Нормальное распределение.

► Нормальным называют распределение вероятностей непрерывной случайной величины X, плотность которого имеет вид

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

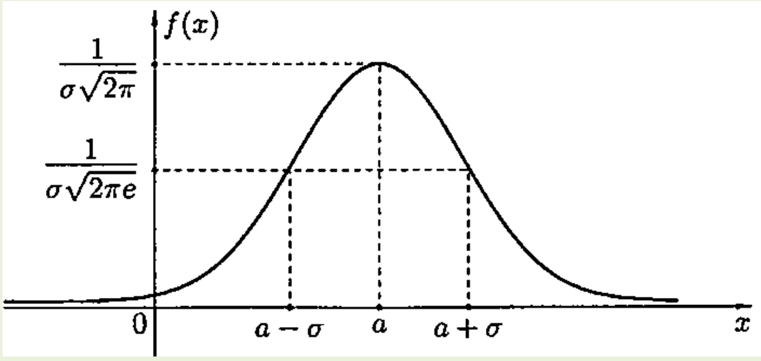
где параметры a и σ имеют смысл:

$a=M(X)$ - математическое ожидание;

$\sigma=\sigma(X)$ - среднее квадратическое отклонение ($\sigma^2=D(X)$ - дисперсия).

Нормальное распределение с параметрами a и σ кратко записывают как $N(a;\sigma)$. Нормальный закон распределения называют также законом Гаусса.

► График плотности распределения вероятности нормального закона называется нормальной кривой или кривой Гаусса:



Нормальное распределение $N(0;1)$ с параметрами $a=0$ и $\sigma=1$ называется нормированным или стандартным. Плотность нормированного нормального распределения

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

называется функцией Гаусса, а её график - кривой вероятностей.

► Вероятность того, что абсолютная величина отклонения с.в. от математического ожидания не превышает величину $\delta>0$:

$$P(|X-a|<\delta)=2\cdot\Phi\left(\frac{\delta}{\sigma}\right),$$

где X - непрерывная случайная величина,

a - математическое ожидание,

δ - граница отклонения абсолютной величины случайной величины,

$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx$ - функция Лапласа,

σ - среднее квадратическое отклонение.

► Вероятность того, что с.в. X примет значение, принадлежащее интервалу $(\alpha;\beta)$:

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),$$ где $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx$ - функция Лапласа.

Литература:

1) Гмурман В.Е. "Теория вероятностей и математическая статистика", 2005, стр. 127;
1) Гмурман В.Е. "Руководство к решению задач по теории вероятностей и математической статистике", 2005, стр. 109;
3) Письменный Д.Т. "Конспект лекций по теории вероятностей и математической статистике", 2004, стр. 96;
4) Кремер Н.Ш. " Теория вероятностей и математическая статистика", 2006, стр. 61;
5) Кремлёв А.Г. "Математика. Раздел "Статистика" ", 2001, издание УрГЮА (г. Екатеринбург).

1) Распределение с.в. X подчинено нормальному закону с параметрами $a=15$ и $\sigma=10$. Записать $f(x)$, $F(x)$, вычислить $P(3;30)$, $P(|X-15|<9)$.

Известны математическое ожидание и среднеквадратическое отклонение данного нормального распределения:

$$M(X)=a=15$$

$$\sigma=10$$

$$D(X)=\sigma^2=10^2=100$$

► Плотность вероятности (плотность распределения вероятностей):

$$f(x)=\frac{1}{\sigma\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}=\frac{1}{10\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-15)^2}{2\cdot 10^2}}=\frac{1}{10\sqrt{2\pi}}\cdot e^{-\frac{(x-15)^2}{200}}$$

► Функция распределения вероятностей для нормального распределения в общем случае

$$F(x)=\int_{-\infty}^{x} f(z)\,dz=\frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{-\infty}^{x} e^{-\frac{(z-a)^2}{2\sigma^2}}dz=\left[\begin{array}{l} t=\dfrac{z-a}{\sigma} \\ z=a+t\sigma \\ dz=\sigma\,dt \\ z\to-\infty:\ t\to-\infty \end{array}\right]=$$

$$=\frac{1}{\sqrt{2\pi}}\cdot\int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}}\cdot\int_{-\infty}^{0} e^{-\frac{t^2}{2}}dt+\frac{1}{\sqrt{2\pi}}\cdot\int_{0}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}}dt$$

Первый интеграл

$$\int_{-\infty}^{0} e^{-\frac{t^2}{2}}dt=\frac{1}{2}\cdot\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt=\frac{\sqrt{2}}{2}\cdot\int_{-\infty}^{\infty} e^{-\left(\frac{t}{\sqrt{2}}\right)^2}d\left(\frac{t}{\sqrt{2}}\right)=\frac{\sqrt{2}}{2}\cdot\sqrt{\pi}=\sqrt{\frac{\pi}{2}}$$

(в силу чётности подынтегральной функции и того, что интеграл Эйлера-Пуассона равен $\sqrt{\pi}$).

Второй интеграл - функция (интеграл вероятностей) Лапласа

$$\frac{1}{\sqrt{2\pi}}\cdot\int_{0}^{\frac{x-a}{\sigma}} e^{-\frac{t^2}{2}}dt=\Phi\left(\frac{x-a}{\sigma}\right)$$

Итак,

$$F(x)=\frac{1}{\sqrt{2\pi}}\cdot\sqrt{\frac{\pi}{2}}+\Phi\left(\frac{x-a}{\sigma}\right)=\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right).$$

В данном случае

$F(x)=\frac{1}{2}+\Phi\left(\frac{x-10}{4}\right)$, где функция Лапласа $\Phi(x)$ определена как $\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_{0}^{x} e^{-\frac{t^2}{2}}dt$

(необходимо уточнить, какое используем определение функции Лапласа, т.к. существуют три различных определения функции Лапласа)

► Вероятность попадания случайной величины X на промежуток $[3;30]$:

$$P(3;30)=\int_3^{30} f(x)\,dx=\int_3^{30}\frac{1}{10\sqrt{2\pi}}\cdot e^{-\frac{(x-15)^2}{200}}dx=\frac{1}{10\sqrt{2\pi}}\cdot\int_3^{30}e^{-\frac{(x-15)^2}{200}}dx\approx 0,818$$

Этот же результат можно получить, используя табулированную (т.е. просчитанную и занесённую в таблицы) функцию Лапласа $\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx$:

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right)$$

$$P(3<X<30)=\Phi\left(\frac{30-15}{10}\right)-\Phi\left(\frac{3-15}{10}\right)=\Phi(1,5)-\Phi(-1,2)=$$

$$=\Phi(1,5)+\Phi(1,2)=0,4332+0,3849\approx 0,8181$$

► Вероятность того, что абсолютная величина отклонения меньше числа $\delta=9$:

$$P\left(|X-a|<\delta\right)=2\cdot\Phi\left(\frac{\delta}{\sigma}\right),$$

где X - непрерывная случайная величина,
a - математическое ожидание случайной величины,
δ - граница отклонения абсолютной величины случайной величины,
$\Phi(x)$ - функция Лапласа,
σ - среднее квадратическое отклонение.

$$P\left(|X-15|<9\right)=2\cdot\Phi\left(\frac{9}{10}\right)=2\cdot\Phi(0,9)=2\cdot 0,3159=0,6318$$

Значение функции Лапласа $\Phi\left(\frac{9}{10}\right)$ найдём по таблице или вычислим:

$$\Phi\left(\frac{9}{10}\right)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^{0,9}e^{-\frac{t^2}{2}}dt\approx 0,31594$$

Примечание. В литературе встречаются три различных определения функции Лапласа

а) $\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^x e^{-\frac{t^2}{2}}dt$ (в 1, 3, 5)

б) $\Phi(x)=\frac{2}{\sqrt{2\pi}}\cdot\int_0^x e^{-\frac{t^2}{2}}dt$ (в 2, 4)

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

отличающиеся по величине результата в два раза. Будем использовать первое (а) определение функции Лапласа.

Литература:

1) Ниворожкина Л.И., Морозова З.А. "Математическая статистика с элементами теории вероятностей в задачах с решениями", 2005, стр. 5, 29;
2) Белько И.В., Свирид Г.Л. "Теория вероятностей и математическая статистика: примеры и задачи", 2004, стр. 106... ;
3) Письменный Д.Т. "Конспект лекций по теории вероятностей и математической статистике", 2004;
4) Кремер Н.Ш. " Теория вероятностей и математическая статистика", 2006, стр. 165;
5) Гмурман В.Е. "Руководство к решению задач по теории вероятностей и математической статистике", 2005, стр. 109;
6) Кремлёв А.Г. "Математика. Раздел "Статистика" ", 2001, издание УрГЮА (г. Екатеринбург).

2) Коробки с конфетами упаковываются автоматически со средней массой 540 г. Полагая, что средняя масса коробок распределена по нормальному закону со средним квадратическим отклонением 30 г, найти:
а) в какой интервал, симметричный относительно математического ожидания, укладываются 88 % коробок;
б) какой процент коробок имеет массу в пределах от 500 до 540 г.

а)

Вероятность того, что абсолютная величина отклонения меньше положительного числа δ :

$$P\left(\left|X-a\right|<\delta\right)=2\cdot\Phi\left(\frac{\delta}{\sigma}\right),$$

где X - непрерывная случайная величина,
a - математическое ожидание,
δ - граница отклонения абсолютной величины случайной величины,
$\Phi(t)$ - функция Лапласа,
σ - среднее квадратическое отклонение.

По условию задачи:
$\sigma=\sigma(X)=30$ г,
$a=M(X)=540$ г.

Обозначим A событие "попадание массы коробки в интервал, в который попадают 88 % коробок"; т.е. $P(A)=0,88$. Тогда

$$P(A)=2\cdot\Phi\left(\frac{\delta}{30}\right)=0,88$$

$$\Phi\left(\frac{\delta}{30}\right)=0,44$$

$$\frac{\delta}{30}=1,555$$

$$\delta\approx 47$$

Так что интервал по массе, симметричный относительно математического ожидания, в который попадают 88 % коробок, равен

$$(a-\delta\ ;\ a+\delta)=(540-47\ ;\ 540+47)=(493;587)\text{ г.}$$

б)

Вероятность того, что с.в. X примет значение, принадлежащее интервалу $(\alpha;\beta)$:

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right),\quad \text{где } \Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx \text{ - функция Лапласа.}$$

$$P(500<X<540)=\Phi\left(\frac{540-540}{30}\right)-\Phi\left(\frac{500-540}{30}\right)=\Phi(0)-\Phi\left(-\frac{4}{3}\right)=$$

$$=\Phi(0)+\Phi(1,33)=0+0,4082\approx 0,41$$

Т.е. 41% коробок имеет массу в пределах от 500 до 540 г.

3) Ошибка измерителя дальности подчинена нормальному закону с систематической ошибкой 20м и средним квадратическим отклонением 60м. Найти вероятность того, что измеренное значение дальности будет отклоняться от истинного не более, чем на 30м.

Если бы систематической ошибки не было, то вычисляли бы вероятность отклонения измеренного значения дальности от истинного от $a-30$ м до $a+30$ м. Но с учётом систематической погрешности этот промежуток смещается на 20 м (неважно, в какую сторону) и вычисляем вероятность отклонения измеренного значения дальности от истинного, например, от $a-10$ м до $a+50$ м.

