ТЕМА 3. ОСНОВНЫЕ ВИДЫ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН

ПОНЯТИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

- Случайной величиной (СВ) называют величину, которая в результате наблюдения принимает то или иное значение, заранее неизвестное и зависящее от случайных обстоятельств с определенной вероятностью

Объем ВНП, количество реализованной продукции, прибыль фирмы, размер чистого экспорта за год и т. д. являются случайными величинами.

ВИДЫ СЛУЧАЙНЫХ ВЕЛИЧИН

- **Дискретные** – область значений счетная
- **Непрерывные** - область значений бесконечна

Например, можно считать, что число покупателей в магазине, побывавших там в течение дня; число автомобилей, ремонтируемых еженедельно в данной мастерской; число находящихся в аэропорту самолетов являются дискретными СВ. Однако большинство СВ, рассматриваемых в экономике, имеют настолько большое число возможных значений, что их удобнее представлять в виде непрерывных СВ. Например, курсы валют, доход, объемы ВНП, ВВП и т. п. обычно рассматриваются как непрерывные СВ.

Распределением случайной величины

- называется закономерность встречаемости разных ее значений (Плохинский Н.А., 1970, с. 12).

ТИПЫ РАСПРЕДЕЛЕНИЙ

- *Равномерное распределение* — когда все значения встречаются одинаково (или почти одинаково) часто.
- *Симметричное распределение* — когда одинаково часто встречаются крайние значения.
- *Нормальное распределение* — симметричное распределение, у которого крайние значения встречаются редко и частота постепенно повышается от крайних к серединным значениям признака.
- *Асимметричные распределения* — *левосторонние* (с преобладанием частот малых значений), *правосторонние* (с преобладанием частот больших значений).

ЗАКОН РАСПРЕДЕЛЕНИЯ СВ (табличная форма)

Для описания СВ необходимо установить соответствие между всеми возможными значениями СВ и их вероятностями. Такое соответствие называется *законом распределения СВ*.

Его можно задать таблично, аналитически (в виде формулы) либо графически.

При табличном задании закона распределения дискретной СВ X первая строка таблицы содержит ее возможные значения, а вторая – их вероятности:

X	x_1	x_2	……	x_k
p_i	p_1	p_2	……	p_k

Обычно $x_1 < x_2 < ... < x_k$. Обязательно $p_1 + p_2 + \ldots + p_k = 1$.

ЗАКОН РАСПРЕДЕЛЕНИЯ СВ (аналитическая форма)

Аналитически СВ задается либо функцией распределения, либо плотностью вероятностей.

Функцией распределения СВ X называют функцию F(x), определяющую вероятность того, что случайная величина X принимает значение меньше, чем x, т. е.

$$F(x) = P(X < x). \tag{1.2}$$

Из определения вытекают *свойства функции распределения*:

1. $0 \le F(x) \le 1$;
2. F(x) – неубывающая функция, т. е. $(x_1 < x_2) \Rightarrow F(x_1) \le F(x_2)$;
3. $\lim\limits_{x \to -\infty} F(x) = 0, \quad \lim\limits_{x \to +\infty} F(x) = 1$;
4. $P(a \le X < b) = F(b) - F(a)$;
5. $P(X \ge x) = 1 - F(x)$;
6. Если возможные значения СВ X принадлежат отрезку [a, b], то

$$F(x) = \begin{cases} 0, & \text{если } x \le a; \\ 1, & \text{если } x > b. \end{cases}$$

Пример 1.1. На станции технического обслуживания анализируются затраты времени на ремонт автомобилей. На основании данных, полученных по 100 автомобилям, выяснилось, что для 25 из них требуется 1 ч для проведения профилактических работ. Мелкий ремонт требуется для 40 автомобилей, что занимает 2 ч. Для 20 автомобилей требуется ремонт с заменой отдельных узлов, что занимает в среднем 5 ч. 10 автомобилей могут быть отремонтированы за 10 ч. Для 5 автомобилей необходимое время ремонта составляет 20 ч. Построить закон распределения СВ X – времени обслуживания случайно выбранного автомобиля.

X	1	2	5	10	20
p_i	0.25	0.40	0.20	0.10	0.05

ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ

Для примера 1.1 функция распределения F(x) и ее график имеют вид:

$$F(x) = \begin{cases} 0, & \text{при } x \le 1; \\ 0.25, & \text{при } 1 < x \le 2; \\ 0.65, & \text{при } 2 < x \le 5; \\ 0.85, & \text{при } 5 < x \le 10; \\ 0.95, & \text{при } 10 < x \le 20; \\ 1, & \text{при } x > 20. \end{cases}$$

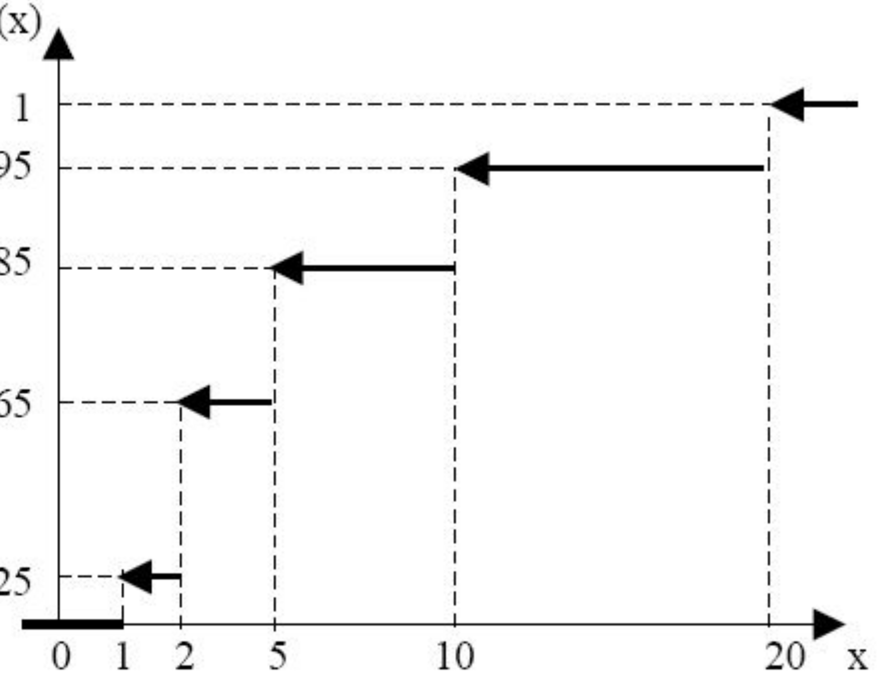


Рис. 1.1

НЕПРЕРЫВНАЯ СВ

Для непрерывной СВ нельзя определить вероятность того, что она примет некоторое конкретное значение (точечную вероятность). Так как в любом интервале содержится бесконечное число значений, то вероятность выпадения одного из них асимптотически равна нулю. В результате непрерывную СВ нельзя задать таблично. Однако функция распределения может быть использована для описания непрерывной СВ. При этом она является непрерывной неубывающей функцией, изменяющейся от 0 до 1

Плотностью вероятности (плотностью распределения вероятностей) непрерывной СВ X называют функцию

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \le X < x + \Delta x)}{\Delta x}. \qquad (1.3)$$

Из свойства 4 функции распределения имеем

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x). \qquad (1.4)$$

Итак, плотность вероятности равна производной от функции распределения (поэтому иногда ее называют дифференциальной функцией распределения).

Свойства плотности вероятности

1. $f(x) \ge 0$;
2. $P(a \le X \le b) = \int\limits_a^b f(x)dx$;
3. для непрерывной СВ справедливы равенства

   $P(\ a \le X \le b) = P(\ a < x < b) = P(\ a \le X < b) = P(a < X \le b)$;
4. если $f(x)$ – плотность вероятности непрерывной СВ, то

   функция распределения $F(x) = \int\limits_{-\infty}^{x} f(t)dt$;
5. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ (условие нормировки).

На рис. 1.2 и 1.3 изображены характерные графики функции распределения и плотности вероятности непрерывной СВ.

