), содержащей несколько неизвестных параметров. В ходе регрессионного анализа на основании выборочных данных находят оценки этих параметров, определяются статистические ошибки оценок или границы доверительных интервалов и проверяется соответствие (адекватность) принятой математической модели экспериментальным данным.

В линейном регрессионном анализе связь между случайными величинами предполагается линейной. В самом простом случае в линейной регрессионной модели имеются две переменные X и Y. И требуется по n парам наблюдений (X_1, $У_1$), (X_2, $У_2$), ...,(X_n,$У_n$)построить (подобрать) прямую линию, называемую линией регрессии, которая «наилучшим образом» приближает наблюдаемые значения. Уравнение этой линии $У = aX + b$ является регрессионным уравнением. С помощью регрессионного уравнения можно предсказать ожидаемое значение зависимой величины Y_0, соответствующее заданному значению независимой переменной X_0.

Линейный регрессионный анализ заключается в подборе графика и его уравнения для набора наблюдений. В регрессионном анализе все признаки (переменные), входящие в уравнение, должны иметь непрерывную, а не дискретную природу.

Когда рассматривается зависимость между одной зависимой переменной У и несколькими независимыми X_1, X_2,..., Xn, говорят о множественной линейной регрессии. В этом случае регрессионное уравнение имеет вид

$$У = a_0 + a_1X_1 + a_2X_2 + ... + a_nXn,$$

где $a_1, a_2, \ldots, a_n$— требующие определения коэффициенты при независимых переменных $X_1, X_2, \ldots, Xn$; a_0 — константа.

Мерой эффективности регрессионной модели является *коэффициент детерминации* R^2 (R-квадрат). Коэффициент детерминации (R-квадрат) определяет, с какой степенью точности полученное регрессионное уравнение описывает (аппроксимирует) исходные данные.

Исследуется также значимость регрессионной модели с помощью *F*-критерия (Фишера). Если величина F-критерия значима ($p < 0{,}05$), то регрессионная модель является значимой.

Достоверность отличия коэффициентов a_1, a_2, a_3 ..., an от нуля проверяется с помощью критерия Стьюдента. В случаях, когда $p > 0{,}05$, коэффициент может считаться нулевым, а это означает, что влияние соответствующей независимой переменной на зависимую переменную недостоверно, и эта независимая переменная может быть исключена из уравнения,

В MS Excel экспериментальные данные аппроксимируются линейным уравнением до 16 порядка:

$$У= a_0 + a_1X_l + a_2X_2 + \ldots + a_{16}X_{16},$$

где Y — зависимая переменная,$X1,\ldots,X_{16}$ — независимые переменные, $a_0, a_1,\ldots, a_{16}$ — искомые коэффициенты регрессии.

Для получения коэффициентов регрессии используется процедура Регрессия из пакета анализа. Кроме того, могут быть использованы функция ЛИНЕЙН для получения параметров регрессионного уравнения и функция ТЕНДЕНЦИЯ для получения предсказанных значений У в требуемых точках Для реализации процедуры Регрессия необходимо:

- выполнить команду Сервис > Анализ данных;

- в появившемся диалоговом окне Анализ данных в списке Инструменты анализа выбрать строку Регрессия, указав курсором мыши и щелкнув левой кнопкой мыши. Затем нажать кнопку ОК;

- в появившемся диалоговом окне задать *Входной интервал Y*, то есть ввести ссылку на диапазон анализируемых зависимых данных, содержащий один столбец данных. Для этого следует навести указатель мыши на верхнюю ячейку столбца зависимых данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к нижней ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши;

- указать *Входной интервал X*, то есть ввести ссылку на диапазон независимых данных, содержащий до 16 столбцов анализируемых данных. Для этого следует навести указатель мыши на поле ввода Входной интервал X и щелкнуть левой кнопкой мыши, затем навести указатель мыши на верхнюю левую ячейку диапазона независимых данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к нижней правой ячейке, содержащей анализируемые данные, затем отпустить левую кнопку мыши;