По условию задачи $\sigma=60$ м, $\delta=30$ м; следовательно

$$P(a-10<X<a+50)=\Phi\left(\frac{50}{60}\right)-\Phi\left(\frac{-10}{60}\right)=\Phi(0,83)+\Phi(0,17)=0,2977+0,0662=0,3639$$

Ответ: $P=0,3639$.

4) Диаметр валика - случайная величина, распределённая по нормальному закону с параметрами $M(X)=10$ мм, $\sigma(X)=0,61$. Найти интервал, в который с вероятностью $P=0,9973$ будут заключены диаметры изготовляемых валиков.

1-й способ решения.

Вероятность того, что абсолютная величина отклонения меньше положительного числа δ :

$$P\left(\left|X-a\right|<\delta\right)=2\cdot\Phi\left(\frac{\delta}{\sigma}\right),$$

где X - непрерывная случайная величина,
a - математическое ожидание,
δ - граница отклонения абсолютной величины случайной величины,
$\Phi(t)$ - функция Лапласа,
σ - среднее квадратическое отклонение.

По условию задачи:

$\sigma(X)=0,61$ мм,

$a=M(X)=10$ мм.

Обозначим A событие "абсолютная величина отклонения не превзойдёт величины δ мм". Вероятность события A:

$$P\left(\left|X-10\right|<\delta\right)=2\cdot\Phi\left(\frac{\delta}{0,61}\right)=0,9973$$

$$\Phi\left(\frac{\delta}{0,61}\right)=0,49865$$

$$\frac{\delta}{0,61}=3,00 \quad \text{(по таблице)}$$

$$\delta=3\cdot0,61=1,83 \quad \text{(мм)}$$

Следовательно, интервал, в который с вероятностью 0,9973 будут заключены диаметры изготовляемых валиков:

$$(10-1,83\ ;\ 10+1,83)=(8,17\ ;\ 11,83) \text{ мм.}$$

2-й способ решения.

Для нормального закона распределения плотность вероятности:

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

где (по условию задачи) $a=M(X)=10$ мм,

$\sigma(X)=0{,}61$ мм.

Отклонение от математического ожидания $M(X)$, дающее искомый интервал, обозначим δ. Тогда

$$P(10-\delta<x<10+\delta)=\int\limits_{10-\delta}^{10+\delta}\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}dx=0{,}9973$$

$$\int\limits_{10-\delta}^{10+\delta}\frac{1}{0{,}61\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-10)^2}{2\cdot 0{,}61^2}}dx=0{,}9973$$

Решаем это уравнение в Mathcad (путём подбора) и находим значение δ, дающее требуемую вероятность 0,9973:

$\delta=1{,}830$ мм.

Следовательно, интервал, в который с вероятностью 0,9973 будут заключены диаметры изготовляемых валиков:

$(10-1{,}83\ ;\ 10+1{,}83)=(8{,}17\ ;\ 11{,}83)$ мм.

3-й способ решения.

Величина вероятности 0,9973 соответствует правилу "трёх сигм":

$P(|X-a|<3\sigma)=0{,}9973$

Следовательно,

$\delta=3\sigma=3\cdot 0{,}61=1{,}83$.

Искомый интервал:

$(10-1{,}83\ ;\ 10+1{,}83)=(8{,}17\ ;\ 11{,}83)$ мм.

5) Случайная величина X имеет нормальное распределение, $M(X)=20$. Вероятность того, что X примет значение, заключённое в интервале $(15;20)$, равна вероятности того, что значение X будет менее 15. Определите вероятность того, что X примет значение, большее 25.

► Вероятность того, что с.в. X примет значение, принадлежащее интервалу $(\alpha;\beta)$:

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right)$$

где X - непрерывная случайная величина,

a - математическое ожидание,

$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int\limits_0^t e^{-\frac{x^2}{2}}dx$ - функция Лапласа,

σ - среднее квадратическое отклонение.

В данном случае:

$P(15<X<20)=P(-\infty<X<15)$

$$\Phi\left(\frac{20-20}{\sigma}\right)-\Phi\left(\frac{15-20}{\sigma}\right)=\Phi\left(\frac{15-20}{\sigma}\right)-\Phi\left(\frac{-\infty-20}{\sigma}\right)$$

$$\Phi\left(\frac{0}{\sigma}\right)-\Phi\left(\frac{-5}{\sigma}\right)=\Phi\left(\frac{-5}{\sigma}\right)-\Phi\left(\frac{-\infty}{\sigma}\right)$$

$$\Phi(0)+\Phi\left(\frac{5}{\sigma}\right)=-\Phi\left(\frac{5}{\sigma}\right)+\Phi(+\infty)$$

$$\Phi(0)+2\Phi\left(\frac{5}{\sigma}\right)=\Phi(+\infty)$$

Поскольку $\Phi(0)=0,\;\; \Phi(+\infty)=0,5$, то

$$2\Phi\left(\frac{5}{\sigma}\right)=0,5$$

$$\Phi\left(\frac{5}{\sigma}\right)=0,25$$

Искомая вероятность:

$$P(25<X<+\infty)=\Phi\left(\frac{+\infty-20}{\sigma}\right)-\Phi\left(\frac{25-20}{\sigma}\right)=\Phi(+\infty)-\Phi\left(\frac{5}{\sigma}\right)=0,5-0,25=0,25$$

Ответ: $P(X>25)=0,25$.

6) Какова абсолютная величина отклонения частости события A от его вероятности, если при 100 повторных испытаниях наиболее вероятная частота появления события равна 36, а вероятность абсолютной величины указанного отклонения равна 0.9545.

Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события равна p, абсолютная величина отклонения относительной частоты (частости) появления события от вероятности появления события не превысит положительного числа ε, приближённо равна удвоенной функции Лапласа при $x=\varepsilon\cdot\sqrt{\frac{n}{pq}}$:

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=2\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right).$$

В данном случае

$$P\left\{\left|\frac{36}{100}-0,9545\right|\le\varepsilon\right\}=2\Phi\left(\varepsilon\cdot\sqrt{\frac{100}{0,9545\cdot(1-0,9545)}}\right)$$

$$2\Phi\left(\varepsilon\cdot\sqrt{\frac{100}{0,9545\cdot 0,0455}}\right)=0,9545$$

$$\Phi(\varepsilon\cdot 47,985)=0,47725$$

$$\varepsilon\cdot 47,985=2$$

$$\varepsilon=0,042$$

Ответ: $\varepsilon=0,042$.

7) Игральную кость бросают 80 раз. Найти с вероятностью 0,99 границы, в которых будет заключено число m выпадений шестёрки.

Вероятность появления шестёрки в одном бросании равна $p=\frac{1}{6}$.

Вычисляем:

$$2\Phi\left(\varepsilon\cdot\sqrt{\frac{80}{\frac{1}{6}\cdot\frac{5}{6}}}\right)=0,99$$

$$\Phi\left(\varepsilon\cdot\sqrt{576}\right)=0,495$$

$$\Phi(24\varepsilon)=0,495$$

$$24\varepsilon=2,58$$

$$\varepsilon=0,1075$$

Таким образом, с вероятностью 0,99 отклонение относительной частоты числа m выпадений шестёрки от вероятности $p=\frac{1}{6}\approx 0,167$ удовлетворяет неравенству

$$\left|\frac{m}{80}-\frac{1}{6}\right|\le 0,1075$$

$$-0,1075\le\frac{m}{80}-\frac{1}{6}\le 0,1075$$

$$-8,6\le m-13,3\le 8,6$$

$$4,7\le m\le 21,9$$

или

$$5\le m\le 22$$

Ответ: с вероятностью 0,99 число m выпадений шестёрки заключено в промежутке $5\le m\le 22$.

8) Подлежат исследованию 400 проб руды. Вероятность промышленного содержания металла в каждой пробе для всех проб одинакова и равна 0,8. Найти вероятность того, что доля проб с промышленным содержанием металла отклонится от вероятности промышленного содержания металла в каждой пробе не более, чем на 0,05.

Обозначим событие:

A - доля проб с промышленным содержанием металла отклонится от вероятности промышленного содержания металла в каждой пробе не более, чем на 0,05.

По условию задачи

$$n=400$$

$$p=0,8$$

$$q=1-p=1-0,8=0,2$$

$$\varepsilon=0,05$$

Вычисляем

$$P(A)=P\left(\left|\frac{m}{400}-0,8\right|\le 0,05\right)=2\Phi\left(0,05\cdot\sqrt{\frac{400}{0,8\cdot 0,2}}\right)=2\Phi(2,5)$$

По таблице [1, стр. 391] находим $\Phi(2,5)=0,4938$; следовательно, искомая вероятность

$$P(A)=2\Phi(2,5)=2\cdot 0,4938=0,9876.$$

Ответ: $P(A)=0,9876$

Литература:

1) Гмурман В.Е. "Руководство к решению задач по теории вероятностей и математической статистике", 2005, стр. 43 (задача 131);
2) Гмурман В.Е. "Теория вероятностей и математическая статистика", 2005, стр. 61;
3) Письменный Д.Т. "Конспект лекций по теории вероятностей и математической статистике", 2004, стр. 58;
4) Шириков В.Ф., Зарбалиев С.М. "Теория вероятностей", 2008, стр. 108.

9) Для определения средней урожайности пшеницы на поле произведена оценка урожая на участках небольшой площади. Полученные результаты сведены в таблицу:

Урожай (ц)	10	13	16	19	22	25	28	31	34
Количество участков	2	6	10	23	31	31	23	15	

Считая, что величина урожая на участках распределена нормально, найти вероятность того, что выборочное среднее значение урожая на участках отклоняется от среднего урожая на всём поле не более, чем на 0,5 ц.

Задан интервальный вариационный ряд:

Урожай (ц)	10 - 13	13 - 16	16 - 19	19 - 22	22 - 25	25 - 28	28 - 31	31 - 34
Количество участков	2	6	10	23	31	31	23	15

Перейдём к дискретному вариационному ряду, заменив интервалы их средним значением:

Урожай (ц)	11,5	14,5	17,5	20,5	23,5	26,5	29,5	32,5
Количество участков	2	6	10	23	31	31	23	15

Объём выборки:

$$n=\sum_{i=1}^{n} n_i = 2+6+10+23+31+31+23+15=141$$

Выборочное среднее (среднее арифметическое выборки)

$$\bar{x}_в = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i\, n_i =$$

$$=\frac{1}{141}\cdot\left(2\cdot 11,5+6\cdot 14,5+10\cdot 17,5+23\cdot 20,5+31\cdot 23,5+31\cdot 26,5+23\cdot 29,5+15\cdot 32,5\right)=$$

$$=24,628 \text{ (ц)}$$

Исправленная выборочная дисперсия

$$s^2=\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(x_i-\bar{x}_в\right)^2 n_i =$$

$$=\frac{1}{141-1}\cdot\left(\left(11,5-24,628\right)^2+\left(14,5-24,628\right)^2+\left(17,5-24,628\right)^2+\left(20,5-24,628\right)^2+\right.$$

$$\left.+\left(23,5-24,628\right)^2+\left(26,5-24,628\right)^2+\left(29,5-24,628\right)^2+\left(32,5-24,628\right)^2\right)=24,884$$

Исправленное среднее квадратическое отклонение

$$s=\sqrt{s^2}=\sqrt{24,884}=4,988 \text{ (ц)}$$

► Для нормального распределения вероятность того, что абсолютная величина отклонения меньше числа $\varepsilon>0$:

$$P\left(\left|X-a\right|<\varepsilon\right)=2\cdot\Phi\left(\frac{\varepsilon}{\mu}\right),$$

где X - непрерывная случайная величина,
a - математическое ожидание случайной величины,
ε - граница отклонения абсолютной величины случайной величины,
$\Phi(x)$ - функция Лапласа,

$\mu = \sqrt{\dfrac{s^2}{n}}$ - ошибка выборки,

s^2 - выборочное среднее квадратическое отклонение,
n - объём выборки.

$$P\left(\left|X-a\right|<0,5\right) = 2\cdot\Phi\left(\frac{0,5}{\sqrt{\dfrac{24,884}{141}}}\right) \approx 2\cdot\Phi(1,190) = 2\cdot 0,3830 = 0,7660$$

Значение функции Лапласа $\Phi\left(\dfrac{0,5}{\sqrt{\dfrac{24,884}{141}}}\right) \approx \Phi(1,190)$ найдём по таблице или вычислим:

$$\Phi(1,19) = \frac{1}{\sqrt{2\pi}}\cdot\int_0^{1,19} e^{-\frac{t^2}{2}}dt \approx 0,3830$$

Ответ: $p = 0,7660$.