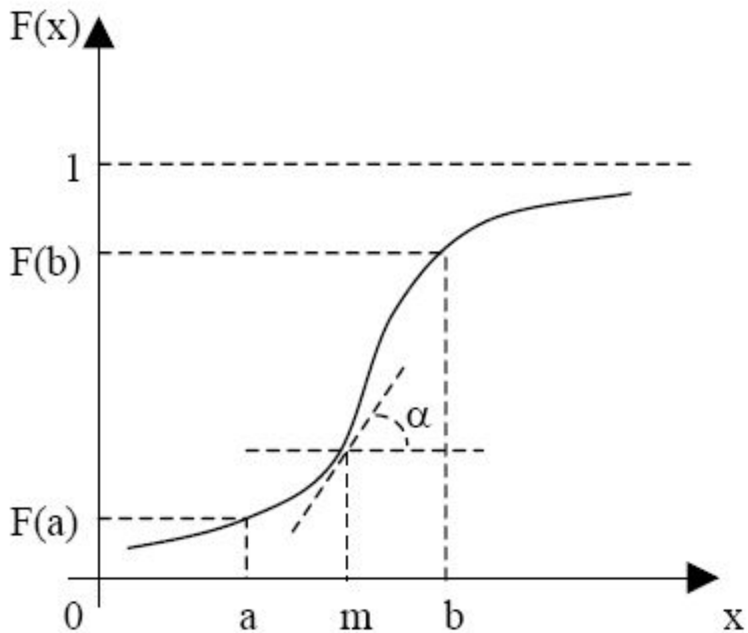


Рис. 1.2

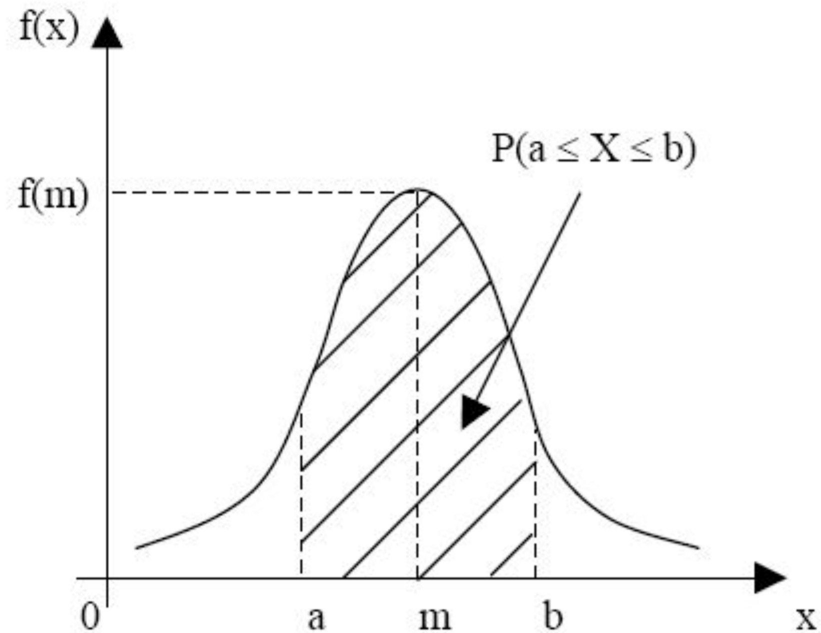
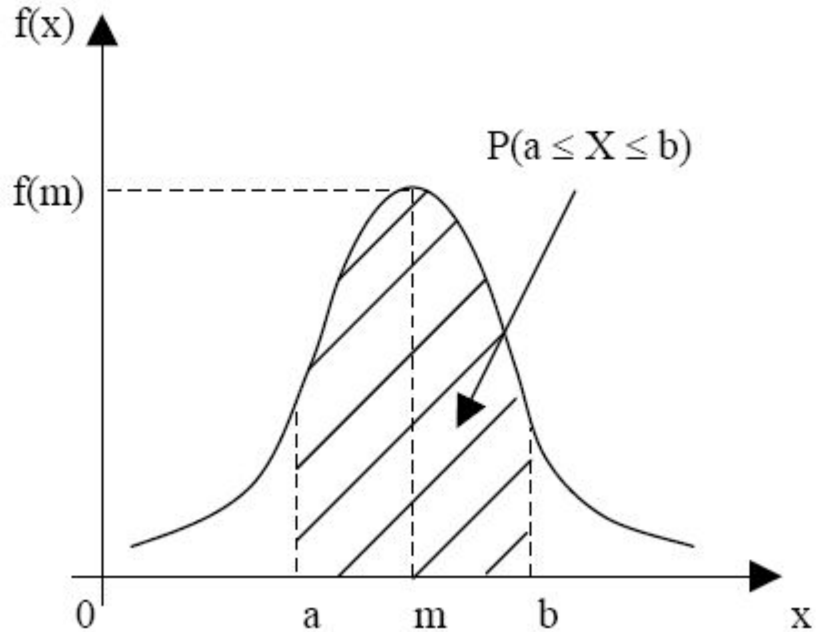


Рис. 1.3

СМЫСЛ ПЛОТНОСТИ ВЕРОЯТНОСТИ



Площадь под графиком кривой
плотности вероятности f(x) равна единице. Площадь заштрихованной фигуры $S = \int_a^b f(x)dx = P(a \le X \le b)$. Вероятность попадания в “хвосты” распределения СВ равна $1 - P(a \le X \le b)$.

Таким образом, с помощью плотности вероятности f(x) непрерывной СВ X можно определить вероятность ее попадания в заданный интервал, т. е. $P(a \le X \le b)$, что имеет большое прикладное значение.

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СВ

Во многих практических случаях информация о СВ, которую дает закон распределения, функция распределения или плотность вероятностей, является избыточной. Иногда даже выгоднее пользоваться числами, которые описывают СВ суммарно. Такие числа называют *числовыми характеристиками СВ*. Условно их подразделяют на характеристики положения (математическое ожидание, мода, медиана, начальные моменты различных порядков) и характеристики рассеивания (дисперсия, среднее квадратическое отклонение, центральные моменты различных порядков). Важнейшими из них являются математическое ожидание, дисперсия, среднее квадратическое отклонение.

Параметры распределения

- - это его числовые характеристики, указывающие, где "в среднем" располагаются значения признака, насколько эти значения изменчивы и наблюдается ли преимущественное появление определенных значений признака.
- Наиболее практически важными параметрами являются математическое ожидание, дисперсия, показатели асимметрии и эксцесса.
- Параметры распределения оказывается возможным определить только по отношению к данным, представленным по крайней мере в интервальной шкале.

Математическое ожидание характеризует среднее ожидаемое значение СВ, т. е. приближенно равно ее среднему значению. Для решения многих задач достаточно знать эту величину. Например, при оценивании покупательной способности населения вполне может хватить знания среднего дохода. При анализе выгодности двух видов деятельности можно ограничиться сравнением их средних прибыльностей. Знание того, что выпускники данного университета зарабатывают в среднем больше выпускников другого университета, может послужить основанием для принятия решения о поступлении в высшее учебное заведение и т. д.

Математическое ожидание M(X) определяется следующим образом.

Для дискретной СВ:

$$M(X) = \sum_{i=1}^{k} x_i \cdot p_i \,, \qquad (1.5)$$

где k – число всех возможных значений СВ X.

Для непрерывной СВ:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f(x)dx. \qquad (1.6)$$

Однако для детального анализа поведения СВ знание лишь среднего значения явно недостаточно. Существуют отличные друг от друга случайные величины, имеющие одинаковые математические ожидания. Например, средний уровень жизни в Швеции и США приблизительно одинаков, однако разброс в доходах в этих странах существенно отличается.

Следовательно, нужна числовая характеристика, которая будет оценивать разброс возможных значений СВ относительно ее среднего значения (математического ожидания). Такой характеристикой является дисперсия.

Дисперсией D(X) СВ X называется математическое ожидание квадрата отклонения СВ от ее математического ожидания. Она рассчитывается по формулам:

$$D(X) = M(X - M(X))^2 = M(X^2) - M^2(X). \qquad (1.7)$$

При этом для дискретных СВ

$$D(X) = \sum_{i=1}^{k}(x_i - M(X))^2 \cdot p_i = \sum_{i=1}^{k} x_i^{\,2} \cdot p_i - M^2(X). \qquad (1.8)$$

Для непрерывных СВ

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot f(x)dx = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx - M^2(X). \qquad (1.9)$$

Дисперсия имеет размерность, равную квадрату размерности СВ. Для того чтобы представить разброс значений СВ в тех же единицах, что и саму СВ, вводится другая числовая характеристика – среднее квадратическое отклонение.

Средним квадратическим отклонением $\sigma(X)$ СВ X называют квадратный корень из дисперсии D(X):

$$\sigma(X) = \sqrt{D(X)}\,. \tag{1.10}$$

Чтобы оценить разброс значений СВ в процентах относительно ее среднего значения вводится *коэффициент вариации* V(X), рассчитываемый по формуле:

$$V(X) = \frac{\sigma(X)}{|M(X)|} \cdot 100\,\%\,. \tag{1.11}$$

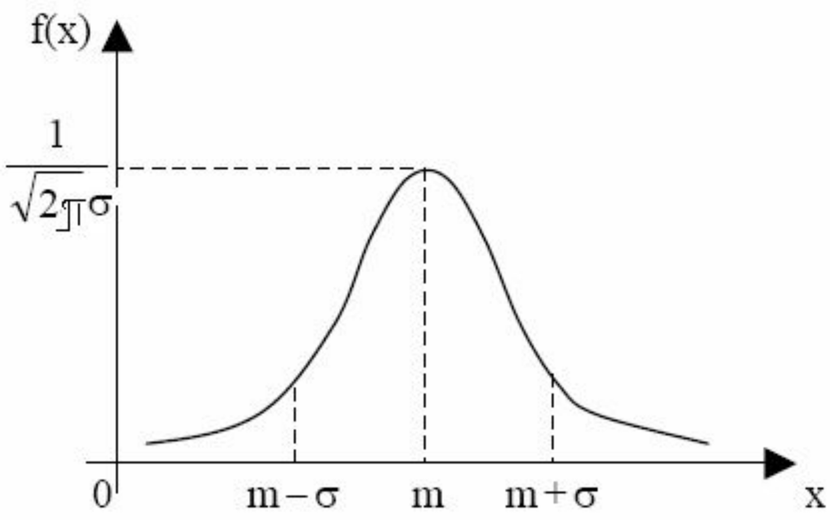
число степеней свободы (ν) исследуемой СВ определяется числом случайных величин, ее составляющих, уменьшенным на число линейных связей между ними. Например, число степеней свободы исследуемой СВ, являющейся композицией n случайных величин, которые в свою очередь связаны m линейными уравнениями, определяется числом $\nu = n - m$.

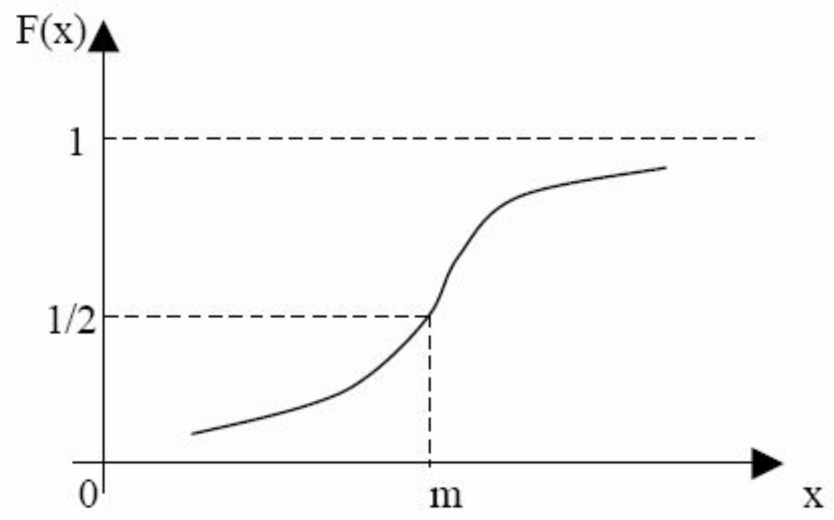
Нормальное распределение (распределение Гаусса) является предельным случаем почти всех реальных распределений вероятности. Поэтому оно используется в очень большом числе реальных приложений теории вероятностей. Говорят, что СВ X имеет *нормальное распределение*, если ее плотность вероятности имеет вид

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{1.12}$$

Это равносильно тому, что функция распределения будет

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt. \tag{1.13}$$





Пример расчета характеристик дискретной случайной величины

По многолетним статистическим данным известно, что вероятность рождения мальчика равна 0,515. Составить закон распределения случайной величины X – числа мальчиков в семье из 4 детей. Найти математическое ожидание и дисперсию этой случайной величины.

Решение

Число мальчиков в семье из $n = 4$ представляет случайную величину X с множеством значений $X = m = 0, 1, 2, 3, 4$, вероятности которых определяются по формуле Бернулли:

$P(X = m) = C_n^m p^m q^{n-m}$, где q = 1-p

В нашем случае n = 4, p = 0,515, q = 1-p = 0,485

Вычислим

$P(X = 0) = C_4^0 * 0{,}515^0 * 0{,}485^4 = 0{,}055$;

$P(X = 1) = C_4^1 * 0{,}515^1 * 0{,}485^3 = 0{,}235$;

$P(X = 2) = C_4^2 * 0{,}515^2 * 0{,}485^2 = 0{,}375$;

$P(X = 3) = C_4^3 * 0{,}515^3 * 0{,}485^1 = 0{,}265$;

$P(X = 4) = C_4^4 * 0{,}515^4 * 0{,}485^0 = 0{,}070$.