- указать выходной диапазон, то есть ввести ссылку на ячейки, в которые будут выведены результаты анализа. Для этого следует поставить переключатель в положение Выходной интервал (навести указатель мыши и щелкнуть левой кнопкой), далее навести указатель мыши на правое поле ввода Выходной интервал и щелкнуть левой кнопкой мыши, затем указатель мыши навести на левую верхнюю ячейку выходного диапазона и щелкнуть левой кнопкой мыши (рис. 6.23). Размер выходного диапазона будет определен автоматически, и на экран будет выведено сообщение в случае возможного наложения выходного диапазона на исходные данные;

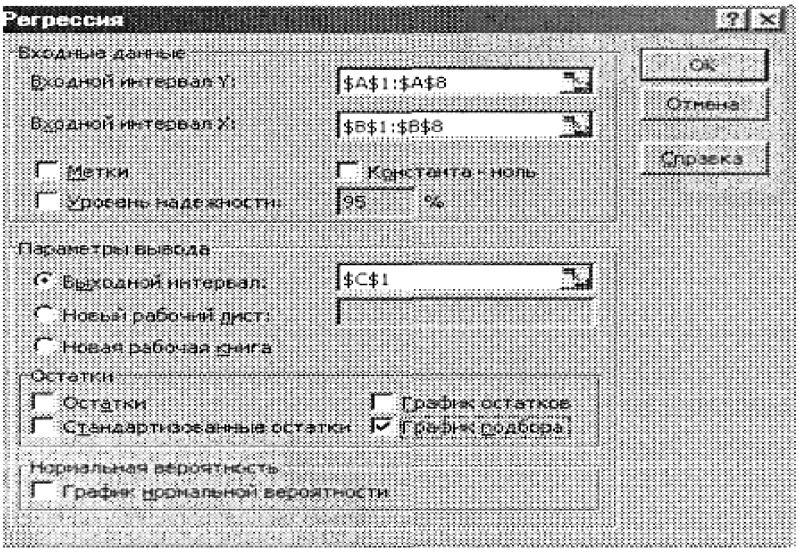


Рис. 6.23. Пример заполнения диалогового окна Регрессия

-если необходимо визуально проверить отличие экспериментальных точек от предсказанных по регрессионной модели, следует установить флажок в поле График подбора;

- нажать кнопку ОК.

Результаты анализа. Выходной диапазон будет включать в себя результаты дисперсионного анализа, коэффициенты регрессии, стандартную погрешность вычисления У, среднеквадратичные отклонения, число наблюдений, стандартные погрешности для коэффициентов.

Интерпретация результатов. Значения коэффициентов регрессии находятся в столбце Коэффициенты и соответствуют:

-У-пересечение – a_0;

- переменная *XI* — a_1;

- переменная *X2* — a_2 и т. д.

В столбце Р-Значение приводится достоверность отличия соответствующих коэффициентов от нуля. В случаях, когда $P >$ 0,05, коэффициент может считаться нулевым, что означает, что соответствующая независимая переменная практически не влияет на зависимую переменную.

Приводимое значение R-квадрат (коэффициент детерминации) определяет, с какой степенью точности полученное регрессионное уравнение аппроксимирует исходные данные. Если R-квадрат > 0,95, говорят о высокой точности аппроксимации (модель хорошо описывает явление). Если R-квадрат лежит в диапазоне от 0,8 до 0,95, говорят об удовлетворительной аппроксимации (модель в целом адекватна описываемому явлению). Если R-квадрат < 0,6, принято считать, что точность аппроксимации недостаточна и модель требует улучшения (введения новых независимых переменных, учета нелинейностей и т. д.).

Пример. В отделе снабжения гостиницы имеется информация об изменении стоимости стирального порошка за длительный период времени. Сопоставляя его с изменениями курса доллара за этот же период времени, можно построить регрессионное уравнение. Ниже приведены стоимость пачки стирального порошка (в руб.) и соответствующий курс доллара (руб./USD).