Литература:

1) Гмурман В.Е. "Руководство к решению задач по теории вероятностей и математической статистике", 2005, стр. 174;
2) Кремер Н.Ш. "Теория вероятностей и математическая статистика", 2006, стр. 321;
3) Кремлёв А.Г. "Математика. Раздел “Статистика” “, методичка УрГЮА, 2001, стр. 47.

10) Из генеральной совокупности, распределённой по нормальному закону, взята выборка. Найти:

а) выборочную среднюю $\overline{x}_в$;

б) выборочное среднее квадратическое отклонение $\sigma_в$;

в) с надёжностью $\gamma = 0,95$ доверительный интервал для оценки математического ожидания a генеральной совокупности при известной дисперсии σ^2.

x_i	130	140	150	160	170	180	190
n_i	5	10	30	25	15	10	5

а)

Вычисляем выборочную среднюю $\overline{x}_â$ по формуле средней взвешенной:

$$\overline{x}_в = \frac{1}{n}\cdot\sum_{i=1}^{k} n_i\cdot x_i$$

(выборочная средняя есть средняя взвешенная значений признака с весами, равными соответствующим частотам).

$$\overline{x}_в = \frac{1}{100}\cdot(130\cdot 5+140\cdot 10+150\cdot 30+160\cdot 25+170\cdot 15+180\cdot 10+190\cdot 5) = 158,5$$

б)

Для вычисления выборочной дисперсии $D_в$ при разных частотах признака используем формулу

$$D_в = \frac{1}{n}\cdot\sum_{i=1}^{k} n_i\cdot\left(x_i-\overline{x}_в\right)^2$$

(выборочная дисперсия есть средняя взвешенная квадратов отклонений с весами, равными соответствующим частотам).

$$D_в = \frac{1}{100}\cdot(\ 5\cdot(130-158,5)^2 + 10\cdot(140-158,5)^2 + 30\cdot(150-158,5)^2 + 25\cdot(160-158,5)^2 +$$
$$+15\cdot(170-158,5)^2 + 10\cdot(180-158,5)^2 + 5\cdot(190-158,5)^2 \ ) = 212,75$$

Выборочное среднее квадратическое отклонение

$$\sigma_в = \sqrt{D_в} = \sqrt{212,75} \approx 14,586$$

(выборочное среднее квадратическое отклонение [стандартное отклонение] есть корень квадратный из выборочной дисперсии).

в)

Доверительный интервал для оценки математического ожидания a генеральной совокупности при известной дисперсии σ^2 с надёжностью γ находим по формуле

$$\overline{x}_в - \delta < a < \overline{x}_в + \delta,$$

где a - оцениваемое неизвестное математическое ожидание генеральной совокупности,

$\overline{x}_в$ - выборочная средняя,

δ - точность оценки (предельная ошибка выборки), $\delta = \dfrac{t\cdot\sigma_в}{\sqrt{n}}$,

$\sigma_в$ - выборочное среднее квадратическое отклонение.

Величину t найдём из равенства

$$2\Phi(t) = \gamma,$$

где $\Phi(t)$ - функция Лапласа,

γ - заданная надёжность.

Доверительный интервал $\left(\overline{x}_в - \delta < a < \overline{x}_в + \delta\right)$ покроет неизвестный параметр a с надёжностью γ.

Вычисляем статистику t (коэффициент, связывающий среднюю ошибку выборки $\dfrac{\sigma_в}{\sqrt{n}}$ с предельной ошибкой δ):

$$2\Phi(t) = 0,95$$
$$\Phi(t) = 0,475$$
$$t = 1,96$$

Получаем точность оценки:

$$\delta = \frac{t\cdot\sigma}{\sqrt{n}} = \frac{1,96\cdot 14,586}{\sqrt{100}} \approx 2,9$$

С надёжностью $\gamma = 0,95$ доверительный интервал

$$\left(\overline{x}_в - \delta \ ; \ \overline{x}_в + \delta\right) \approx \left(158,5 - 2,9 \ ; \ 158,5 + 2,9\right) = \left(155,6 \ ; \ 161,4\right)$$

покрывает неизвестный параметр a (математическое ожидание генеральной совокупности).

11) Для изучения общественного мнения о работе правоохранительных органов в порядке механического отбора было опрошено 3500 человек (1% общей численности городского населения). Из числа опрошенных 2980 человек положительно оценили работу правоохранительных органов. С вероятностью 0.997 определить пределы, в которых находится доля лиц, положительно оценивающих работу правоохранительных органов.

Механический способ заключается в отборе элементов из генеральной совокупности с заданной цикличностью (в данном случае опрашивался каждый 100-й человек).

Обозначим p - доля оптимистов в генеральной совокупности (генеральная доля).

Доверительный интервал для генеральной доли определяется неравенством

$$w-\varepsilon < p < w+\varepsilon$$

где выборочная доля (доля оптимистов, вычисленная по данным выборки)

$$w = \frac{2980}{3500} = \frac{149}{175} \approx 0,85143 .$$

Средняя ошибка механической выборки для доли:

$$\mu = \sqrt{\frac{w\cdot(1-w)}{n}} = \sqrt{\frac{\frac{149}{175}\cdot\left(1-\frac{149}{175}\right)}{3500}} = 0,006012$$

По заданной вероятности $\gamma = 0,997$ определим статистику t:

$$2\Phi_0(t) = 0,997$$

$$\Phi_0(t) = 0,4985$$

$$t \approx 2,97$$

Предельная ошибка выборочной доли составит

$$\varepsilon = t\cdot\mu = 2,97\cdot 0,006012 \approx 0,01786$$

Вычислим границы для генеральной доли p:

$$w-\varepsilon < p < w+\varepsilon$$

$$0,85143-0,01786 < p < 0,85143+0,01786$$

$$0,83357 < p < 0,86929$$

Итак, с вероятностью 0,997 можно утверждать, что доля лиц, положительно оценивающих работу правоохранительных органов, находится в пределах от 0,834 до 0,869 (или от 83,4% до 86,9%).

Литература:

1) Кремлёв А.Г. "Математика. Раздел "Статистика" ", 2001, издание УрГЮА (г. Екатеринбург), стр. 53 (задача 12);
2) Ниворожкина Л.И., Морозова З.А. "Математическая статистика с элементами теории вероятностей в задачах с решениями", 2005, стр. 193.

12) Данные наблюдений сведены в группы и представлены в виде интервального статистического ряда. Первая строка - интервалы наблюдавшихся значений случайной величины X. Вторая - соответствующие им частоты. Требуется:

а) построить гистограмму и полигон относительных частот;

б) найти эмпирическую функцию распределения и построить её график;

в) вычислить числовые характеристики выборки: выборочную среднюю, выборочную дисперсию, выборочное среднее квадратическое отклонение;

г) предполагая, что исследуемая случайная величина распределена по нормальному закону, записать плотность вероятности случайной величины X и построить её график на одном чертеже с гистограммой относительных частот (график выравнивающей кривой);

д) найти теоретические частоты нормального закона распределения. При уровне значимости $\alpha = 0,05$ проверить по критерию согласия хи-квадрат гипотезу о нормальном законе распределения;

е) найти интервальные оценки параметра a нормального закона распределения. Доверительную вероятность принять равной 0,95.

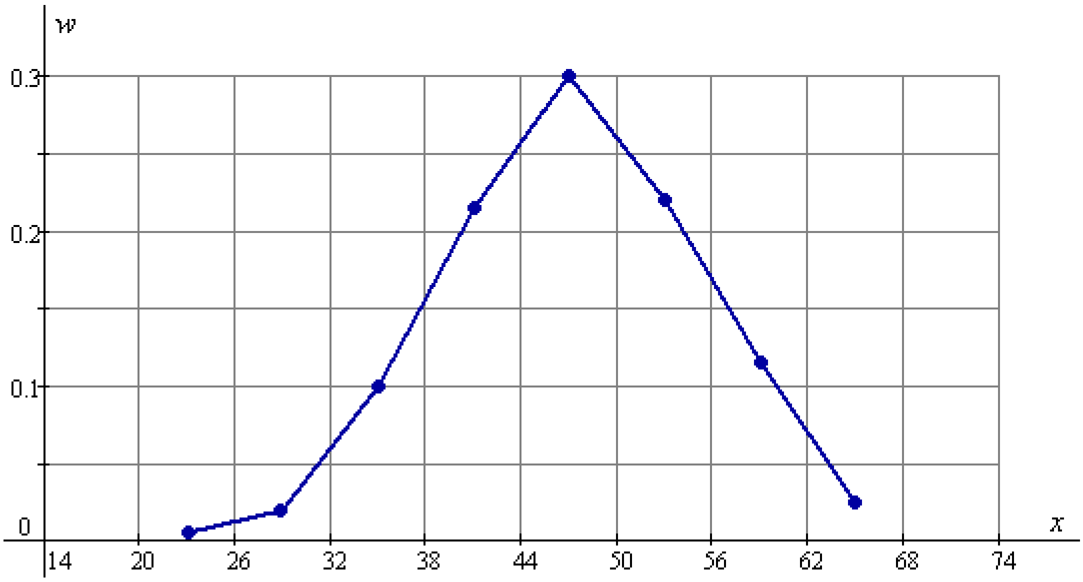
Интервалы	(20; 26)	(26; 32)	(32; 38)	(38; 44)	(44; 50)	(50; 56)	(56; 62)	(62; 68)
Частоты	1	4	20	43	60	44	23	5

а)

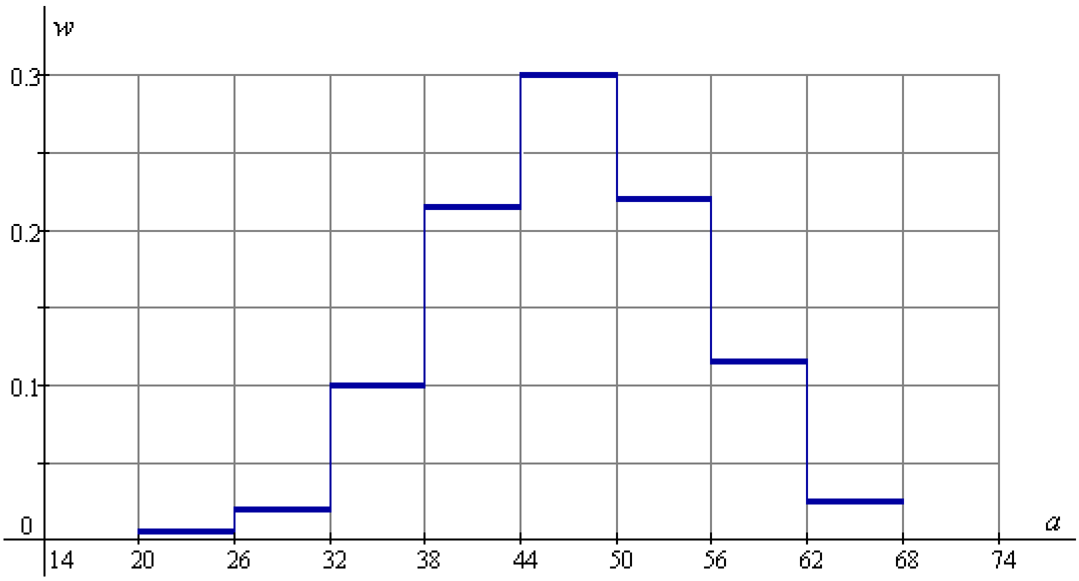
таблица 1

Интервалы a	(20; 26)	(26; 32)	(32; 38)	(38; 44)	(44; 50)	(50; 56)	(56; 62)	(62; 68)
Частоты n	1	4	20	43	60	44	23	5
Относительные частоты w	0,005	0,020	0,100	0,215	0,300	0,220	0,115	0,025

Полигон относительных частот ($w_i(x_i)$); x_i - середины интервалов.