(Здесь учтено, что $C_4^0 = 1$, $C_4^1 = 4$, $C_4^2 = \frac{4*3}{1*2} = 6$, $C_4^3 = C_4^1 = 4$, $C_4^4 = 1$)

Ряд распределения имеет вид

$X = m$	x_i	0	1	2	3	4
	p_i	0,055	0,235	0,375	0,265	0,070

Убеждаемся, что $\sum_{i=1}^{5} p_i = 0{,}055 + 0{,}235 + \ldots + 0{,}070 = 1$

Математическое ожидание *M(X)* и дисперсию D(X) можно найти,
учитывая, что закон распределения случайной величины X биномиальный, можно воспользоваться простыми формулами

M(X) = np = 4*0,515 = 2,06,

D(X) = npq = 4*0,515*0,485 = 0,999.

ПРИМЕР 2

Ряд распределения дискретной случайной величины состоит из двух неизвестных значений. Вероятность того, что случайная величина примет одно из этих значений, равна 0,8. Найти функцию распределения случайной величины, если ее математическое ожидание равно 3,2, а дисперсия 0,16.

Решение

Ряд распределения имеет вид

X:	x_i	x_1	x_2
	p_i	0,8	0,2

где $p_i = 0{,}8$, а $p_2 = 1 - p_1 = 1 - 0{,}8 = 0{,}2$.

По условию

$$\begin{cases} a = M(X) = \sum_{i=1}^{2} x_i p_i = 3{,}2 \\ D(X) = M(X^2) - a^2 = \sum_{i=1}^{2} x_i^2 - a^2 = 0{,}16 \end{cases}$$

или

$$\begin{cases} 0{,}8x_1 + 0{,}2x_2 = 3{,}2 \\ 0{,}8x_1^2 + 0{,}2x_2^2 - 3{,}2^2 = 0{,}16 \end{cases}$$

ПРИМЕР 3

Записываем выражение функции распределения:

$$F(x)=\begin{cases}0 & при \quad x\le 3\\ 0{,}8 & при \quad 3<x\le 4\\ 1 & при \quad x>4\end{cases}$$

или

$$F(x)=\begin{cases}0 & при \quad x\le 2{,}4\\ 0{,}2 & при \quad 2{,}4<x\le 3{,}4\\ 1 & при \quad x>3{,}4\end{cases}$$

Дана функция распределения случайной величины X:

$$F(x)=\begin{cases}0 & при \quad x\le 0\\ x^2/4 & при \quad 0<x\le 2\\ 1 & при \quad x>2\end{cases}$$

а) найти плотность вероятности $\varphi(x)$;

г) найти вероятности $P(X=1)$, $P(X<1)$, $P(1\le X<2)$ (две последние вероятности показать на графиках $\varphi(x)$ и $F(x)$);

д) вычислить математическое ожидание $M(X)$, дисперсию D(X), моду $M_0(X)$ и медиану $M_e(X)$.

Решение а) Плотность вероятности

$$\varphi(x)=F'(x)=\begin{cases}0 & при \quad x\le 0 \;\; и \; при \; x>2\\ x/2 & при \;\; 0<x\le 2\end{cases}$$

в) Случайная величина X – непрерывная, так как функция распределения $F(x)$ непрерывна, а ее производная – плотность вероятности $\varphi(x)$ – непрерывна во всех точках, кроме одной *($x = 2$)*.

г) $P(X = 1) = 0$ как вероятность отдельно взятого значения непрерывной случайной величины.

$P(X < 1)$ можно найти либо по определению функции распределения, либо по формуле (3.21) через плотность вероятности $\varphi(x)$:

$P(X < 1) = F(1) = \frac{1^2}{4} = \frac{1}{4}$ (ордината графика $F(1)$ – см. рис. 3.20б)

или

$$P(X < 1) = \int_{-\infty}^{1} \varphi(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} \frac{x}{2}dx = 0 + \frac{x^2}{4}\Big|_0^1 = \frac{1}{4}$$

(площадь под кривой распределения $\varphi(x)$ на отрезке [0,1] – см.рис.3.20а).

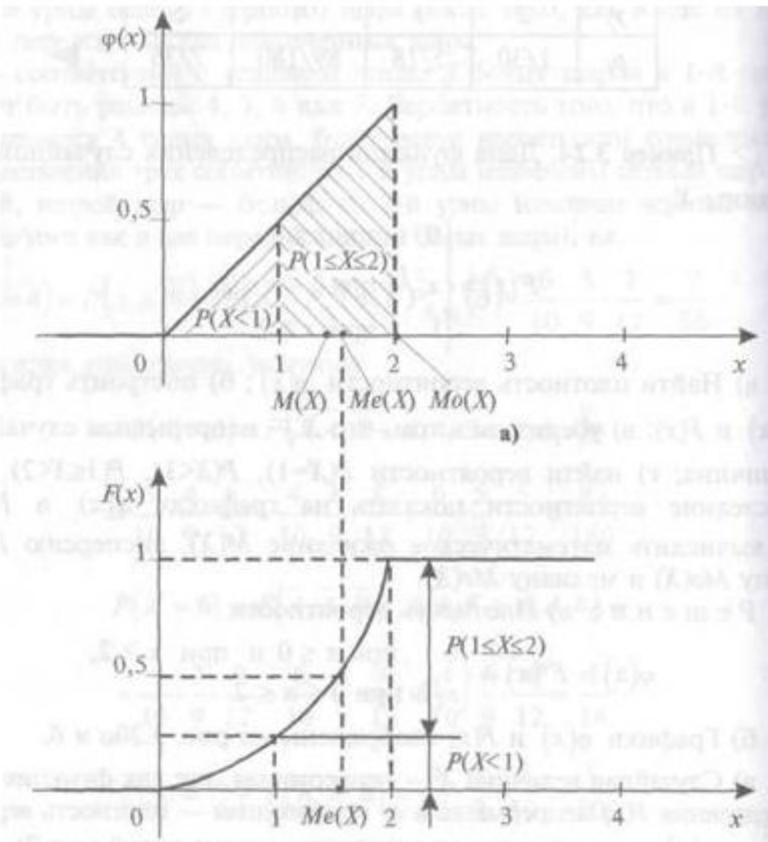


Рис. 3.20

$P(1 \leq X \leq 2)$ можно найти либо как приращение функции распределения по формуле (3.20), либо по формуле (3.22) через

$$P(1 \leq X \leq 2) = F(2) - F(1) = \frac{2^2}{4} - \frac{1^2}{4} = \frac{3}{4}$$

(площадь под кривой распределения $\varphi(x)$ на отрезке [1,2] – рис. 3.20а)

д) По формуле (3.25) математическое ожидание

$$a = M(X) = \int_{-\infty}^{\infty} x * \varphi(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{2} x * \left(\frac{x}{2}\right)dx + \int_{2}^{\infty} 0dx = 0 + \frac{x^3}{6}\bigg|_0^2 + 0 = \frac{1}{6} * 2^3 = \frac{4}{3}$$

плотность вероятности $\varphi(x)$.

Если представить распределение случайной величины X *в* виде единичной массы, распределенной по треугольнику (рис. 3.20а), то значение $M(X)=4/3$ означает абсциссу центра массы треугольника.

По формуле (3.27) дисперсия $D(X) = M(X^2) - a2$.

$$M(X^2) = \int_{-\infty}^{\infty} x^2 * \varphi(x)dx = 0 + \int_{0}^{2} x^2 * \left(\frac{x}{2}\right)dx + 0 = 2$$

Теперь

$$D(X) = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}$$

Плотность вероятности $\varphi(x)$ максимальна при x = 2 (см. рис. 3.20а), следовательно, $M_0(X) = 2$.

Медиану $M_e(X) = b$ найдем из условия $F(b) = \frac{1}{2}$, т.е. $\frac{b^2}{4} = \frac{1}{2}$, откуда $b = M_e(X) = \sqrt{2}$, или через плотность вероятности

$\int_{-\infty}^{b} \varphi(x)dx = \frac{1}{2}$, т.е. $0 + \int_{0}^{b} \frac{x}{2}dx = \frac{x^2}{4}\Big|_0^b = \frac{b^2}{4} = \frac{1}{2}$, откуда $b = M_e(X) = \sqrt{2}$

ПРИМЕР 4

Дано:

Дано следующее распределение дискретной случайной величины X

X	1	2	4	5
P	0.31	0.1	0.29	0.3

Найти ее математическое ожидание, дисперсию и среднеквадратичное отклонение, используя формулы для их определения.

Решение:

1) Найдем математическое ожидание **M(x)**

$$M(x)=\sum_{i=1}^{n} x_i p_i = 1\times 0.31+2\times 0.1+4\times 0.29+5\times 0.3=3.17$$

2) Найдем дисперсию **D(x)**

$$D(x)=\sum_{i=1}^{n}\left(x_i-M(x)\right)^2 p_i=\left(1-3.17\right)^2\times 0.31+\left(2-3.17\right)^2\times 0.1+\left(4-3.17\right)^2\times 0.29+\left(5-3.17\right)^2\times 0.3=2.8011$$

3) Найдем среднеквадратичное отклонение

$$\sigma(x)=\sqrt{D(x)}=\sqrt{2.8011}\approx 1.6736$$

Пример биноминального распределения

В магазин поступила обувь с двух фабрик в соотношении 2:3. Куплено 4 пары обуви. Найти закон распределения числа купленных пар обуви, изготовленной первой фабрикой. Найти математическое ожидание и среднее квадратическое отклонение этой случайной величины.

Решение

Вероятность того, что случайно выбранная пара обуви изготовлена первой фабрикой, равна

p=2/(2+3)=0,4.

Случайная величина X – число пар обуви среди четырех, изготовленных первой фабрикой, имеет биномиальный закон распределения с параметрами $n = 4$, $p = 0,4$. Ряд распределения X имеет вид:

x_i	0	1	2	3	4
p_i	0,1296	0,3456	0,3456	0,1536	0,0256

(Значения p_i – $P(X = m)$, (m = 0,1,2,3,4) вычислены по формуле

$$P(X = m) = C_4^m * 0{,}4^m * 0{,}6^{4-m}$$

Найдем математическое ожидание и дисперсию случайной величины X по формулам:

$$M(X) = n * p = 4 * 0{,}4 = 1{,}6\ ,\ D(X) = n * p * q = 4 * 0{,}4 * 0{,}6 = 0{,}96$$

Геометрическое распределение

Проводится проверка большой партии деталей до обнаружения бракованной (без ограничения числа проверенных деталей). Составить закон распределения числа проверенных деталей. Найти его математическое ожидание и дисперсию, если известно, что вероятность брака для каждой детали равна 0,1.