N	Порошок	Курс
1	5	6,3
2	7	9
3	9	12
4	12	15
5	15	19
6	16	21
7	20	25
8	25	29,3

Необходимо па основании этих данных построить регрессионное уравнение, позволяющее по курсу доллара определять предполагаемую стоимость пачки стирального порошка.

Решение.

1. Введите данные в рабочую таблицу: стоимость пачки порошка – в диапазон A1:A8; курс доллара в диапазон B1:B8 (заметим, что знаку запятой, отделяющей целую часть от дробной, соответствует «запятая»).

2. В пункте меню Сервис выберите строку Анализ данных и далее укажите курсором мыши на строку Регрессия.

3. В появившемся диалоговом окне (рис. 6.23) задайте *Входной интервал Y*. Для этого наведите указатель мыши на верхнюю ячейку столбца зависимых данных (A1), нажмите левую кнопку мыши и, не отпуская ее, протяните указатель мыши к нижней ячейке (A8), затем отпустите левую кнопку мыши. (Обратите внимание, что зависимые данные — это те данные, которые предполагается вычислять.)

4. Так же укажите *Входной интервал X*, то есть введите ссылку па диапазон независимых данных B1:B8. (Независимые данные — это те данные, которые будут измеряться или наблюдаться.)

5. Установите флажок в поле График подбора.

6. Далее укажите выходной диапазон. Для этого поставьте переключатель в положение Выходной интервал (наведите указатель мыши и щелкните левой кнопкой), затем наведите указатель мыши на правое поле ввода Выходной интервал и, щелкнув левой кнопкой мыши, указатель мыши наведите на левую верхнюю ячейку выходного диапазона (C1). Щелкните левой кнопкой мыши (рис. 6.23).

Нажмите кнопку ОК.

Результаты анализа. В выходном диапазоне появятся следующие результаты и график подбора (рис. 6.24).

ВЫВОД ИТОГОВ								
Регрессионная статистика								
Множественный R	0,99601535							
R-квадрат	0,99204353							
Нормированный R-квад	0,99072333							
Стандартная ошибка	0,65108335							
Наблюдения	8							
Дисперсионный анализ								
	df	*SS*	*MS*	*F*	*Значимость F*			
Регрессия	1	317,3315	317,3315	746,5832	1,57573E-07			
Остаток	6	2,543457	0,42391					
Итого	7	319,875						
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	-0,830909	0,576322	-1,44174	0,199462	-2,241118321	0,579301	-2,24112	0,57930
Переменная X 1	0,84661253	0,030943	27,36025	1,58E-07	0,770397294	0,922328	0,77089	0,92232

Рис. 6.24. Результаты анализа и график соответствия экспериментальных точек и предсказанных по регрессионной модели из примера

Интерпретация результатов. В таблице Дисперсионный анализ оценивается общее качество полученной модели: ее достоверность по уровню значимости критерия Фишера — *p*, который должен быть меньше, чем 0,05 (строка Регрессия, столбец Значимость F, в примере- 1.58E-07 (0,000000158), то есть *p* = 0,000000158 и модель значима) и степень точности описания моделью процесса — R-квадрат (вторая строка сверху в таблице Регрессионная статистика, в примере R-квадрат = 0,992). Поскольку R-квадрат > 0,95, можно говорить о высокой точности аппроксимации (модель хорошо описывает явление (рис. 6.24)).

Далее необходимо определить значения коэффициентов модели. Они определяются из таблицы в столбце Коэффициенты — в строке Y-пересечение приводится свободный член; в строках соответствующих переменных приводятся значения коэффициентов при этих переменных. В столбце р-значение приводится достоверность отличия соответствующих коэффициентов от нуля. В случаях, когда *p* > 0,05, коэффициент может считаться нулевым. Это означает, что соответствующая независимая переменная практически не влияет на зависимую переменную и коэффициент может быть убран из уравнения.