Гистограмма относительных частот ($w_i(a_i)$).



б)

Построим эмпирическую (по опытным данным) функцию распределения случайной величины X:

при $x \le 20$ $\quad F(x) = 0$;

при $20 < x \le 26$ $\quad F(x) = 0 + 0,005 = 0,005$;

при $26 < x \le 32$ $\quad F(x) = 0,005 + 0,02 = 0,025$;

при $32 < x \le 38$ $\quad F(x) = 0,025 + 0,1 = 0,125$;

при $38 < x \le 44$ $\quad F(x) = 0,125 + 0,215 = 0,34$;

при $44 < x \le 50$ $\quad F(x) = 0,34 + 0,3 = 0,64$;

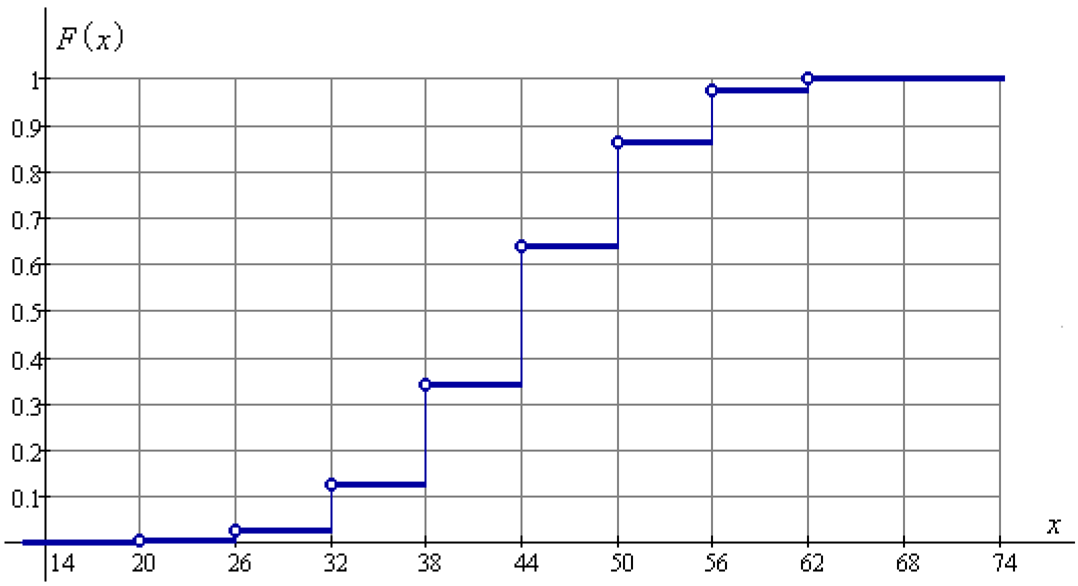
при $50 < x \le 56$ $\quad F(x) = 0,64 + 0,22 = 0,86$;

при $56 < x \le 62$ $\quad F(x) = 0,86 + 0,115 = 0,975$;

при $x > 62$ $\quad F(x) = 0,975 + 0,025 = 1$.

$$F(x) = \begin{cases} 0 & при\ x \le 20 \\ 0,005 & при\ 20 < x \le 26 \\ 0,025 & при\ 26 < x \le 32 \\ 0,125 & при\ 32 < x \le 38 \\ 0,34 & при\ 38 < x \le 44 \\ 0,64 & при\ 44 < x \le 50 \\ 0,86 & при\ 50 < x \le 56 \\ 0,975 & при\ 56 < x \le 62 \\ 1 & при\ x > 62 \end{cases}$$

Изобразим график функции распределения:



в)

Вычислим числовые характеристики выборки:

выборочную среднюю,

выборочную дисперсию,

выборочное среднее квадратическое отклонение.

таблица 2

Середины интервалов x	23	29	35	41	47	53	59	65
Относительные частоты w	0,005	0,020	0,100	0,215	0,300	0,220	0,115	0,025
Относит. частота на единицу варианты $b = w/6$	0,00083	0,00333	0,01667	0,03583	0,05	0,03667	0,01917	0,00417

Выборочная средняя

$$\overline{x}_в = \sum_{i=1}^{8} x_i w_i = 47,18$$

Выборочная дисперсия

$$D_в = \sum_{i=1}^{8} \left(x_i - \overline{x}\right)^2 \cdot p_i = 64,0476$$

Исправленная дисперсия

$$S^2 = \frac{n}{n-1} \cdot D_в = \frac{200}{199} \cdot 64,0476 = 64,3694$$

Выборочное среднее квадратическое отклонение

$$\sigma_в = \sqrt{D_в} = \sqrt{64,0476} = 8,003$$

г)

Предполагая, что исследуемая случайная величина распределена по нормальному закону, записать плотность вероятности случайной величины X и построить её график на одном чертеже с гистограммой относительных частот (график выравнивающей кривой).

Для нормального закона распределения плотность вероятности:

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

где (по условию задачи) $a = \overline{x} = 47,18$,

$\sigma(X) = 8,003$,

т.е.

$$f(x) = \frac{1}{8,003 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-47,18)^2}{2 \cdot 64,0476}}$$

На графике для сопоставимости изобразим плотность вероятности случайной величины X и гистограмму относительных частот **на единицу варианты** (вторая строка таблицы 2).

Для построения выравнивающей (сглаживающей) кривой, а также для проведения дальнейшей статистической обработки составим следующую таблицу:

таблица 3

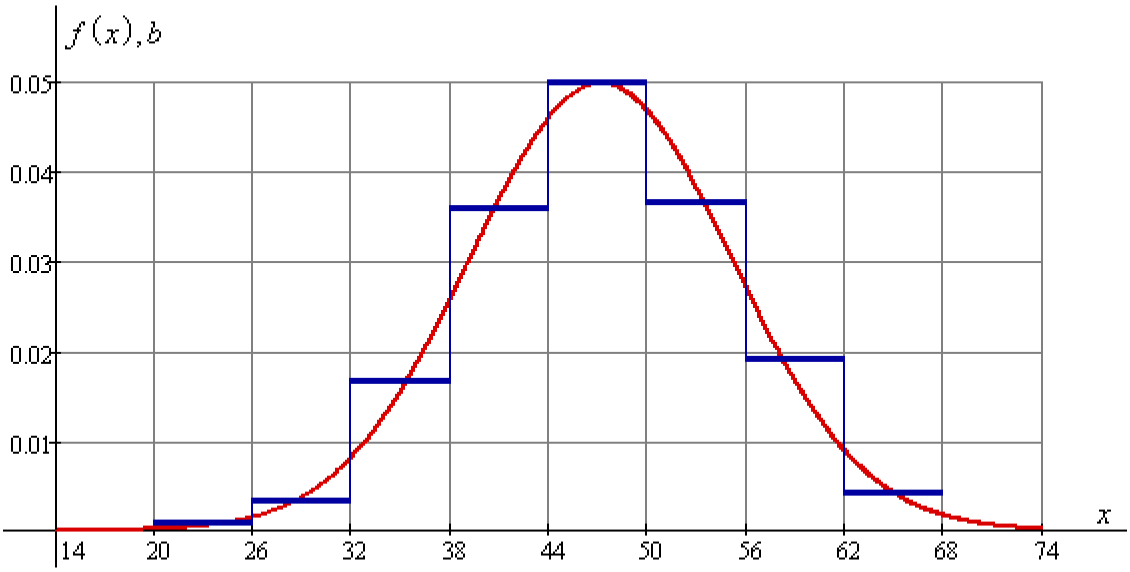
1	2	3	4	5	6	7	8	9	10
x_i	m_i	$x_i - \overline{x}$	$z_i = \frac{x_i - \overline{x}}{s}$	$\varphi(z_i)$	$y_i = f(x_i) = \varphi(z_i)/s$	$p_i = h \cdot y_i$	$n \cdot p_i$	$m_i - n \cdot p_i$	$\frac{(m_i - np_i)^2}{np_i}$
23	1	-24,18	-3,021	0,0042	0,00052	0,00312	0,624	0,376	0,227
29	4	-18,18	-2,272	0,0303	0,00379	0,02274	4,548	-0,548	0,066
35	20	-12,18	-1,522	0,1257	0,01571	0,09426	18,852	1,148	0,070
41	43	-6,18	0,772	0,2966	0,03706	0,22236	44,472	-1,472	0,049
47	60	-0,18	0,022	0,3989	0,04984	0,29904	59,808	0,192	0,001
53	44	5,82	0,727	0,3056	0,03819	0,22914	45,828	-1,828	0,073
59	23	11,82	1,477	0,1334	0,01667	0,10002	20,004	2,996	0,449
65	5	17,82	2,227	0,0332	0,00415	0,02489	4,978	0,022	0,000
									0,935

(здесь s - с.к.о.)

Здесь $\varphi(z_i)$ вычисляется по формуле плотности стандартного нормального распределения

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$$ (можно использовать таблицы значений этой функции).

Кривая $y_i = f(x_i)$ изображена далее в Mathcad 14 для наглядности на одном чертеже с гистограммой распределения и хорошо согласуется с данными выборки (пока это только визуальная оценка).



д)

Найти теоретические частоты нормального закона распределения. При уровне значимости $\alpha = 0,05$ проверить по критерию согласия хи-квадрат гипотезу о нормальном законе распределения.

Оценим согласованность выбранного теоретического распределения с опытными данными в соответствии с критерием Пирсона (критерий χ^2 [хи квадрат]). Критерий согласия Пирсона - наиболее часто употребеляемый критерий для проверки простой гипотезы о законе распределения (гипотезу, однозначно фиксирующую распределение наблюдений, называют простой (в ней речь идёт об одном значении параметра), в противном случае - сложной). Суммируя величины колонки 10 в таблице 3, получаем величину $\chi_{\text{р}}^2 = \sum_{i=1}^{9} \frac{(m_i - np_i)^2}{np_i} \approx 0,94$, характеризующую меру расхождения теоретического и статистического распределений. Чем больше величина χ^2, тем больше это расхождение.