Решение

Случайная величина X – число проверенных деталей до обнаружения бракованной – имеет геометрическое распределение (4.11) с параметром p =0,1. Поэтому ряд распределения имеет вид

$X = m$:	x_i	1	2	3	4	…	m	…
	p_i	0,1	0,09	0,081	0,0729	…	$0{,}9^m*0{,}1$	…

По формулам:

$$M(X) = \frac{1}{p} = \frac{1}{0{,}1} = 10, \quad D(X) = \frac{q}{p^2} = \frac{0{,}9}{0{,}1^2} = 90$$

Гипергеометрическое распределение

В лотерее «Спортлото 6 из 45» денежные призы получают участники, угадавшие 3, 4, 5 и 6 видов спорта из отобранных случайно 6 видов из 45 (размер приза увеличивается с увеличением числа угаданных видов спорта). Найти закон распределения случайной величины X – числа угаданных видов спорта среди случайно отобранных шести. Какова вероятность получения денежного приза? Найти математическое ожидание и дисперсию случайной величины X.

Решение

Число угаданных видов спорта в лотерее «6 из 45» есть случайная величина, имеющая гипергеометрическое распределение с параметрами $n = 6$, $M = 6$, $N = 45$. Ряд ее распределения, рассчитанный по формуле (4.14), имеет вид:

X:	x_i	0	1	2	3	4	5	6
	p_i	0,40056	0,42413	0,15147	0,02244	0,00137	0,00003	0,0000001

Вероятность получения денежного приза

$$P(3 \le X \le) = \sum_{i=3}^{6} P(X = i) = 0{,}02244 + 0{,}00137 + 0{,}00003 + 0{,}0000001 = 0{,}02384 \approx 0{,}024$$

По формулам

$$M(X) = 6 * \frac{6}{45} = 0{,}8, \quad D(X) = 6 * \frac{39}{44} * \left(1 - \frac{39}{45}\right) * \left(1 - \frac{6}{45}\right) = 0{,}6145$$

Таким образом, среднее число угаданных видов спорта из 6 всего 0,8, а вероятность выигрыша только 0,024.

Пуассоновское распределение

Установлено, что время ремонта телевизоров есть случайная величина *X*, распределенная по показательному закону. Определить вероятность того, что на ремонт телевизора потребуется не менее 20 дней, если среднее время ремонта телевизоров составляет 15 дней. Найти плотность вероятности, функцию распределения и среднее квадратическое отклонение случайной величины *X*.

Решение

По условию математическое ожидание $M(x)=1/\lambda = 15$, откуда параметр λ - 1/15 и по формулам (4.21) и (4.22) плотность вероятности и функция распределения имеют вид:

$$\varphi(x) = \frac{1}{15} * e^{-\frac{1}{15} * x}, \quad F(x) = 1 - e^{-\frac{1}{15} * x}, \quad (x \geq 0).$$

Искомую вероятность $P(X \geq 20)$ можно было найти по формуле (3.22), интегрируя плотность вероятности, т.е.

$$P(X \geq 20) = P(20 \leq X < \infty) = \int_{20}^{\infty} \frac{1}{15} * e^{-\frac{1}{15} * x} dx,$$

но проще это сделать, используя функцию распределения:

$$P(X \geq 20) = 1 - P(X < 20) = 1 - F(20) = 1 - (1 - e^{-\frac{20}{15}}) = e^{-\frac{20}{15}} = 0{,}264$$

Осталось найти среднее квадратическое отклонение $\sigma_x = M(X) = 15$ дней.

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ - НОРМАЛЬНОЕ

- Нормальным распределение называется потому, что оно очень часто встречалось в естественно-научных исследованиях и казалось "нормой" всякого массового случайного проявления признаков. Это распределение следует закону, открытому тремя учеными в разное время: Муавром в 1733 г. в Англии, Гауссом в 1809 г. в Германии и Лапласом в 1812 г. во Франции. График нормального распределения представляет собой привычную глазу психолога-исследователя так называемую колоколообразную кривую.

Для практических расчетов специально разработаны таблицы функций f(u), F(u) стандартизированного нормального распределения, однако чаще используется так называемая *таблица значений функции Лапласа Ф(u)*

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{t^2}{2}} dt = F(u) - 0.5. \qquad (1.15)$$

Эти таблицы можно использовать для любой нормальной СВ X ($X \sim N(m, \sigma)$) при расчете соответствующих вероятностей:

$$P(a \le X \le b) = F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right). \quad (1.16)$$

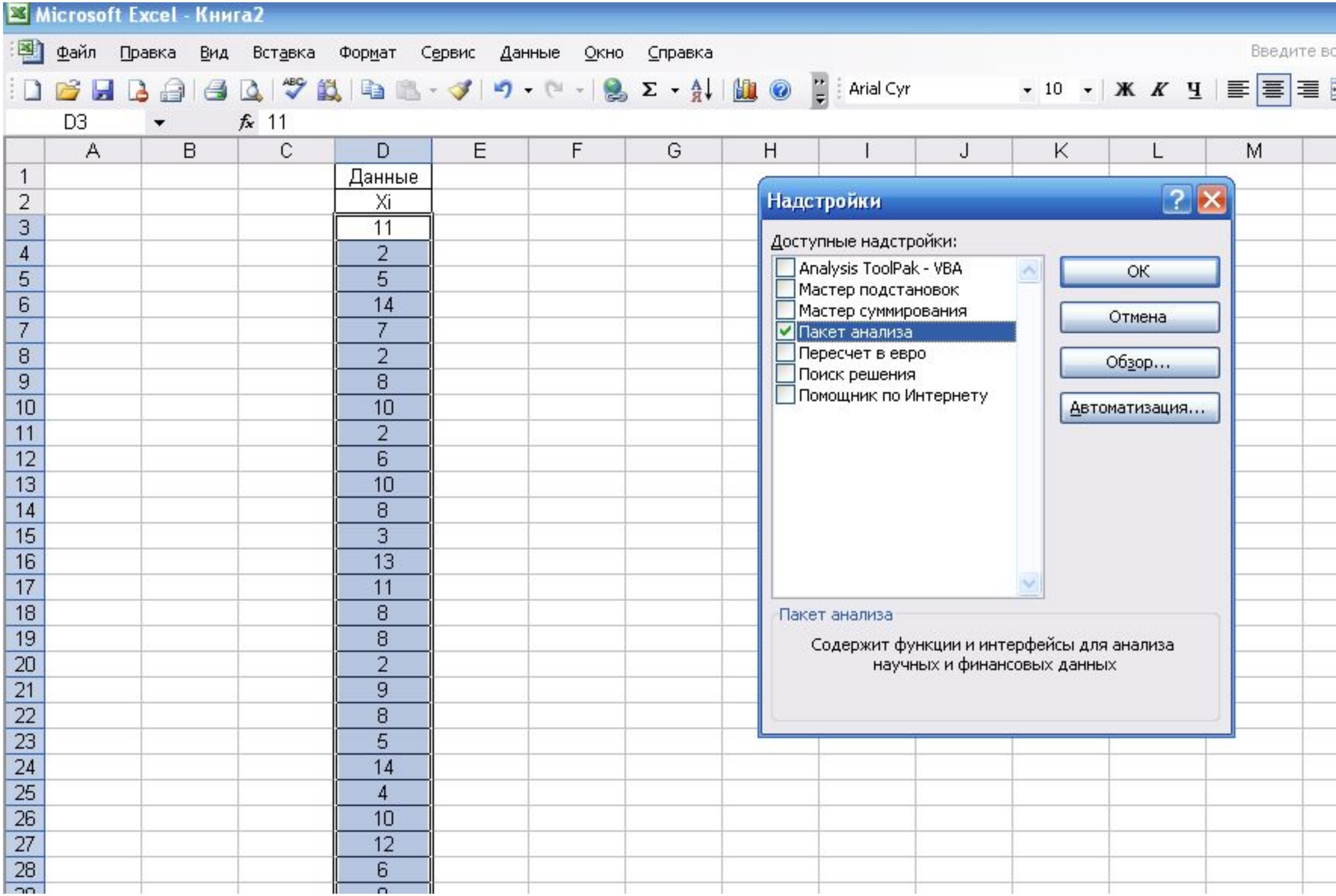
Заметим, что если $X \sim N(m, \sigma)$, то $U = \dfrac{X-m}{\sigma} \sim N(0,1)$.