Отсюда выражение для определения стоимости пачки порошка в рублях будет иметь следующий вид: *-0,83 + 0,847*(Курс доллара, руб./USD).*

Полученная модель с высокой точностью позволяет определять стоимость пачки стирального порошка (R^2 = 99,2%).

Воспользовавшись полученным уравнением, можно рассчитать ожидаемую стоимость пачки стирального порошка при изменениях курса доллара. Например, для расчета при курсе доллара 35 руб./USD необходимо поставить табличный курсор в любую свободную ячейку (A10); ввести с клавиатуры знак =, щелкнуть указателем мыши по ячейке D17, ввести с клавиатуры знак +, щелкнуть по ячейке D18, ввести с клавиатуры знак * и число *35*. В результате в ячейке A10 будет получена ожидаемая стоимость пачки порошка — *28,8* руб.

Пример. Построить регрессионную модель для предсказания изменений уровня заболеваемости органов дыхания (У) в зависимости от содержания в воздухе двуокиси углерода *(X_1)* и степени запыленности (X_2). В таблице приведены данные наблюдений в течение 29 месяцев.

X_1	X_2	Y
1	1,3	1160
1	1,3	1155
1,1	1,4	1158
1,1	1,4	1157
1,1	1,5	1160
1,1	1,5	1161
1	1,4	1157
1	1,5	1159
1,2	1,6	1256
1,2	1,7	1260
0,6	1	1040
0,6	1	1039
0,7	1,1	1039
0,7	1,15	1040
0,75	1,2	1040
0,7	1,2	1039
0,7	1,3	1040
0,7	1,3	1039
0,8	1,4	1140
0,8	1,4	1138

x_1	x_2	Y
0,78	1,5	1240
0,8	1,5	1239
0,78	1,5	1241
0,78	1,6	1240
0,8	1,7	1239
0,8	1,8	1239
0,75	1,8	1240
0,78	1,9	1238
0,75	1,9	1238

Решение.

1. Введите данные наблюдений в диапазон A1:C30 рабочей таблицы Excel.
2. В пункте меню Сервис выберите строку Анализ данных и далее укажите курсором мыши на строку Регрессия. Нажмите кнопку ОК.
3. В появившемся диалоговом окне задаем. *Входной интервал Y*. Для этого наведите указатель мыши на верхнюю ячейку столбца зависимых данных (C2), нажмите левую кнопку мыши и, не отпуская ее, протяните указатель мыши к нижней ячейке (C30), затем отпустите левую кнопку мыши. (Обратите внимание, что зависимые данные — это те данные, которые предполагается вычислять).
4. Так же укажите *Входной интервал X*, то есть введите ссылку на диапазон независимых данных A2:B30. (Независимые данные – это те данные, которые будут измеряться или наблюдаться).
5. Установите флажок в поле График подбора.
6. Далее укажите выходной диапазон. Для этого поставьте переключатель в положение Выходной интервал (наведите указатель мыши и щелкните левой кнопкой), затем наведите указатель мыши на правое поле ввода Выходной интервал и, щелкнув левой кнопкой мыши, указатель мыши наведите на левую верхнюю ячейку выходного диапазона (D1). Щелкните левой кнопкой мыши. Нажмите кнопку ОК.
7. В выходном диапазоне появятся результаты регрессионного анализа и графики предсказанных точек (рис. 6.25).