Задана величина критического значения уровня значимости $\alpha_{\text{кр.}} = P\left(\chi^2 \geq \chi_{\text{р}}^2\right) = 0,05$, который должен быть превышен для принятия гипотезы о законе распределения. Отметим, что принятие гипотезы всегда происходит на некотором субъективно принятом уровне значимости и основывается на значениях конечной выборки (в данном случае $n = 200$). Обычно ориентируются на критическое значение уровня значимости $\alpha = 0,01 \ldots 0,05$ ($\alpha = 0,01$ соответствует более мягкому подходу в оценке гипотезы, $\alpha = 0,05$ соответствует более строгому подходу в оценке гипотезы).

По таблице критических точек χ^2 - распределения исходя из числа степеней свободы $r = k - 3 = 8 - 3 = 5$ (k - число разрядов) и реализовавшегося значения $\chi_{\text{р}}^2 = 0,94$ находим величину $\alpha = P\left(\chi^2 \geq \chi_{\text{р}}^2\right) = P\left(\chi^2 \geq 0,94\right) \approx 0,97$, т.е. вероятность того, что величина, распределённая по закону χ^2, превысит значение $\chi_{\text{р}}^2$.

Таблицы довольно грубы; проще и быстрее значение α вычисляется, например в Mathcad 14:

$$\varphi(x,r) := \frac{1}{2^{\frac{r}{2}} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}} \qquad \alpha := \int_{0.94}^{\infty} \varphi(x,5)\,dx = 0.967$$

Поскольку вероятность $\alpha > \alpha_{\text{кр.}}$ $(0,97 > 0,05)$, то можно считать, что эмпирически принятое теоретическое нормальное распределение не противоречит опытным данным и гипотеза о виде распределения и о его параметрах может быть принята. Другими словами, принятая гипотеза не противоречит имеющимся выборочным данным на уровне значимости $0,05$, или уровне надёжности (уровне доверия) $1 - 0,05 = 0,95$.

е)

Найти интервальные оценки параметра a нормального закона распределения. Доверительную вероятность принять равной 0,95.

Вероятность того, что абсолютная величина отклонения меньше положительного числа δ:

$$P\left(\left|X - a\right| < \delta\right) = 2 \cdot \Phi\left(\frac{\delta}{\sigma}\right),$$

где X - непрерывная случайная величина,
a - математическое ожидание,
δ - граница отклонения абсолютной величины случайной величины,
$\Phi(t)$ - функция Лапласа (берём из таблиц),
σ - среднее квадратическое отклонение.

По условию задачи:

$\sigma(X) = 8$,

$a = M(X) = 47,18$

Величина вероятности задана:

$$2 \cdot \Phi\left(\frac{\delta}{8}\right) = 0,95$$

$$\Phi\left(\frac{\delta}{8}\right) = 0,475$$

$$\frac{\delta}{8} = 1,96$$

$$\delta \approx 15,68$$

Так что с вероятностью 0,95 параметр a нормального распределения (мат. ожидание генеральной совокупности) находится в интервале

$$(a - \delta\ ;\ a + \delta) = (47,18 - 15,68\ ;\ 47,18 + 15,68) = (31,5\ ;\ 62,86).$$

13) Для имеющейся совокупности опытных данных (выборки) требуется:
1) Построить статистический ряд и гистограмму распределения.
2) Вычислить следующие статистики распределения:
выборочную среднюю,
выборочное среднеквадратическое отклонение,
коэффициент вариации,
асимметрию,
экцесс.
Раскрыть смысловую сторону каждой статистики.
3) Обосновать выбор теоретического распределения и методом моментов найти его параметры.
4) Построить теоретическую кривую распределения.
5) Проверить согласованность теоретического и выборочного распределений, применяя критерий согласия Пирсона.

В результате эксперимента реализовались 100 значений измеряемой величины, которые получены случайным выбором из генеральной совокупности. Итак, задана выборка (или простая статистическая совокупность), состоящая из 100 значений ($n = 100$):

19,9	14,0	17,2	15,3	14,3	18,8	23,1	6,6	12,1	14,5
16,7	12,7	15,4	15,0	10,5	17,5	13,4	14,7	12,6	18,8
17,9	19,6	17,1	11,7	17,5	9,4	16,5	11,0	14,3	16,1
15,4	15,7	8,1	11,7	15,2	14,8	11,5	14,8	11,8	17,4
13,0	10,6	15,7	22,0	20,1	14,8	10,0	19,7	17,4	17,6
19,7	15,9	8,7	13,3	13,1	21,0	22,7	20,0	22,9	7,9
20,3	14,3	9,9	12,1	13,7	19,1	19,5	15,7	17,7	17,1
13,0	19,0	8,0	21,9	16,6	21,7	18,8	11,4	29,2	21,7
15,6	12,0	9,8	25,0	20,3	14,3	10,6	12,8	20,8	12,8
19,9	16,1	9,8	15,7	21,0	10,2	6,2	12,0	12,2	18,5

1

Выбираем из этой совокупности наименьшее и наибольшее значения:

$$x_{наим.} = 6,2$$

$$x_{наиб.} = 29,2$$

Округляем граничные значения до 6 и 30. Разбиваем этот интервал на частичные интервалы (разряды) и подсчитываем количество наблюдений, приходящееся на каждый разряд. Обычно число разрядов выбирают в промежутке 7...15. Пусть количество разрядов $k = 8$, и их длина h постоянна; тогда

$$h = \frac{30-6}{8} = 3 .$$

Составим интервальный статистический ряд (табл. 1).

таблица 1

Номер i	1	2	3	4	5	6	7	8
i-й разряд	6-9	9-12	12-15	15-18	18-21	21-24	24-27	27-30
Число наблюдений m_i	6	16	24,5	25,5	18	8	1	1
Частота p_i^*	0,03	0,075	0,08	0,16	0,16	0,185	0,185	0,085
Частость $b_i^* = p_i^* / h$	0,015	0,0375	0,04	0,08	0,08	0,0925	0,0925	0,0425

$$(n = \sum_{i=1}^{8} m_i = 100 ,\ h = 3)$$

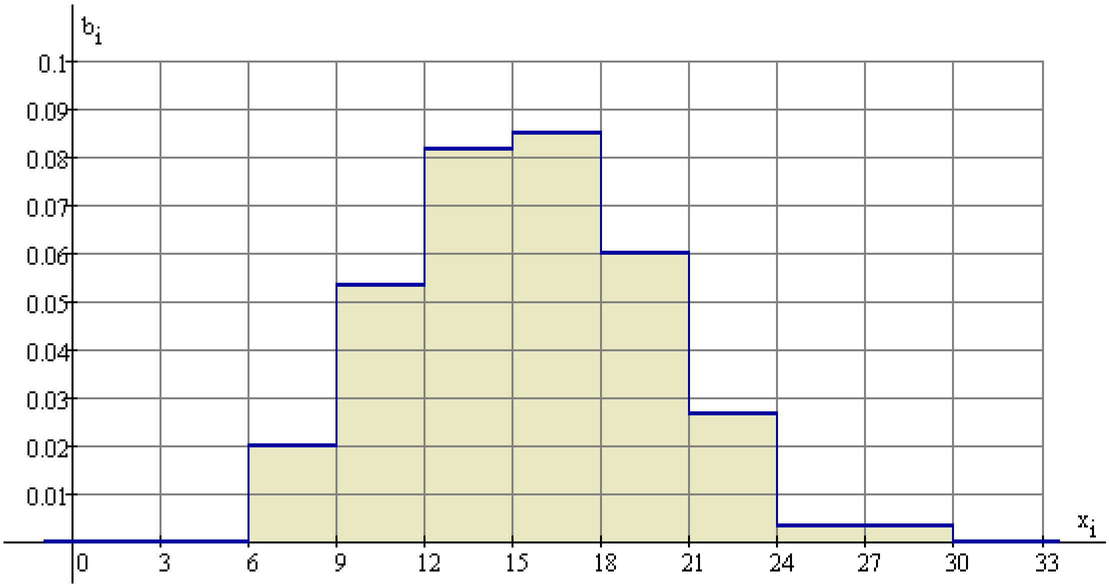
В последних разрядах оказалось слишком мало наблюдений ($m_i < 3$), поэтому объединим их:

таблица 2

Номер i	1	2	3	4	5	6	7
i-й разряд	6-9	9-12	12-15	15-18	18-21	21-24	24-30
Число наблюдений m_i	6	16	24,5	25,5	18	8	2
Частота p_i^*	0,06	0,16	0,245	0,255	0,18	0,08	0,02
Частость b_i^*	0,02	0,0533	0,0817	0,085	0,06	0,0267	0,0033

Здесь $p_i^* = \dfrac{m_i}{n}$, $b_i^* = \dfrac{p_i^*}{h}$.

Используя данные таблицы 2, построим гистограмму распределения (зависимость $b_i\left(x_i\right)$).



2

Для вычисления статистик распределения интервальный ряд табл.2 заменим дискретным статистическим рядом, ставя в соответствие каждому разряду его середину:

таблица 3

Номер i	1	2	3	4	5	6	7
Середина интервала x_i	7,5	10,5	13,5	16,5	19,5	22,5	27
Число наблюдений m_i	6	16	24,5	25,5	18	8	2
Частота p_i^*	0,06	0,16	0,245	0,255	0,18	0,08	0,02

По данным табл. 3 вычислим выборочные начальные моменты распределения по следующим формулам:

$$a_1 = \sum_{i=1}^{k} x_i \cdot p_i^* , \quad a_2 = \sum_{i=1}^{k} x_i^2 \cdot p_i^* , \quad a_3 = \sum_{i=1}^{k} x_i^3 \cdot p_i^* , \quad a_4 = \sum_{i=1}^{k} x_i^4 \cdot p_i^* .$$

Получаем:

$$a_1 = 15,50$$

$$a_2 = 258,62$$

$$a_3 = 4598,40$$

$$a_4 = 86331,12$$

Вычисления в Mathcad 14:

$$B := \begin{pmatrix} 7.5 \\ 10.5 \\ 13.5 \\ 16.5 \\ 19.5 \\ 22.5 \\ 27 \end{pmatrix} \qquad C := \begin{pmatrix} 0.06 \\ 0.16 \\ 0.245 \\ 0.255 \\ 0.18 \\ 0.08 \\ 0.02 \end{pmatrix}$$

$$a_1 := B \cdot C \qquad a_1 = 15.495$$

$$a_2 := \overrightarrow{B^2} \cdot C \qquad a_2 = 258.615$$

$$a_3 := \overrightarrow{B^3} \cdot C \qquad a_3 = 4598.40375$$

$$a_4 := \overrightarrow{B^4} \cdot C \qquad a_4 = 86331.11625$$

$$a_2 - a_1^2 = 18.51997 \qquad \sqrt{a_2 - a_1^2} = 4.303$$

$$a_3 - 3 \cdot a_1 \cdot a_2 + 2 \cdot a_1^3 = 17.2303$$

$$a_4 - 4 \cdot a_3 \cdot a_1 + 6 \cdot a_2 \cdot a_1^2 - 3 \cdot a_1^4 = 938.23808$$

Определяем статистики выборочного распределения.

Выборочная средняя $\bar{x} = a_1 = 15,50$. Выборочная средняя - статистический аналог математического ожидания.