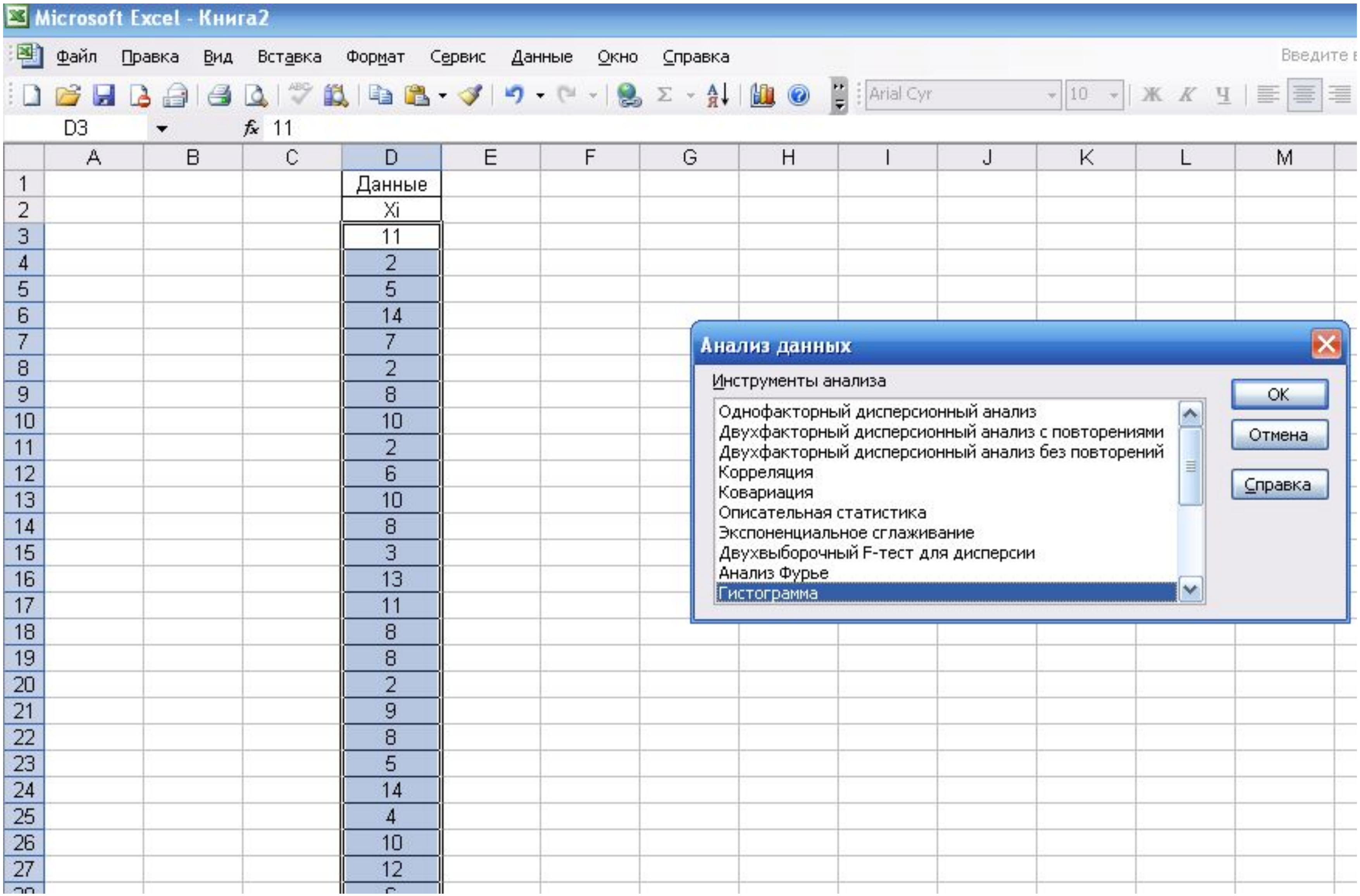
Кроме того, справедливы следующие

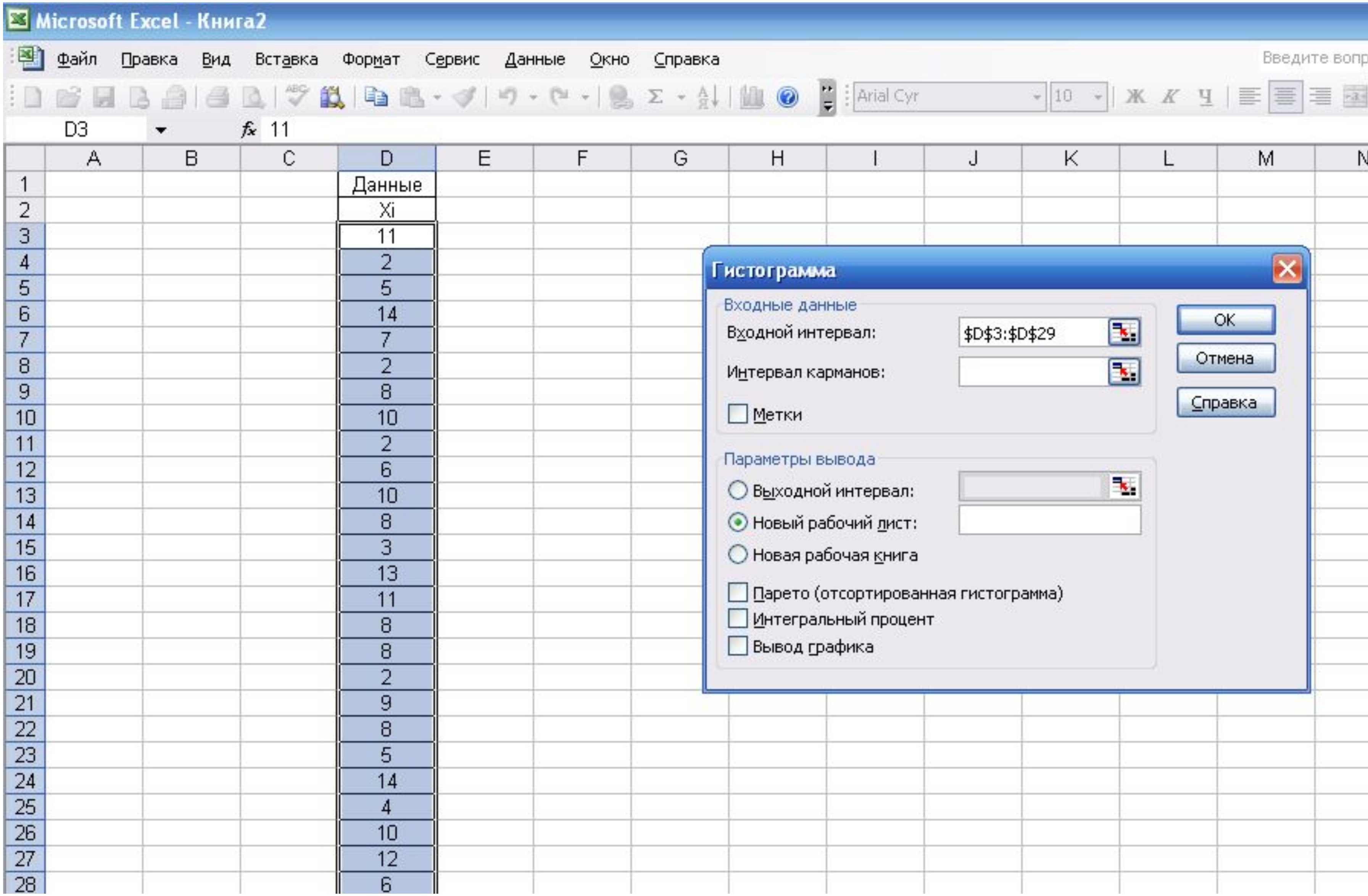
соотношения. $P(\,|X - M(X)| < \sigma\,) = 0.68$; $P(|X - M(X)| < 2\sigma\,) = 0.95$; $P(\,|X - M(X)| < 3\sigma\,) = 0.9973$. Другими словами, значения нормально распределенной СВ X ($X \sim N(m, \sigma)$) на 99.73 % сосредоточены в области $[m - 3\sigma, m + 3\sigma]$.

Методы проверки распределения на нормальность

- 1. **Визуальный метод** включает построение гистограммы эмпирического распределения (такая возможность реализована во всех популярных статистических пакетах) и ее визуальное сравнение с теоретической кривой нормального распределения.
- 2. **Метод оценки показателей асимметрии и эксцесса** основан на вычислении этих показателей и их стандартных ошибок.
- 3. **Критерии согласия**


Microsoft Excel - Книга2
Файл
Правка
Вид
Вставка
Формат
Сервис
Данные
Окно
Справка
Введите воп
Arial Cyr
10
Ж
К
Ч
D3
11
A
B
C
D
E
F
G
H
I
J
K
L
M
Данные
Xi
11
2
5
14
7
2
8
10
2
6
10
8
3
13
11
8
8
2
9
8
5
14
4
10
12
6
Надстройки
Доступные надстройки:
Analysis ToolPak - VBA
Мастер подстановок
Мастер суммирования
Пакет анализа
Пересчет в евро
Поиск решения
Помощник по Интернету
OK
Отмена
Обзор...
Автоматизация...
Пакет анализа
Содержит функции и интерфейсы для анализа научных и финансовых данных

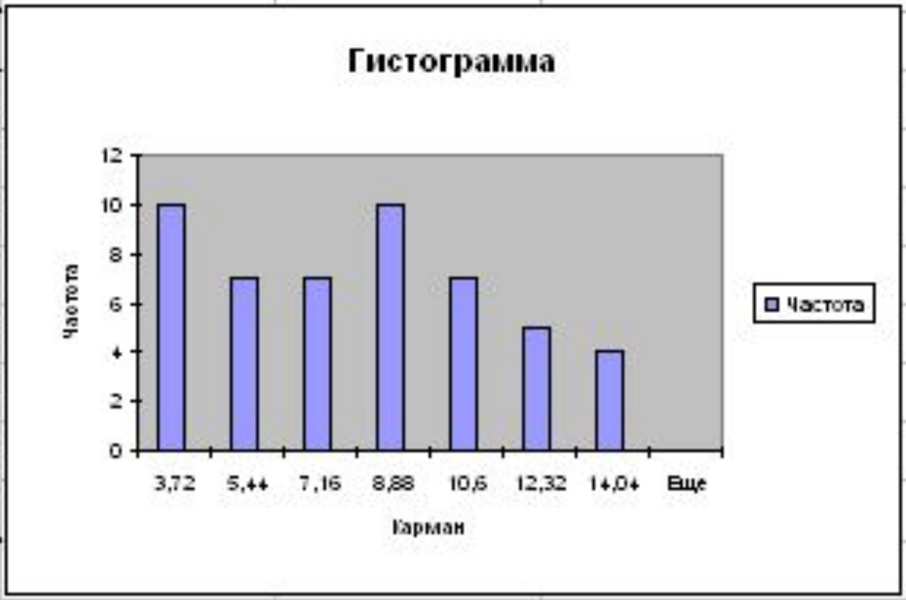

Microsoft Excel - Книга2
Файл
Правка
Вид
Вставка
Формат
Сервис
Данные
Окно
Справка
Arial Cyr
10
D3
11
A
B
C
D
E
F
G
H
I
J
K
L
M
Данные
Xi
11
2
5
14
7
2
8
10
2
6
10
8
3
13
11
8
8
2
9
8
5
14
4
10
12
Анализ данных
Инструменты анализа
Однофакторный дисперсионный анализ
Двухфакторный дисперсионный анализ с повторениями
Двухфакторный дисперсионный анализ без повторений
Корреляция
Ковариация
Описательная статистика
Экспоненциальное сглаживание
Двухвыборочный F-тест для дисперсии
Анализ Фурье
Гистограмма
OK
Отмена
Справка


Microsoft Excel - Книга2
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
Введите вопро
Arial Cyr
10
D3
11
Данные
Xi
11
2
5
14
7
2
8
10
2
6
10
8
3
13
11
8
8
2
9
8
5
14
4
10
12
6
Гистограмма
Входные данные
Входной интервал:
$D$3:$D$29
Интервал карманов:
Метки
OK
Отмена
Справка
Параметры вывода
Выходной интервал:
Новый рабочий лист:
Новая рабочая книга
Парето (отсортированная гистограмма)
Интегральный процент
Вывод графика

	A	B	C	D	E	F
10	Число интервалов m определяется по формуле Стерджесса					
11	Число интервалов= 1+3,322*LOG10(50)=					7
12	Величина интервала определяется по формуле					
13	I =	(МАКС(B16:B65)-МИН(B16:B65))/E11=			(14-2)/7=	h=
14		Данные	№ интервала	Нижняя граница интервала	Верхняя граница интервала= нижняя +h	Частота повторения, f
15		Xi	1	2	3,72	10
16		11	2	3,72	5,44	7
17		2	3	5,44	7,16	7
18		5	4	7,16	8,88	10
19		14	5	8,88	10,6	7
20		7	6	10,6	12,32	5
21		2	7	12,32	14,04	4
22		8		сумма		50
23		10	ИНТЕРВАЛЬНЫЙ РЯД РАСПРЕДЕЛЕНИЯ			
24		2				
25		6		n	50	
26		10		среднее	7,16	
27		8		сигма	3,495536512	
28		3		i=	7	
29		13		Xмин	2	
30		11		Xмах	14	
31		8				
32		8				
33		2				

Визуальный метод

Карман	*Частота*
3,72	10
5,44	7
7,16	7
8,88	10
10,6	7
12,32	5
14,04	4
Еще	0



Асимметрией (Skewness)

- или **выборочным коэффициентом скошенности,** называют меру отклонения эмпирического распределения частот от распределения, симметричного относительно максимальной ординаты.
- Для симметричных распределений показатель асимметрии равен нулю. Отрицательный показатель асимметрии означает, что кривая распределения скошена **влево**, положительный – **вправо** от теоретической симметричной кривой распределения.