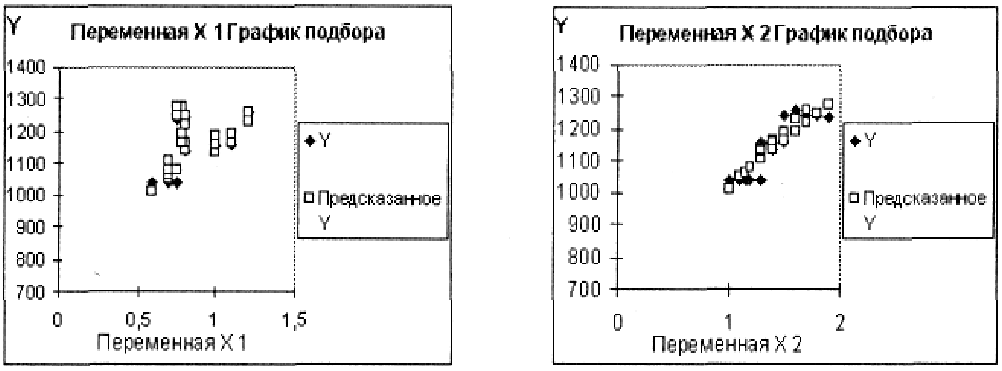


Рис. 6.25. Графики расположения фактических и предсказанных точек

Интерпретация результатов. В таблице Дисперсионный анализ оценивается достоверность полученной модели по уровню значимости критерия Фишера (строка Регрессия, столбец Значимость F, в примере — *1,4E-09 (1,4*10-9),* то естьр « 0,05 и модель значима) и степень описания моделью процесса — R-квадрат (вторая строка сверху в таблице Регрессионная статистика, в примере R-квадрат = 0,89). Поскольку -R-квадрат > 0,8, можно говорить о довольно высокой точности аппроксимации (модель хорошо описывает зависимость заболеваемости от содержания углекислого газа и запыленности воздуха (рис. 6.25)).

Далее необходимо определить значения коэффициентов модели. Они определяются из таблицы в столбце Коэффициенты — в строке Y-пересечение приводится свободный член a_0 = 682; в строках соответствующих переменных приводятся значения коэффициентов при этих переменных a_1= 91 и a_2 = 275. В столбце р-значение приводится достоверность отличия соответствующих коэффициентов от нуля. Все коэффициенты значимы, то есть $p < 0{,}05$, и коэффициенты могут считаться не равными нулю.

Поэтому выражение для определения уровня заболеваемости органов дыхания в зависимости от содержания углекислого газа и пыли в воздухе будет иметь вид:

$$Y = 682 + 91 * X_1 + 275 * X_2$$

УПРАЖНЕНИЯ.

1. Постройте зависимость зарплаты (руб.) от возраста сотрудника гостиницы по следующим данным:

Возраст	Зарплата
20	800
50	2500
45	2500
40	2000
25	1200
30	1800

23. Постройте зависимость жизненной емкости легких в литрах (Y) от роста в метрах (X_1) и возраста в годах (X_2) для группы мужчин:

X_1	X_2	Y
1,85	18	5,4
1,8	25	5,7
1,75	20	4,8
1,7	24	5,1
1,68	21	4,5
1,73	19	4,8
1,77	22	5,1
1,81	23	5,6
1,76	18	4,7

2. Определите должное значение жизненной емкости легких для мужчины возраста 22 лет и роста 183 см из регрессионного уравнения, полученного в предыдущем упражнении.

3. Имеются данные о цене на нефть x (ден. ед.) и индексе акций нефтяных компаний y (усл. ед.):

x	y
17,28	537
17,05	534
18,30	550
18,80	555
19,20	560
18,50	552

Постройте зависимость индекса акций нефтяных компаний от цены на нефть.

Технический редактор – Т.И. Медведева
Отпечатано в издательско-полиграфическом центре
ФГОУ ВПО МичГАУ
Подписано в печать 19.03.07. г. Формат 60х84 $^{1}/16$,

Заказ №
Бумага офсетная № 1. Усл.печ.л. 7,7   Тираж 75 экз.   Ризограф

Издательско-полиграфический центр
Мичуринского государственного аграрного университета

393760, Тамбовская обл., г. Мичуринск, ул. Интернациональная, 101,
тел. +7 (47545) 5-55-12
E-mail: vvdem@mgau.ru