Выборочная дисперсия $s^2 = a_2 - a_1^2 = 258,62 - 15,50^2 = 18,5200$. Выборочная дисперсия - аналог дисперсии; характеризует рассеяние случайной величины вокруг её среднего значения.

Выборочное среднеквадратическое отклонение $s = \sqrt{s^2} = \sqrt{18,5200} \approx 4,303$.

Коэффициент вариации $V = \dfrac{s}{\bar{x}} \cdot 100\% = \dfrac{4,303}{15,50} \cdot 100\% \approx 28\%$. Коэффициент вариации 28% относительно невысок, что говорит о небольшом разбросе значений признака относительно некоторой средней величины.

Асимметрия $A_s = \dfrac{\mu_3}{s^3}$, где третий центральный момент

$$\mu_3 = a_3 - 3a_1 a_2 + 2a_1^3 = 4598,40 - 3 \cdot 15,50 \cdot 258,62 + 2 \cdot 15,50^3 \approx 17,230$$

$$A_s = \frac{17,230}{4,303^3} \approx 0,216$$

Экцесс $E_x = \dfrac{\mu_4}{s^4} - 3$, где четвёртый момент

$$\mu_4 = a_4 - 4a_3 a_1 + 6a_2 a_1^2 - 3a_1^4 = 86331,12 - 4 \cdot 4598,4 \cdot 15,50^2 + 6 \cdot 258,62 \cdot 15,50^2 - 3 \cdot 15,50^4 \approx 938,238$$

$$E_x = \frac{938,238}{4,303^4} - 3 \approx -0,263$$

Третий центральный момент и коэффициент асимметрии служат для характеристики асимметрии (скошенности) распределения. В нашем случае асимметрия близка к нулю, что может быть характерно для нормального или равномерного распределения, когда распределение симметрично относительно математического ожидания. (например, для показательного распределения асимметрия может быть около 2) .

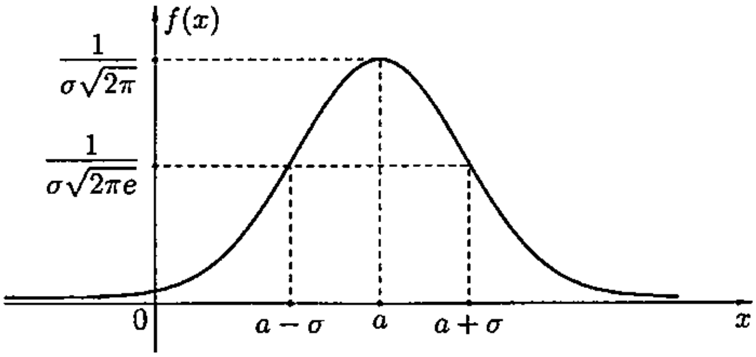
Четвёртый центральный момент характеризует крутость (островершинность или плосковершинность) распределения. Кривые, более островершинные, чем нормальная, обладают положительным экцессом, более плосковершинные - отрицательным экцессом. В нашем случае экцесс распределения отрицателен, что говорит о плосковершинности кривой распределения, что в совокупности с малой асимметрией может быть характерно и для равномерного распределения.

Для нормального распределения и асимметрия и экцесс близки к нулю.

3

Исходя из вида гистограммы и значений вычисленных статистик (экцесс и асимметрия малы; хотя коэффициент вариации небольшой 28%, но отклонения вариант согласованные - по краям ряда значения малы, в середине - побольше), выбираем в качестве теоретического распределения нормальное (с параметрами m и σ). Т.е. принимаем, что неизвестная плотность теоретического распределения имеет вид:

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



Теперь возникает задача о наилучшем выборе параметров распределения (в данном случае это параметры m и σ). Для решения этой задачи будем руководствоваться методом моментов, который предполагает, что наилучшими значениями параметров распределения являются те, для которых теоретические значения первых двух моментов распределения совпадают с выборочными (статистиками распределения).

В случае нормального распределения параметры m и σ равны соответственно выборочному математическому ожиданию и выборочной дисперсии:

$$m=\bar{x}=15{,}50$$

$$\sigma=s=4{,}303$$

Итак, в качестве теоретической (сглаживающей) кривой распределения возьмём график функции

$$f(x)=\frac{1}{4{,}303\cdot\sqrt{2\pi}}\cdot e^{-\frac{(x-15{,}50)^2}{2\cdot 4{,}303^2}}=0{,}09271\cdot e^{-\frac{(x-15{,}5)^2}{37{,}032}}$$

4

Для построения этой кривой, а также для проведения дальнейшей статистической обработки составим следующую таблицу:

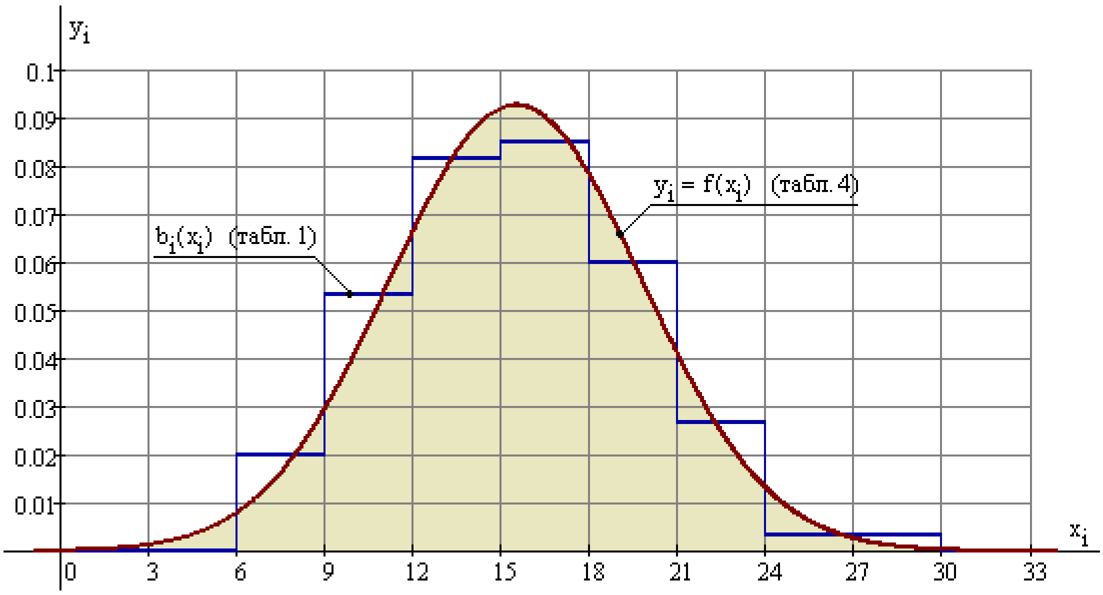
таблица 4

1	2	3	4	5	6	7	8	9	10
x_i	m_i	$x_i-\bar{x}$	$z_i=\frac{x_i-\bar{x}}{s}$	$\varphi(z_i)$	$y_i=f(x_i)=\varphi(z_i)/s$	$p_i=h\cdot y_i$	$n\cdot p_i$	$m_i-n\cdot p_i$	$\frac{(m_i-np_i)^2}{np_i}$
7,5	6	-8	-1,859	0,0709	0,01648	0,04944	4,944	1,056	0,226
10,5	16	-5	-1,162	0,2031	0,04720	0,14160	14,160	1,840	0,239
13,5	24,5	-2	-0,465	0,3581	0,08322	0,24966	24,966	-0,466	0,009
16,5	25,5	1	0,232	0,3883	0,09024	0,27072	27,072	-1,572	0,091
19,5	18	4	0,930	0,2589	0,06017	0,18051	18,051	-0,051	0,000
22,5	8	7	1,627	0,1062	0,02468	0,07404	7,404	0,596	0,048
27	2	11,5	2,673	0,0112	0,00265	0,01590	1,590	0,410	0,106
									0,719

Здесь $\varphi(z_i)$ вычисляется по формуле плотности стандартного нормального распределения

$$\varphi(z)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{z^2}{2}}$$ (можно использовать таблицы значений этой функции).

Кривая $y_i = f(x_i)$ изображена далее в Mathcad 14 для наглядности на одном чертеже с гистограммой распределения и хорошо согласуется с данными выборки (пока это только визуальная оценка).



5

Оценим согласованность выбранного теоретического распределения с опытными данными в соответствии с критерием Пирсона (критерий χ^2 [хи квадрат]). Критерий согласия Пирсона - наиболее часто употребляемый критерий для проверки простой гипотезы о законе распределения (гипотезу, однозначно фиксирующую распределение наблюдений, называют простой (в ней речь идёт об одном значении параметра), в противном случае - сложной). Суммируя величины колонки 10 в таблице 3, получаем величину $\chi^2_{\text{р}}=\sum_{i=1}^{9}\frac{(m_i-np_i)^2}{np_i}\approx 0{,}72$, характеризующую меру расхождения теоретического и статистического распределений. Чем больше величина χ^2, тем больше это расхождение.

Зададимся некоторым критическим значением уровня значимости $\alpha = P\left(\chi^2 \ge \chi^2_{\text{р}}\right)$, например $\alpha_{\text{кр.}} = P\left(\chi^2 \ge \chi^2_{\text{р}}\right)=0{,}05$, который должен быть превышен для принятия гипотезы о законе распределения. Отметим, что принятие гипотезы всегда происходит на некотором субъективно принятом уровне значимости и основывается на значениях конечной выборки (в данном случае $n=100$). Обычно ориентируются на критическое значение уровня значимости $\alpha = 0{,}01\ ...\ 0{,}05$ ($\alpha = 0{,}01$ соответствует более мягкому подходу в оценке гипотезы, $\alpha = 0{,}05$ соответствует более строгому подходу в оценке гипотезы).

По таблице критических точек χ^2 - распределения исходя из числа степеней свободы $r = k-3 = 7-3 = 4$ (k - число разрядов) и реализовавшегося значения $\chi^2_{\text{р}}=0{,}72$ находим величину $\alpha = P\left(\chi^2 \ge \chi^2_{\text{р}}\right) = P\left(\chi^2 \ge 0{,}72\right)\approx 0{,}9$, т.е. вероятность того, что величина, распределённая по закону χ^2, превысит значение $\chi^2_{\text{р}}$.

Таблицы довольно грубы; проще и быстрее значение α вычисляется, например в Mathcad 14:

$$\varphi(x,r) := \frac{1}{2^{\frac{r}{2}} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}} \qquad \alpha := \int_{0.72}^{\infty} \varphi(x,4)\, dx = 0.949$$

Поскольку вероятность $\alpha > \alpha_{\text{кр.}}$ $(0{,}9 > 0{,}05)$, то можно считать, что эмпирически принятое теоретическое нормальное распределение не противоречит опытным данным и гипотеза о виде распределения и о его параметрах может быть принята. Другими словами, принятая гипотеза не противоречит имеющимся выборочным данным на уровне значимости $0{,}05$, или уровне надёжности (уровне доверия) $1-0{,}05=0{,}95$.