АССИМЕТРИЯ

$$As = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n \cdot \delta^3}.$$

ЕСЛИ ДАННЫЕ НОРМАЛЬНЫЕ, то они имеют симметричное распределение признака, когда As=0

Стандартная ошибка асимметрии (Standard error of skewness) – это отклонение асимметрии от нуля, которое находится по формуле.

$$m_{As} = \sqrt{\frac{6}{n}}.$$

Показатель эксцесса (Kurtosis)

- или **выборочный коэффициент остро вершинности** характеризует степень отклонения эмпирической кривой распределения от теоретической кривой нормального распределения

$$Ex = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4}{n \cdot \delta^4} - 3$$

ЭКСЦЕСС («горб»)

ЕСЛИ ДАННЫЕ НОРМАЛЬНЫЕ, то они имеют такой «горб», когда

Ex=0

Стандартная ошибка эксцесса (Standard error of kurtosis)

$$m_{Ex} = 2m_{As} = 2\sqrt{\frac{6}{n}} .$$

Показатели ассиметрии и эксцесса свидетельствуют о достоверном отличии эмпирического распределения от нормального в том случае, если они превышают свою ошибку репрезентативности более чем в 3 раза (метод Н.А. Плохинского).

$$t_a = \frac{|A|}{m_A} \geq 3; \quad t_E = \frac{|E|}{m_E} \geq 3$$

Одним из способов проверки распределения экспериментальных данных на нормальность является расчет показателей ассиметрии и эксцесса и сопоставление их с критическими значениями (метод Е.И. Пустыльника).

$$A_{кр} = 3 \cdot \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}}; \; E_{кр} = 5 \cdot \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}}$$

Эмпирическое распределение достоверно отличается от нормального распределения, если $A_{эмп} > A_{кр}$, $E_{эмп} > E_{кр}$.

НОРМАЛЬНЫЕ ДАННЫЕ
правило трех сигм

ЕСЛИ В СОВОКУПНОСТИ ВСЕ ВАРИАНТЫ НАХОДЯТСЯ В ИНТЕРВАЛЕ , ТО ОНИ НОРМАЛЬНЫЕ

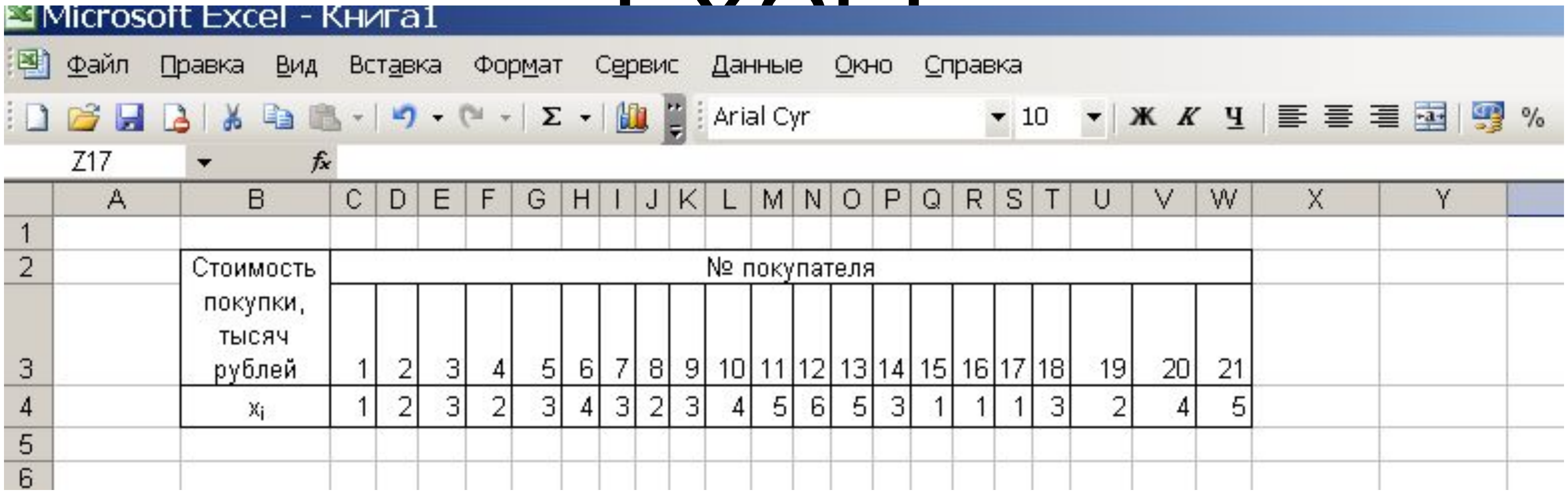
$$\bar{x} \pm 3\sigma$$

АНОМАЛЬНЫЕ ДАННЫЕ ВЫБРОСЫ

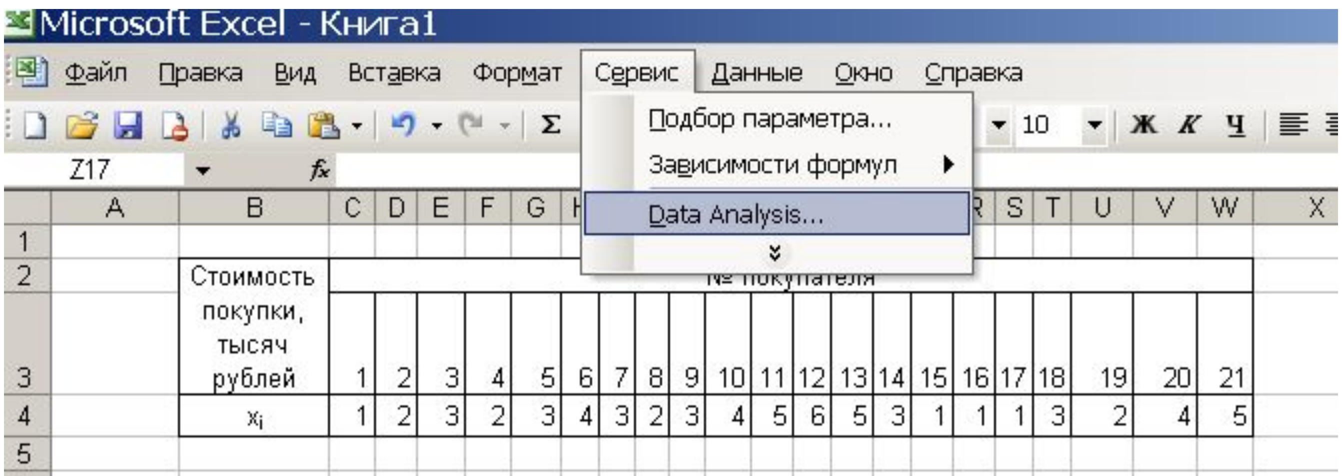
ЕСЛИ В СОВОКУПНОСТИ ЕСТЬ ВАРИАНТЫ, КОТОРЫЕ НАХОДЯТСЯ ЗА ПРЕДЕЛАМИ ИНТЕРВАЛА , ТО ОНИ АНОМАЛЬНЫЕ

$$\bar{x} \pm 3\sigma$$

РАСЧЕТ ПОКАЗАТЕЛЕЙ В EXCEL

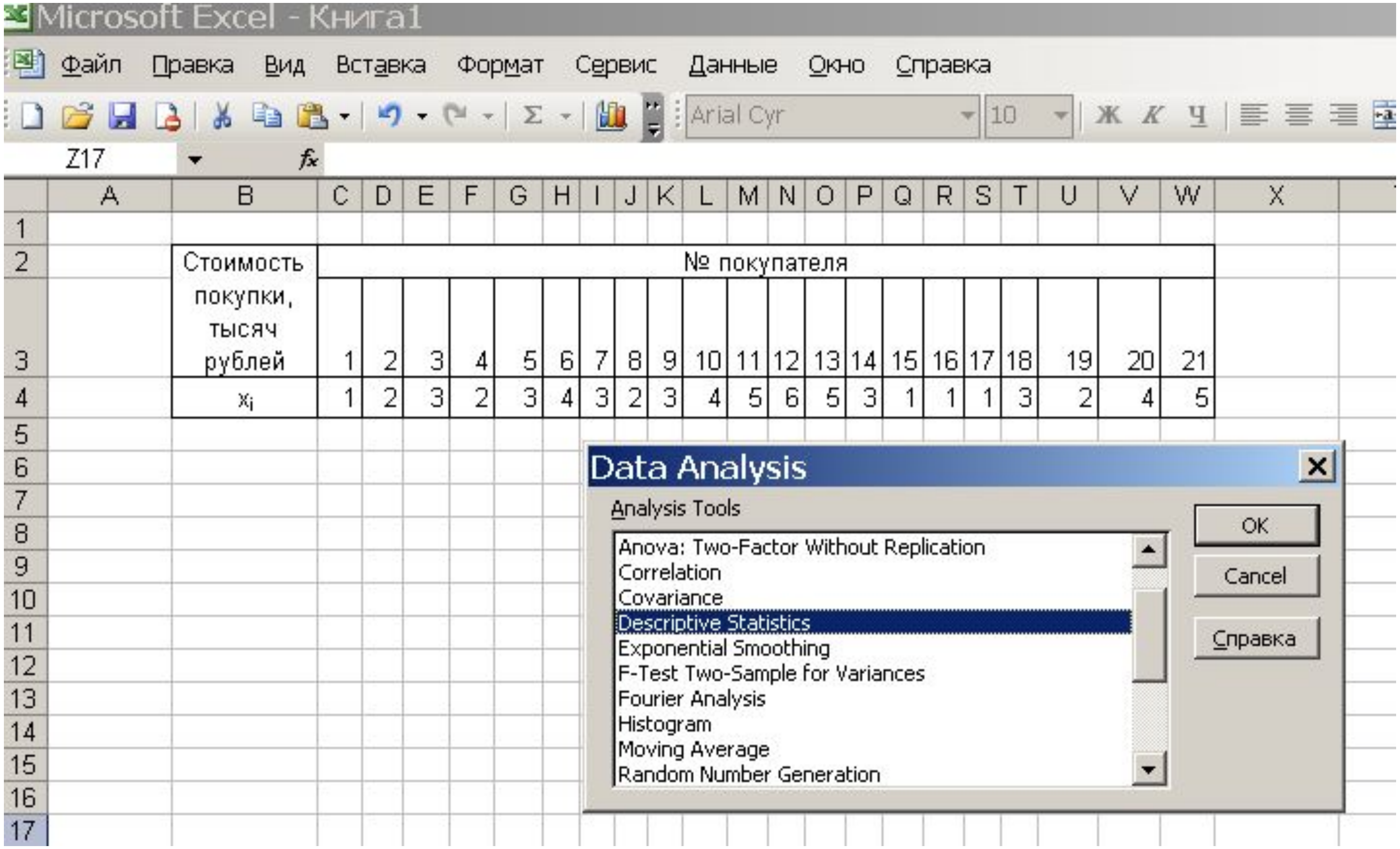


Стоимость покупки, тысяч рублей	№ покупателя																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_i	1	2	3	2	3	4	3	2	3	4	5	6	5	3	1	1	1	3	2	4	5



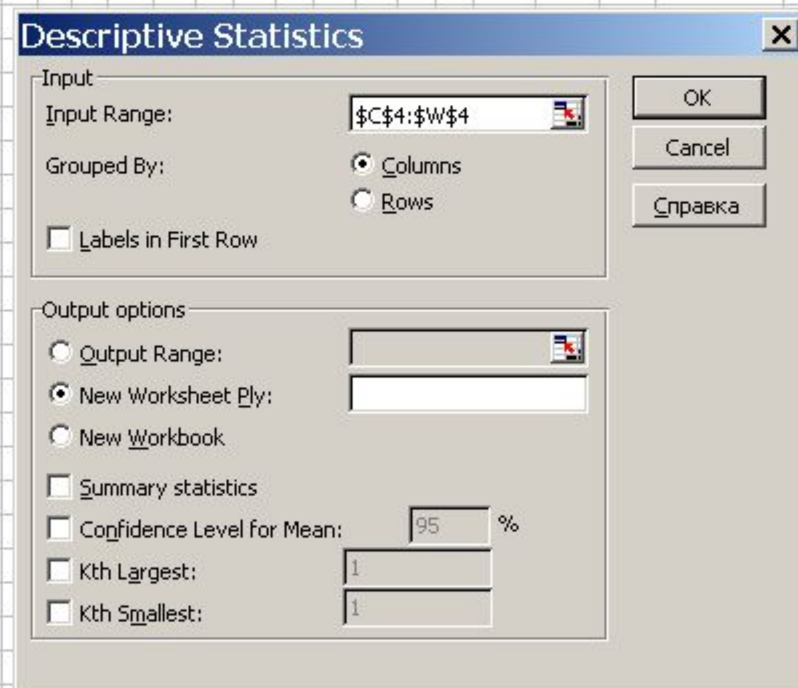
Стоимость покупки, тысяч рублей	№ покупателя																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_i	1	2	3	2	3	4	3	2	3	4	5	6	5	3	1	1	1	3	2	4	5

РАСЧЕТ ПОКАЗАТЕЛЕЙ В EXCEL

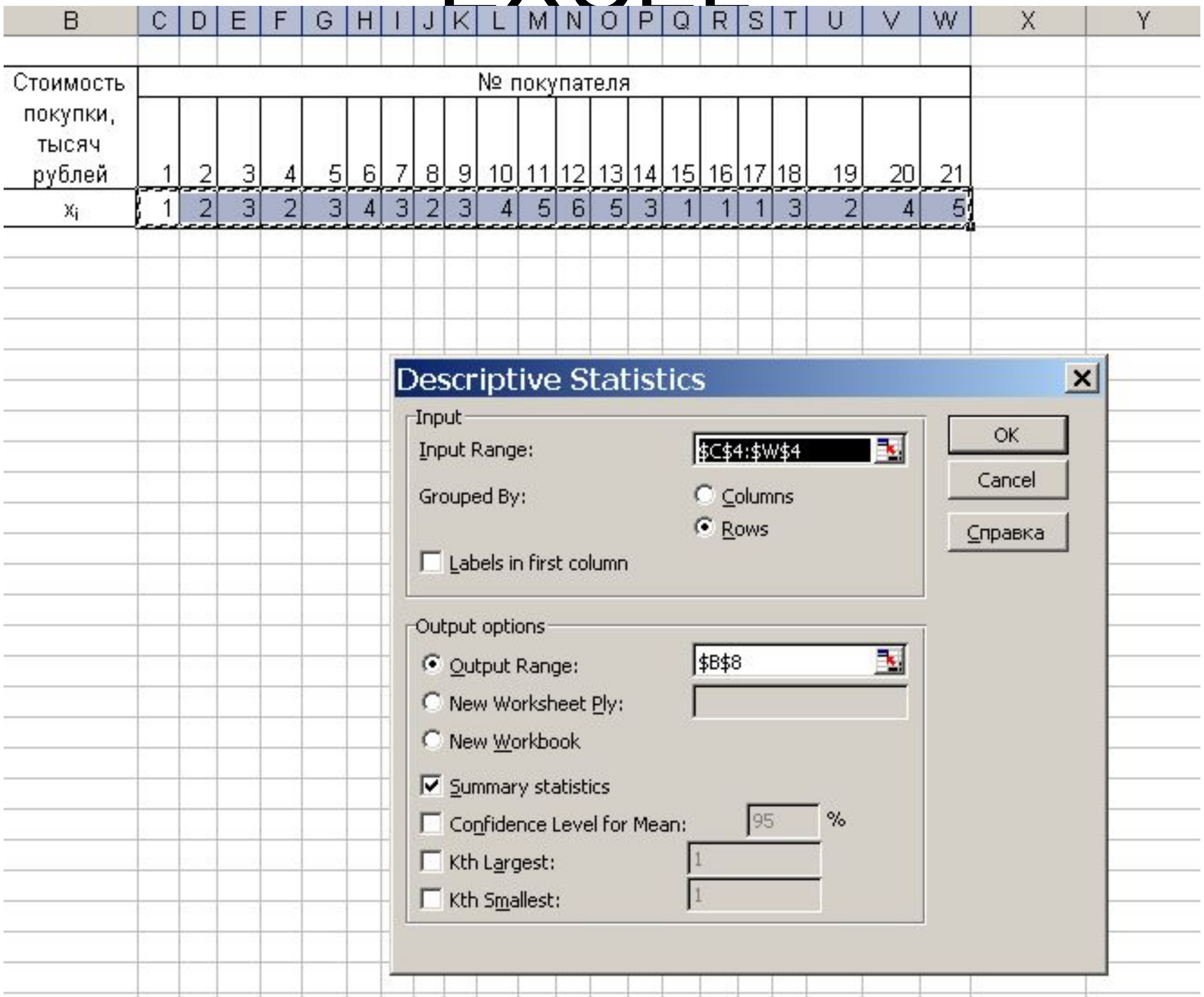


РАСЧЕТ ПОКАЗАТЕЛЕЙ В

Стоимость покупки, тысяч рублей	№ покупателя																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_i	1	2	3	2	3	4	3	2	3	4	5	6	5	3	1	1	1	3	2	4	5



РАСЧЕТ ПОКАЗАТЕЛЕЙ В EXCEL



Стоимость покупки, тысяч рублей	№ покупателя																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_i	1	2	3	2	3	4	3	2	3	4	5	6	5	3	1	1	1	3	2	4	5

РАСЧЕТ ПОКАЗАТЕЛЕЙ В EXCEL

Стоимость покупки, тысяч рублей	№ покупателя																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x_i	1	2	3	2	3	4	3	2	3	4	5	6	5	3	1	1	1	3	2	4	5

Row1		
Mean	3	среднее
Standard Error	0,324	d
Median	3	Me
Mode	3	Mo
Standard Deviation	1,483	сигма
Sample Variance	2,2	
Kurtosis	-0,72	асимметрия
Skewness	0,305	эксцесс
Range	5	
Minimum	1	
Maximum	6	
Sum	63	
Count	21	n

Критерий согласия Пирсона

	A	B	C	D	E	F	G	H	I	J
)										
0	Число интервалов m определяется по формуле Стерджесса									
1	Число интервалс 1+3,322*LOG10(50)=					7		Xмин	2	
2	Величина интервала h определяется по формуле									
3	h=	(МАКС(B16:B65)-МИН(B16:B65))/E:			(14-2)/7=	h=	1,72	Xмах	14	
4		Данные	*№ интервала*	*Нижняя граница интервала*	*Верхняя граница интервала= нижняя +h*	*Частота повторения, f*	*Середина интервала*	*Теоретические частоты=$E$25*$G$13*НОРМРАСП(G15;$F$26;$F$27;0),f'*	Разность частот	$(f-f')^2/f'$
5		Xi	1	2	3,72	10	2,86	4,60572193	5,394278065	6,317845
6		11	2	3,72	5,44	7	4,58	7,4748033	-0,4748033	0,03016
7		2	3	5,44	7,16	7	6,3	9,522495931	-2,522495931	0,668206
8		5	4	7,16	8,88	10	8,02	9,522495931	0,477504069	0,023944
9		14	5	8,88	10,6	7	9,74	7,4748033	-0,4748033	0,03016
0		7	6	10,6	12,32	5	11,46	4,605721935	0,394278065	0,033753
1		2	7	12,32	14,04	4	13,18	2,227638078	1,772361922	1,410133
2		8		сумма		50		45,43368041		8,5142
3		10								
4		2	ИНТЕРВАЛЬНЫЙ РЯД РАСПРЕДЕЛЕНИЯ							
5		6		n	50		Критерий Г	хи-квадрат=		8,5142
6		10		среднее	7,16		табличное значение хи-квадрат			9,487729
7		8		сигма	3,495536512					
8		3		i=	7					

ИНТЕРВАЛЬНЫЙ РЯД РАСПРЕДЕЛЕНИЯ

n	50
среднее	7,16
сигма	3,495536512
i=	7

Критерий Гхи-квадрат=	8,5142
табличное значение хи-квадрат	9,487729

ВЫВОД: нормальное распределение,

так как табличное значение критерия Пирсона больше расчетного значения (9,48>8,51)

ПРИМЕНЕНИЕ СВ

Нормальное распределение используется при проверке различных гипотез в статистике (о величине математического ожидания при известной дисперсии, о равенстве математических ожиданий и т. д.).

Распределение χ^2 применяется для нахождения интервальных оценок, а также при проверке статистических гипотез. При этом активно используется таблица критических точек χ^2-распределения

Распределение Стьюдента применяется для нахождения интервальных оценок, а также при проверке статистических гипотез. При этом активно используется таблица критических точек распределения Стьюдента

Распределение Фишера используется при проверке статистических гипотез, в дисперсионном и регрессионном анализах. При этом активно используется таблица критических точек распределения Стьюдента

ПОЧЕМУ РАСПРЕДЕЛЕНИЕ ГАУССА ГЛАВНОЕ?

1. Оно является предельным законом, к которому приближаются другие законы при росте числа степеней свободы.
2. ЗБЧ доказывают, что сумма достаточно большого числа независимых случайных величин, подчиненных каким-либо законам распределения, приближенно подчиняется нормальному закону, и это тем точнее, чем больше количество случайных величин суммируется, каждая величин в сумме должна играть малую роль.
3. Оно используется, если СВ имеет закон распределения другой, но вычисления по нему сложны, а аппроксимация его нормальным распределением допустима.

ФУНКЦИИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ В EXCEL

- НОРМРАСП(x; среднее;σ;интегральная)

Если интегральная равна Ложь(0), то вычисляется плотность распределения, если Истина, то функция нормального распределения

Пример 6.1. Для закупки и последующей продажи мужских зимних курток фирмой было проведено выборочное обследование мужского населения города в возрасте от 18 до 65 лет в целях определения его среднего роста. В результате было установлено, что средний рост $\overline{x} = 176$ см, стандартное отклонение $\sigma = 6$ см. Необходимо определить, какой процент общего числа закупаемых курток должны составлять куртки 5-го роста (182–186 см). Предполагается, что рост мужского населения города распределен по нормальному закону.

Формула для решения задачи имеет следующий вид:

=НОРМРАСП(186;176;6;ИСТИНА)–НОРМРАСП (182;176;6;ИСТИНА) = 0,95221 – 0,84134 = 0,11086 ≈ 11%.

Таким образом, куртки 5-го роста должны составлять приблизительно 11% общего числа закупаемых курток.

ФУНКЦИИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ В EXCEL

- НОРМОБР(вероятность, среднее, σ) – рассчитывает значение случайной величины, если известна её вероятность и параметры распределения.

На практике часто встречается задача, обратная задаче вычисления вероятности попадания нормально распределенной случайной величины на участок, симметричный относительно математического ожидания $\bar{x}$. Формула для вероятности попадания случайной величины на участок, симметричный относительно математического ожидания, имеет следующий вид:

$$P(|x-\bar{x}|<l)=2\Phi^{*}\left(\frac{l}{\sigma}\right)-1,$$

где l – половина длины участка, симметричного относительно математического ожидания.

Пример 6.2. Для задачи, рассмотренной в примере 6.1, рассчитать границы интервала роста мужского населения города, вероятность попадания в который случайной величины роста составляет 0,95.

Для этого предварительно необходимо преобразовать аргументы НОРМОБР к стандартному виду, в результате чего имеем

$$l = \text{НОРМОБР}\ ((P+1)/2;0;\sigma\,).$$

После подстановки данных получим формулу = НОРМОБР ((0,95 + 1)/2;0;6), которая рассчитает значение 11,7598. Таким образом, границы искомого интервала составят 164,24 и 187,76 см.

Распределение χ^2 (хи-квадрат)

Пусть X_i, $i = 1, \ldots, n$ – независимые нормально распределенные СВ с математическими ожиданиями m_i и средними квадратическими отклонениями σ_i соответственно, т. е. $X_i \sim N(m_i, \sigma_i)$.

Тогда СВ $U_i = (X_i - m_i)/\sigma_i$, $i = 1,\ldots, n$ являются независимыми СВ, имеющими стандартизированное нормальное распределение, $U_i \sim N(0, 1)$.

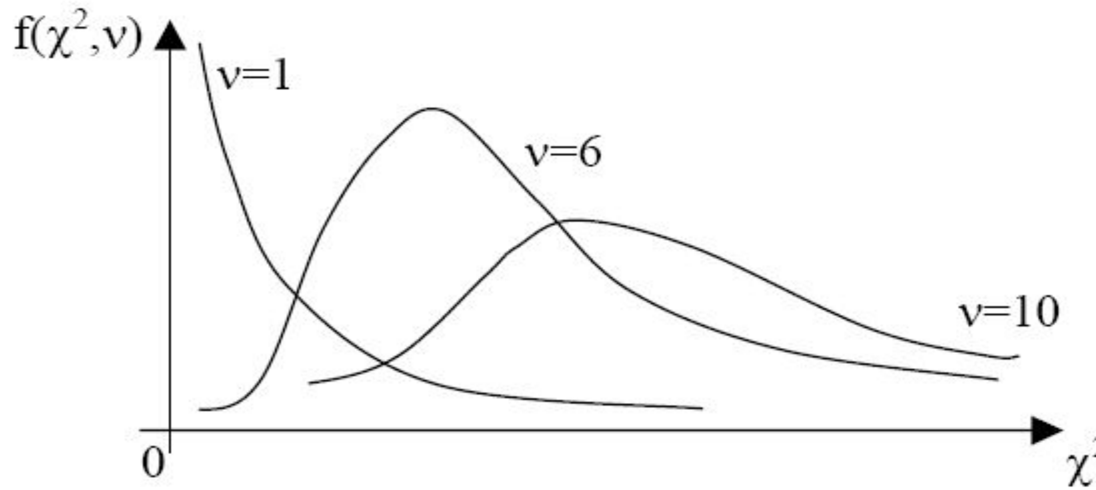
СВ χ^2 имеет *хи-квадрат распределение с n степенями свободы* ($\chi^2 \sim \chi^2_n$), если

$$\chi^2 = \sum^n U_i^2 = U_1^2 + U_2^2 + \ldots + U_n^2.$$

с увеличением числа степеней свободы распределение χ^2 постепенно приближается к нормальному

$M(\chi^2) = \nu = n - m,$

$D(\chi^2) = 2\nu = 2(n - m).$

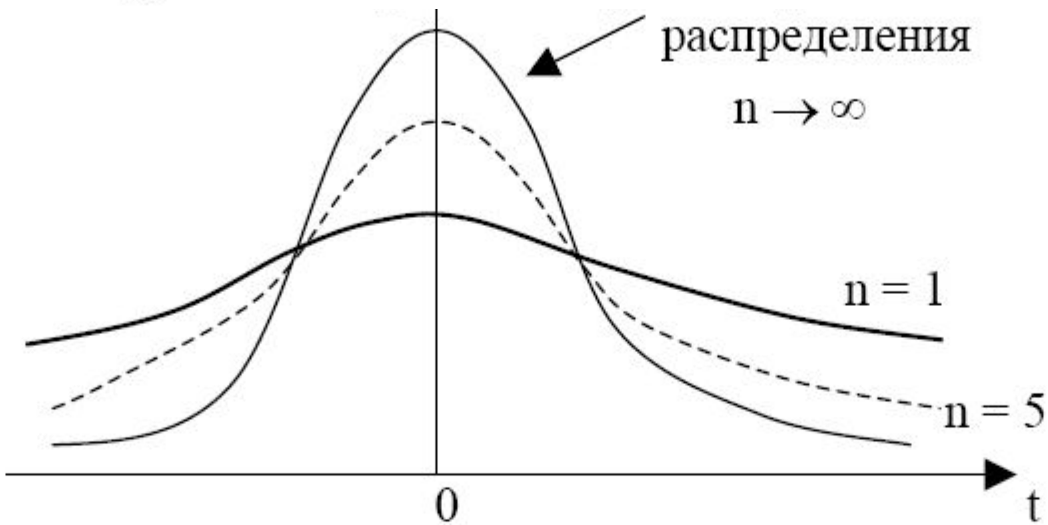


Распределение Стьюдента

Пусть СВ $U \sim N(0, 1)$, СВ V – независимая от U величина, распределенная по закону χ^2 с n степенями свободы. Тогда величина

$$T = \frac{U}{\sqrt{V/n}} \tag{1.18}$$

имеет *распределение Стьюдента (t-распределение) с n степенями свободы* ($T \sim T_n$).



$M(T) = 0,$

$D(T) = \dfrac{n}{n-2}.$

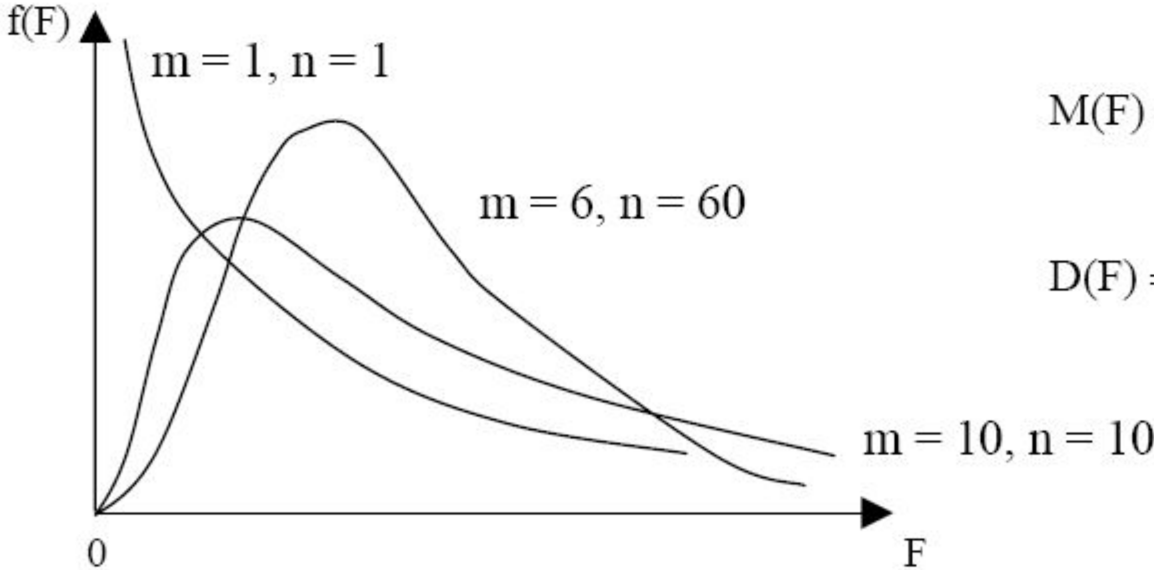
При этом с увеличением числа степеней свободы распределение Стьюдента приближается к стандартизированному нормальному, причем при $n > 30$ распределение Стьюдента практически можно заменить нормальным распределением.

Распределение Фишера

Пусть V и W – независимые СВ, распределение по закону χ^2 со степенями свободы $\nu_1 = m$ и $\nu_2 = n$ соответственно. Тогда величина

$$F = \frac{V/m}{W/n} \qquad (1.19)$$

имеет *распределение Фишера со степенями свободы* $\nu_1 = m$ *и* $\nu_2 = n$ ($F \sim F_{m,n}$). Таким образом, распределение Фишера F определяется двумя параметрами – числами степеней свободы m и n.



$$M(F) = \frac{n}{n-2} \quad (n > 2),$$

$$D(F) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (n > 4).$$