Литература:

1) Белугин В.И., Величко Т.В., Поповский Э.Е. "Высшая математика", часть 4, методичка УрГУПС, 2001, стр. 29;
2) Ниворожкина Л.И., Морозова З.А. "Математическая статистика с элементами теории вероятностей в задачах с решениями", 2005, стр. 284...287, стр. 585;
3) Письменный Д.Т. "Конспект лекций по теории вероятностей и математической статистике", 2004, стр. 212, 216;
4) Кремер Н.Ш. "Теория вероятностей и математическая статистика", 2006, стр. 173 (формула плотности вероятности χ^2-распределения);
5) Белько И.В., Свирид Г.П. "Теория вероятностей и математическая статистика", 2004, стр. 146, 155, 239;
6) Бирюкова Л.Г. и др. "Теория вероятностей и математическая статистика", 2004, стр. 86 (числа степеней свободы).

► Распределение величины χ^2 (хи-квадрат) при большом количестве наблюдений n практически не зависит от числа наблюдений и от теоретической функции распределения, а зависит лишь от числа степеней свободы. Число степеней свободы r определяется равенством $r = k - 3$, где k - число разрядов. Вычитание числа 3 объясняется для нормального и равномерного распределений наличием трёх соотношений, связывающих частоты:

$$\sum_{i=1}^{k} p_i^* = 1, \quad \sum_{i=1}^{k} x_i \cdot p_i^* = m, \quad \sum_{i=1}^{k} \left(x_i - \bar{x}\right)^2 \cdot p_i^* = \sigma^2 .$$

В случае показательного распределения $r = k - 2$, т.к. последнее соотношение при расчёте χ^2 не используется (см. табл. 3).

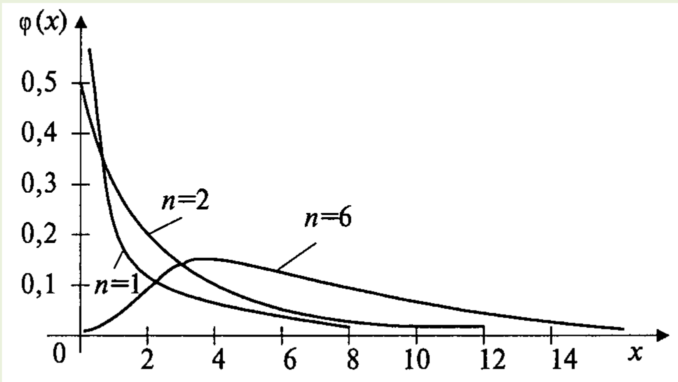
► Плотность вероятности χ^2-распределения имеет вид

$$\varphi(x) = \begin{cases} \dfrac{1}{2^{\frac{r}{2}} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}} & \text{при } x \geq 0, \\ 0 & \text{при } x < 0 \end{cases} \qquad (*)$$

где $\Gamma(y) = \int\limits_0^{+\infty} e^{-t} \cdot t^{y-1} dt$ - гамма-функция Эйлера (для целых положительных значений $\Gamma(y) = (y-1)!$),

r - число степеней свободы.

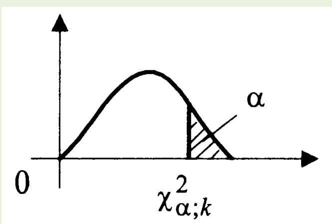
Кривые χ^2-распределения для различных значений числа степеней свободы r приведены на рисунке:



Они показывают, что χ^2-распределение асимметрично, обладает положительной (правосторонней) асимметрией.

При $r > 30$ распределение случайной величины $Z = \sqrt{2\chi^2} - \sqrt{2r-1}$ близко к стандартному нормальному закону, т.е. $N(0;1)$.

Вычислив величину $\chi_{\text{p}}^2 = 17,669$, уровень значимости α можно вычислить по таблице критических точек χ^2 - распределения (как сделано в разделе 4) или посредством формулы (*) в любом калькуляторе, например Mathcad:



$$\phi(x,r) := \frac{1}{2^{\frac{r}{2}} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}}$$

$$\alpha := \int_{17.669}^{\infty} \phi(x,8)\, dx \rightarrow 0.02385$$

В Matcad гамма-функция доступна через клики Insert→Function→Special→Gamma или печатается двойным нажатием клавиш: сначала G, а затем [Ctrl]+G.

Обратное вычисление можно сделать с помощью встроенной в Mathcad функции, например:

$$\mathrm{qchisq}(1 - 0.01, 2) = 9.210$$

(таким образом вычислена приведённая далее таблица критических точек распределения χ^2)

► Критические точки распределения χ^2 .

число степеней свободы k	уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,635	5,024	3,841	0,00393	0,00098	0,00016
2	9,210	7,378	5,991	0,103	0,051	0,020
3	11,345	9,348	7,815	0,352	0,216	0,115
4	13,277	11,143	9,488	0,711	0,484	0,297
5	15,086	12,833	11,070	1,145	0,831	0,554
6	16,812	14,449	12,592	1,635	1,237	0,872
7	18,475	16,013	14,067	2,167	1,690	1,239
8	20,090	17,535	15,507	2,733	2,180	1,647
9	21,666	19,023	16,919	3,325	2,700	2,088
10	23,209	20,483	18,307	3,940	3,247	2,558
11	24,725	21,920	19,675	4,575	3,816	3,053
12	26,217	23,337	21,026	5,226	4,404	3,571
13	27,688	24,735	22,362	5,892	5,009	4,107
14	29,141	26,119	23,685	6,571	5,629	4,660
15	30,578	27,488	24,996	7,261	6,262	5,229
16	32,000	28,845	26,296	7,962	6,908	5,812
17	33,409	30,191	27,587	8,672	7,564	6,408
18	34,805	31,526	28,869	9,390	8,231	7,015
19	36,191	32,852	30,144	10,117	8,907	7,633
20	37,566	34,170	31,410	10,851	9,591	8,260
21	38,932	35,479	32,671	11,591	10,283	8,897
22	40,289	36,781	33,924	12,338	10,982	9,542
23	41,638	38,076	35,172	13,091	11,689	10,196
24	42,980	39,364	36,415	13,848	12,401	10,856
25	44,314	40,646	37,652	14,611	13,120	11,524
26	45,642	41,923	38,885	15,379	13,844	12,198
27	46,963	43,195	40,113	16,151	14,573	12,879
28	48,278	44,461	41,337	16,928	15,309	13,565
29	49,588	45,722	42,557	17,708	16,047	14,256
30	50,892	46,979	43,773	18,493	16,791	14,953
40	63,691	59,342	55,758	26,509	24,433	22,164
50	76,154	71,421	67,505	34,764	32,357	29,707
60	88,379	83,297	79,082	43,188	40,482	37,485
80	112,329	106,629	101,879	60,391	57,153	53,540
100	135,807	129,561	124,342	77,929	74,221	70,064
120	158,950	152,211	146,567	95,705	91,572	86,923

таблицы см. также http://www.statsoft.ru/home/textbook/modules/sttable.html

Литература:

1) Ивановский Р.И. "Теория вероятностей и математическая статистика. Основы, прикладные аспекты с примерами и задачами в среде Mathcad", 2008, стр. 332;

2) http://www.twirpx.com/file/187799/ - Гурский Д.А., Турбина Е.С. "Вычисления в Mathcad 12", полная версия книги (существует только в электронном виде), 2006, стр. 561 (распределение хи-квадрат).

14) Нарисовать диаграммы функции распределения:

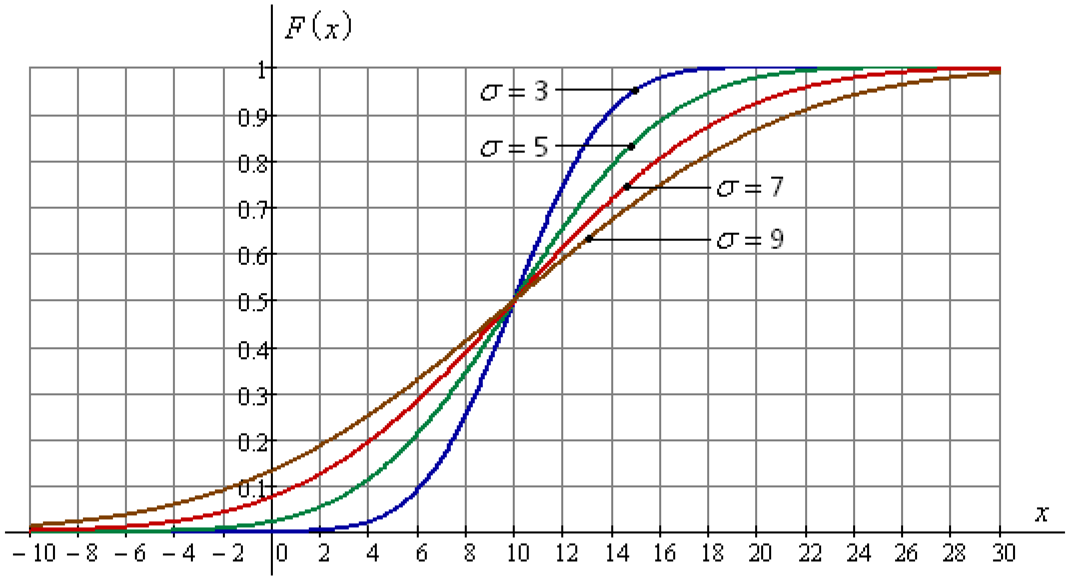
1. нормального распределения, $M(X) = 10, \ \sigma(X) = \{3, 5, 7, 9\}$;
2. показательного распределения, $M(X) = \{4, 6, 8, 10\}$;
3. распределения Пуассона, $M(X) = \{4, 6, 8, 10\}$.

1.

Функция распределения с.в., распределённой по нормальному закону, выражается через функцию Лапласа по формуле:

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right), \text{ где } \Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

Выполним иллюстрацию в Mathcad 15:



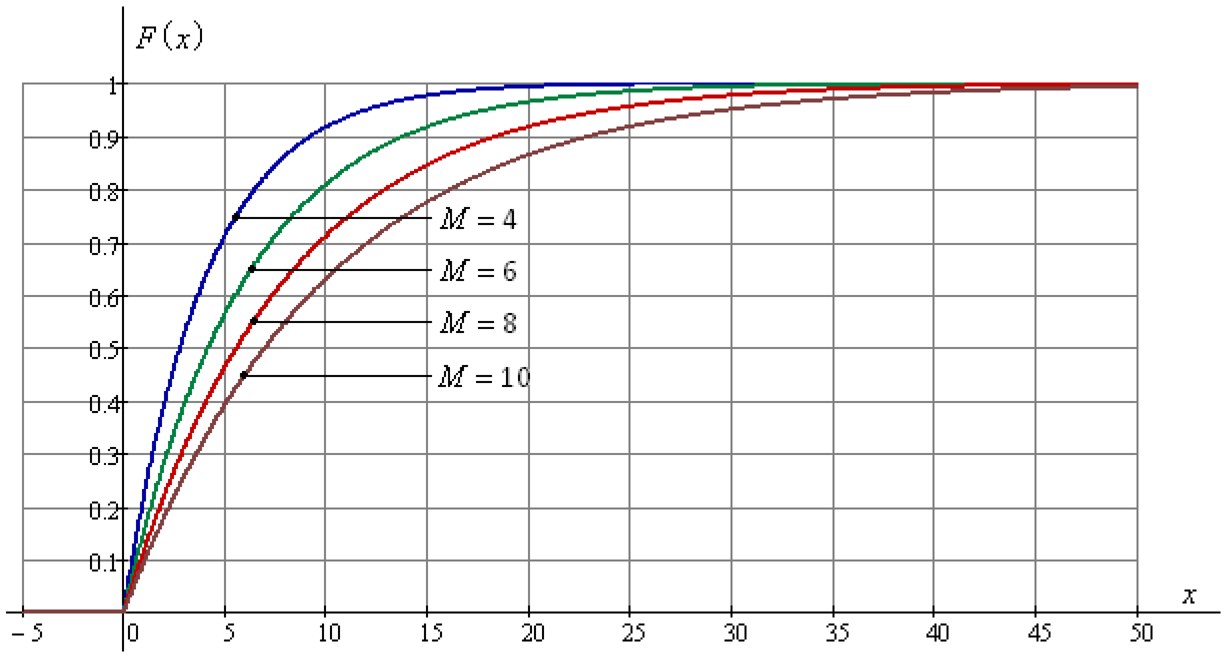
2.

Функция распределения с.в., распределённой по показательному (экспоненциальному) закону, определяется формулой

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

причём $M(X) = \frac{1}{\lambda}$, $D(X) = \frac{1}{\lambda^2}$.

Выполним иллюстрацию в Mathcad 15:



3.

Распределением Пуассона называется распределение вероятностей дискретной случайной величины, определяемое формулой Пуассона

$P_\lambda(k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$, причём параметр распределения Пуассона $\lambda = a = \sigma^2$ равен математическому ожиданию идисперсии этого закона распределения.

(вероятность появления k событий простейшего потока за некоторый промежуток времени t при заданной интенсивности λ/t)

Функция распределения Пуассона имеет вид

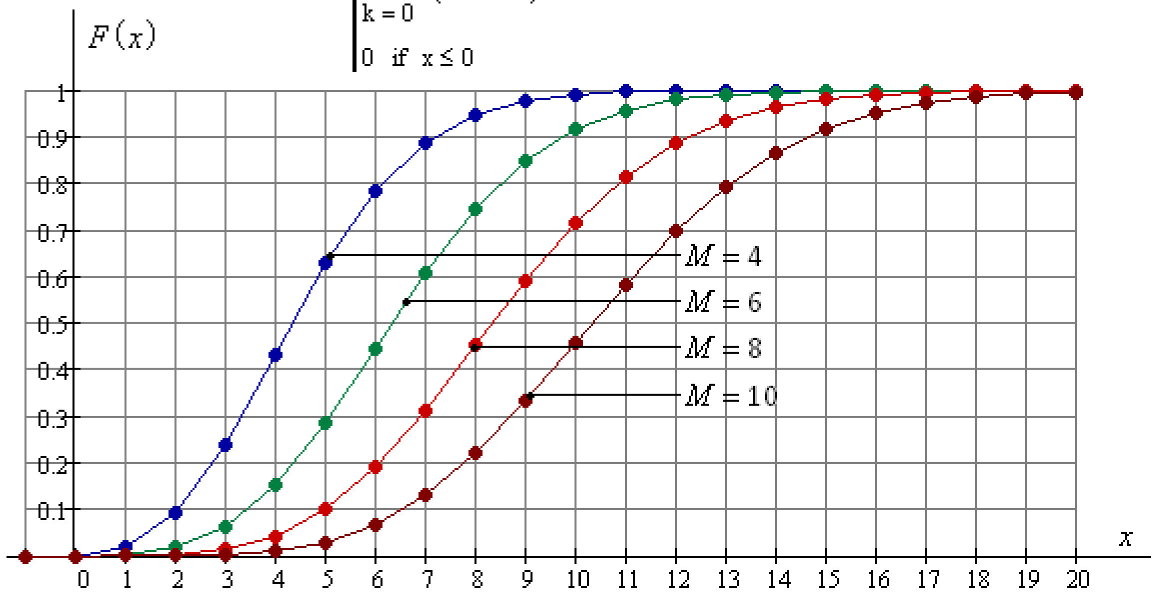
$$F(x) = \begin{cases} \sum\limits_{0 \le k < x} \dfrac{\lambda^k}{k!} e^{-\lambda} & при\ \ x > 0 \\ 0 & при\ \ x \le 0 \end{cases}$$

причём

$$M(X) = \lambda .$$

Выполним иллюстрацию в Mathcad 15:

$$F(x,\lambda) := \begin{vmatrix} \sum\limits_{k=0}^{x-1} \left(\dfrac{\lambda^k}{k!} \cdot e^{-\lambda} \right) & \text{if } x > 0 \\ 0 \ \text{ if } x \le 0 & \end{vmatrix} \qquad x := 0, 1 .. 20$$



Литература:

1) Спирина М.С., Спирин П.А. "Теория вероятностей и математическая статистика", 2007, стр. 127 (распределение Пуассона);

2) Шириков В.Ф., Зарбалиев С.М. "Теория вероятностей", 2008, стр. 198 (распределение Пуассона).

15) Нарисовать диаграммы плотности распределения (каждую на одном графике, обязательно создать графическую иллюстрацию):

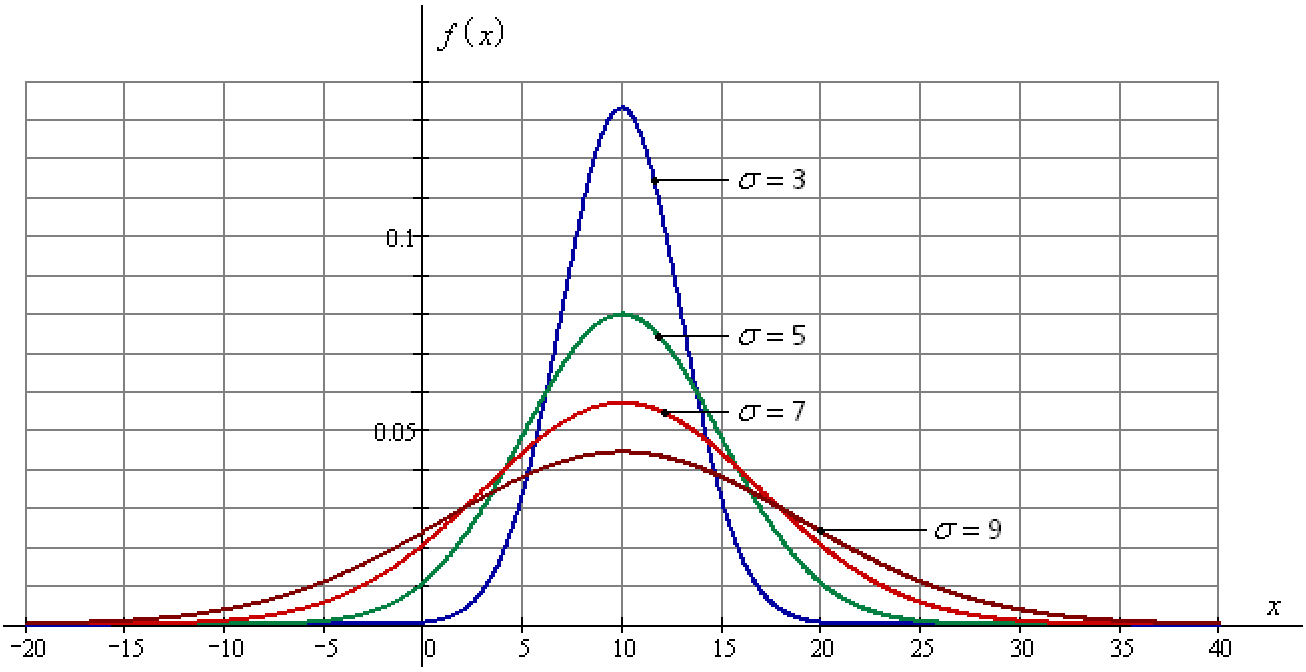
1. нормального распределения, $M(X) = 10$, $\sigma(X) = \{3, 5, 7, 9\}$;
2. показательного распределения, $M(X) = \{4, 6, 8, 10\}$;
3. распределения Пуассона, $M(X) = \{4, 6, 8, 10\}$.

1.

Непрерывная с.в. имеет нормальный закон распределения (закон Гаусса) с параметрами a и σ^2, если её плотность вероятности имеет вид

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

Выполним иллюстрацию в Mathcad 15:



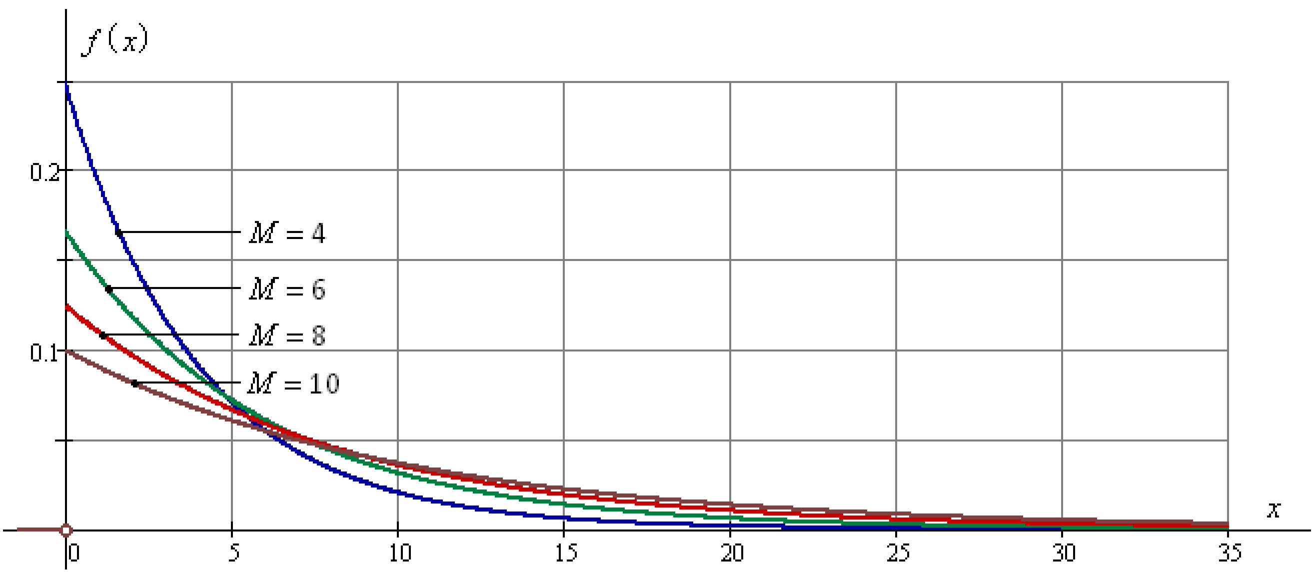
2.

Непрерывная с.в. имеет показательный (экспоненциальный) закон распределения с параметром $\lambda>0$, если её плотность вероятности имеет вид

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

причём $M(X)=\frac{1}{\lambda}$, $D(X)=\frac{1}{\lambda^2}$.

Выполним иллюстрацию в Mathcad 15:



3.
Поскольку плотность вероятности есть функция дифференциальная, а вероятность для распределения Пуассона вычисляется по формуле

$$P(x)=\frac{\lambda^x}{x!}e^{-\lambda}, \quad n\in N_0$$

то для дискретной с.в. плотность вероятности

$$f(x)=\frac{\lambda^x}{x!}e^{-\lambda}-\frac{\lambda^{x-1}}{(x-1)!}e^{-\lambda}=\frac{\lambda^{x-1}}{(x-1)!}e^{-\lambda}\cdot\left(\frac{\lambda}{x}-1\right)$$

причём

$$M(X)=\lambda.$$

Выполним иллюстрацию в Mathcad 15: