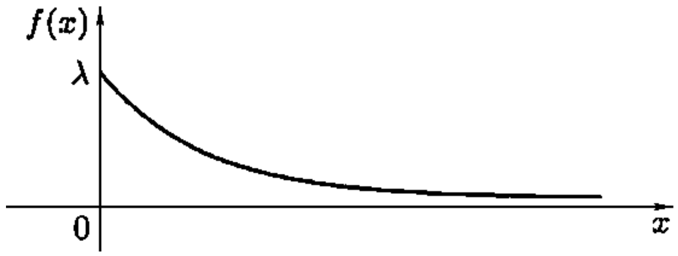
Показательное распределение.

1) Распределение с.в. X подчинено показательному закону с параметром $\lambda = 5,2$. Записать $f(x)$, вычислить $M(X),\ D(X)$.



Показательное распределение с параметром λ имеет плотность вероятности:

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

и функция распределения

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

Математическое ожидание $M(X)$ и дисперсия $D(X)$:

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

$$f(x) = \begin{cases} 0 & ,\ если\ x < 0 \\ 5,2 \cdot e^{-5,2 \cdot x} & ,\ если\ x \geq 0 \end{cases}$$

$$M(X) = \frac{1}{5,2} = \frac{5}{26} \approx 0,192$$

$$D(X) = \frac{1}{5,2^2} = \frac{25}{676} \approx 0,037$$

2) Длительность времени безотказной работы элемента имеет показательное распределение с параметром $\lambda = 0,02$. Найти вероятность того, что за время длительностью $t = 100$ ч:

а) элемент откажет;
б) элемент не откажет.

Обозначим событие:

A - элемент откажет за время работы 100 ч.

В данном случае плотность вероятности:

$$f(t) = \begin{cases} 0 & ,\ если\ t < 0 \\ 0,02 \cdot e^{-0,02 \cdot t} & ,\ если\ t \geq 0 \end{cases}$$

а функция распределения имеет вид

$$F(t) = \begin{cases} 0 & ,\ если\ t < 0 \\ 1 - e^{-0,02 \cdot t} & ,\ если\ t \geq 0 \end{cases}$$

Вероятность того, что элемент откажет в течение времени $t=100$ ч - это вероятность отказа элемента на промежутке $0\le t\le 100$ ч:

$$P(A)=P(0\le t\le 100)=F(100)-F(0)=\left(1-e^{-0{,}02\cdot 100}\right)-\left(1-e^{-0{,}02\cdot 0}\right)=$$
$$=\left(1-e^{-2}\right)-\left(1-e^{0}\right)=1-e^{-2}\approx 0{,}865$$

События "элемент откажет" и "элемент не откажет" - противоположные, поэтому вероятность того, что элемент не откажет на промежутке времени $0\le t\le 100$ ч:

$$P(\bar{A})=1-P(A)=1-\left(1-e^{-2}\right)=e^{-2}\approx 0{,}135$$

3) 98% топливных насосов дизельных тракторов выходят из строя после 3000 моточасов. Какова вероятность, что насос выйдет из строя в интервале времени от 2000 до 2500 моточасов. Оценить среднее время безотказной работы таких машин и дисперсию.

► Полагаем непрерывную случайную величину "время безотказной работы машины" распределённой по экспоненциальному (показательное) закону.

Плотность вероятности при показательном распределении:

$$f(x)=\begin{cases}0\,, & x<0\\ \lambda\cdot e^{-\lambda x}, & x\ge 0\end{cases}$$

Найдём параметр λ:

$$P(t\ge 3000)=1-P(t<3000)=1-F(3000)=1-\int_0^{3000}\lambda\cdot e^{-\lambda x}dx=1-\left(1-e^{-\lambda\cdot 3000}\right)=e^{-3000\cdot\lambda}=0{,}98$$

$$-3000\cdot\lambda\cdot\ln e=\ln 0{,}98$$

$$\lambda=\frac{\ln 0{,}98}{-3000}\approx 0{,}0000067$$

Следовательно

$$f(x)=\begin{cases}0\,, & x<0\\ 0{,}0000067\cdot e^{-0{,}0000067x}, & x\ge 0\end{cases}$$

► Найдём вероятность того, что насос выйдет из строя в интервале времени от 2000 до 2500 моточасов.

$$P(2000<X<2500)=\int_{2000}^{2500}f(x)\,dx=\int_{2000}^{2500}0{,}0000067\cdot e^{-0{,}0000067x}dx=$$
$$=-e^{-0{,}0000067x}\Big|_{2000}^{2500}=-\left(e^{-0{,}01675}-e^{-0{,}0134}\right)\approx 0{,}0033$$

► Оценим среднее время безотказной работы машин (аналог математического ожидания) и дисперсию, используя готовые формулы для показательного распределения.

Для показательного распределения математическое ожидание $M(X)=1/\lambda$; т.е.

$$M(X)=\frac{1}{0{,}0000067}\approx 150000 \text{ (моточасов)}\approx 17 \text{ (лет)}$$

Дисперсия:

$$D(X)=\frac{1}{\lambda^2}=\left(\frac{1}{0{,}0000067}\right)^2\approx 2\cdot 10^{10} \text{ (моточасов}^2\text{)}$$

4) По данным страховых агенств некоторой страны вероятность того, что человек доживёт до 70 лет, равна 0,32. Какова вероятность того, что случайный новорожденный доживёт до свадьбы (до 22 лет) ? Оценить среднее время жизни в данной стране и среднее отклонение от него.

Продолжительность жизни человека подобна длительности работы механизма - до первого отказа. Поэтому используем показательное распределение.

Плотность вероятности при показательном распределении:

$$f(x)=\begin{cases} 0\,, & x\le 0 \\ \lambda\cdot e^{-\lambda x}, & x>0 \end{cases}$$

Найдём параметр λ (вероятность того, что человек доживёт до 70 лет, равна 0,32):

$$P(t\ge 70)=1-P(t<70)=1-F(70)=1-\int_0^{70}\lambda\cdot e^{-\lambda x}dx=1-\left(1-e^{-\lambda\cdot 70}\right)=e^{-\lambda\cdot 70}=0,32$$

$$-70\cdot\lambda\cdot\ln e=\ln 0,32$$

$$\lambda=\frac{\ln 0,32}{-70}\approx 0,0163$$

Вероятность дожить до 22 лет:

$$P(t\ge 22)=e^{-0,0163\cdot 22}\approx 0,70$$

Для показательного распределения математическое ожидание $M(X)=1/\lambda$; т.е. в данной стране среднее время жизни получается

$$M(X)=\frac{1}{0,0163}\approx 61 \text{ (год)}$$

Дисперсия с.в. X и среднеквадратическое отклонение:

$$D(X)=\frac{1}{\lambda^2}\ ;\quad \sigma(X)=\sqrt{D(X)}=\frac{1}{\lambda}=\frac{1}{0,0163}\approx 61 \text{ (год)}$$

- среднее квадратическое отклонение от $M(X)$.

Выборочный метод.

5) 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 ($n=100$):

18,4	0,8	4,9	10,5	8,8	24,9	16,0	6,0	4,6	16,2
7,6	16,1	13,1	3,0	23,6	4,7	1,8	5,6	6,2	11,4
0,3	9,2	2,6	11,5	2,6	18,9	3,8	4,9	22,4	7,9
12,2	2,0	2,5	5,2	11,2	15,4	13,6	2,4	10,5	3,4
0,7	3,6	2,2	10,3	63,7	13,9	7,5	35,0	28,6	0,2
7,7	4,0	3,6	10,4	15,6	48,3	7,5	8,0	0,9	20,3
18,2	2,8	0,8	17,4	4,0	0,9	1,9	22,6	3,4	18,1
15,8	3,6	2,7	21,5	30,2	3,9	19,2	9,3	21,8	25,1
0,0	33,6	2,1	7,4	4,6	37,0	10,4	15,0	2,6	5,0
6,6	16,0	6,1	9,9	7,6	1,8	11,2	3,8	21,3	0,5

1

Выбираем из этой совокупности наименьшее и наибольшее значения:

$$x_{наим.} = 0$$

$$x_{наиб.} = 63{,}7$$

Округляем граничные значения до 0 и 65. 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 $k = 13$, и их длина h постоянна; тогда

$$h = \frac{65 - 0}{13} = 5 .$$

Составим интервальный статистический ряд (табл. 1).

таблица 1

Номер i	1	2	3	4	5	6	7	8	9	10	11	12	13
i-й разряд	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65
Число наблюдений m_i	38,5	18,5	13,5	13,5	8	2	2,5	1,5	0	1	0	0	1
Частота p_i^*	0,385	0,185	0,135	0,135	0,08	0,02	0,025	0,015	0	0,01	0	0	0,01
Частость $b_i^* = p_i^* / h$	0,077	0,037	0,027	0,027	0,016	0,004	0,005	0,003	0	0,002	0	0	0,002

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

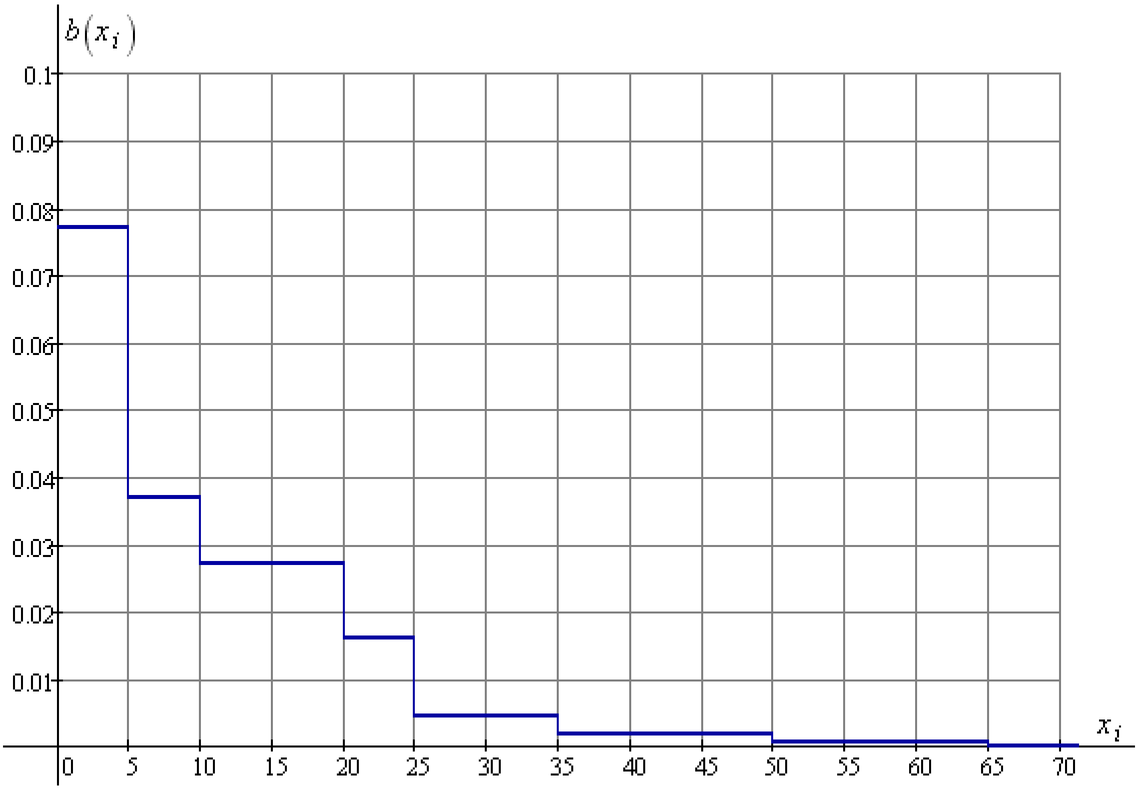
В последних разрядах оказалось слишком мало наблюдений ($m_i < 3$), поэтому объединим их:

таблица 2

Номер i	1	2	3	4	5	6	7	8
i-й разряд	0-5	5-10	10-15	15-20	20-25	25-35	35-50	50-65
Число наблюдений m_i	38,5	18,5	13,5	13,5	8	4,5	2,5	1
Частота p_i^*	0,385	0,185	0,135	0,135	0,08	0,045	0,025	0,01
Частость b_i^*	0,077	0,037	0,027	0,027	0,016	0,0045	0,001(6)	0,000(6)

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

Используя данные таблицы 2, построим гистограмму распределения (зависимость $b_i(x_i)$) в Mathcad 14.



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	8
Середина интервала x_i	2,5	7,5	12,5	17,5	22,5	30	42,5	57,5
Число наблюдений m_i	38,5	18,5	13,5	13,5	8	4,5	2,5	1
Частота p_i^*	0,385	0,185	0,135	0,135	0,08	0,045	0,025	0,01

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:

$$a_1 = 11{,}1875$$

$$a_2 = 234{,}4688$$

$$a_3 = 7017{,}734$$

$$a_4 = 264387{,}3$$

Вычисления в Mathcad 14:

$$B := \begin{pmatrix} 2.5 \\ 7.5 \\ 12.5 \\ 17.5 \\ 22.5 \\ 30 \\ 42.5 \\ 57.5 \end{pmatrix} \qquad C := \begin{pmatrix} 0.385 \\ 0.185 \\ 0.135 \\ 0.135 \\ 0.08 \\ 0.045 \\ 0.025 \\ 0.01 \end{pmatrix}$$

$$a_1 := B \cdot C \qquad a_1 = 11.1875$$

$$a_2 := \overrightarrow{B^2} \cdot C \qquad a_2 = 234.46875$$

$$a_3 := \overrightarrow{B^3} \cdot C \qquad a_3 = 7017.73438$$

$$a_4 := \overrightarrow{B^4} \cdot C \qquad a_4 = 264387.30469$$

$$a_2 - a_1^2 = 109.30859 \qquad \sqrt{a_2 - a_1^2} = 10.455$$

$$a_3 - 3 \cdot a_1 \cdot a_2 + 2 \cdot a_1^3 = 1948.83545$$

$$a_4 - 4 \cdot a_3 \cdot a_1 + 6 \cdot a_2 \cdot a_1^2 - 3 \cdot a_1^4 = 79425.36958$$

Определяем статистики выборочного распределения.

Выборочная средняя $\overline{x} = a_1 = 11{,}1875$. Выборочная средняя - статистический аналог математического ожидания.

Выборочная дисперсия $s^2 = a_2 - a_1^2 = 234{,}4688 - 11{,}1875^2 = 109{,}3086$. Выборочная дисперсия - аналог дисперсии; характеризует рассеяние случайной величины вокруг её среднего значения.

Выборочное среднеквадратическое отклонение $s = \sqrt{s^2} = \sqrt{109{,}3086} \approx 10{,}455$.

Коэффициент вариации $V = \dfrac{s}{\overline{x}} \cdot 100\% = \dfrac{10{,}455}{11{,}1875} \cdot 100\% \approx 93\%$. Коэффициент вариации 93% очень большой, что говорит о большом разбросе значений признака относительно некоторой средней величины.

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

$$\mu_3 = a_3 - 3a_1a_2 + 2a_1^3 = 7017{,}734 - 3 \cdot 11{,}1875 \cdot 234{,}4688 + 2 \cdot 11{,}1875^3 \approx 1948{,}835$$

$$A_s = \frac{1948{,}835}{10{,}455^3} \approx 1{,}705$$

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

$$\mu_4 = a_4 - 4a_3a_1 + 6a_2a_1^2 - 3a_1^4 =$$

$$= 264387{,}3 - 4 \cdot 7017{,}734 \cdot 11{,}1875 + 6 \cdot 234{,}4688 \cdot 11{,}1875^2 - 3 \cdot 11{,}1875^4 \approx 79425{,}37$$

$$E_x = \frac{79425{,}37}{10{,}455^4} - 3 \approx 3{,}648$$

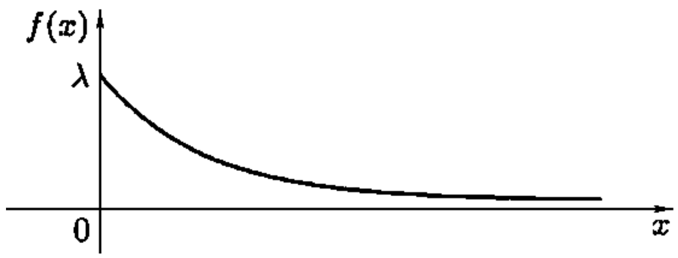
Третий центральный момент и коэффициент асимметрии служат для характеристики асимметрии (скошенности) распределения. В нашем случае асимметрия близка к двум, что характерно для показательного распределения (для нормального или равномерного распределения, когда распределение симметрично относительно математического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

Четвёртый центральный момент характеризует крутость (островершинность или плосковершинность) распределения. Кривые, более островершинные, чем нормальная, обладают положительным эксцессом, более плосковершинные - отрицательным эксцессом. В нашем случае эксцесс распределения положителен, что говорит об островершинности кривой распределения, что в совокупности с большой асимметрией указывает на показательное распределение (для нормального распределения и асимметрия и эксцесс близки к нулю).

3

Исходя из вида гистограммы и значений вычисленных статистик (экцесс и асимметрия велики, коэффициент вариации близок к 100%), выбираем в качестве теоретического распределения показательное (с параметром λ). Т.е. принимаем, что неизвестная плотность теоретического распределения имеет вид:

$$f(x)=\begin{cases} 0 & \text{если } x<0 \\ \lambda\cdot e^{-\lambda x} & \text{если } x\ge 0 \end{cases}$$



Теперь возникает задача о наилучшем выборе параметров распределения (в данном случае это параметр λ). 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).

Для показательного распределения параметр λ на основании метода моментов вычисляем по формуле

$$\lambda=\frac{1}{\overline{x}}=\frac{1}{11,1875}=\frac{16}{179}\approx 0,089385\,.$$

Через параметр λ для показательного распределения легко вычислить математическое ожидание

$M(X)=\dfrac{1}{\lambda}=11,1875$ и дисперсию $D(X)=\dfrac{1}{\lambda^2}=\left(\dfrac{179}{16}\right)^2\approx 125,16\,.$

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

$$f(x)=\begin{cases} 0 & \text{если } x<0 \\ \dfrac{16}{179}\cdot e^{-\frac{16\cdot x}{179}} & \text{если } x\ge 0 \end{cases}$$

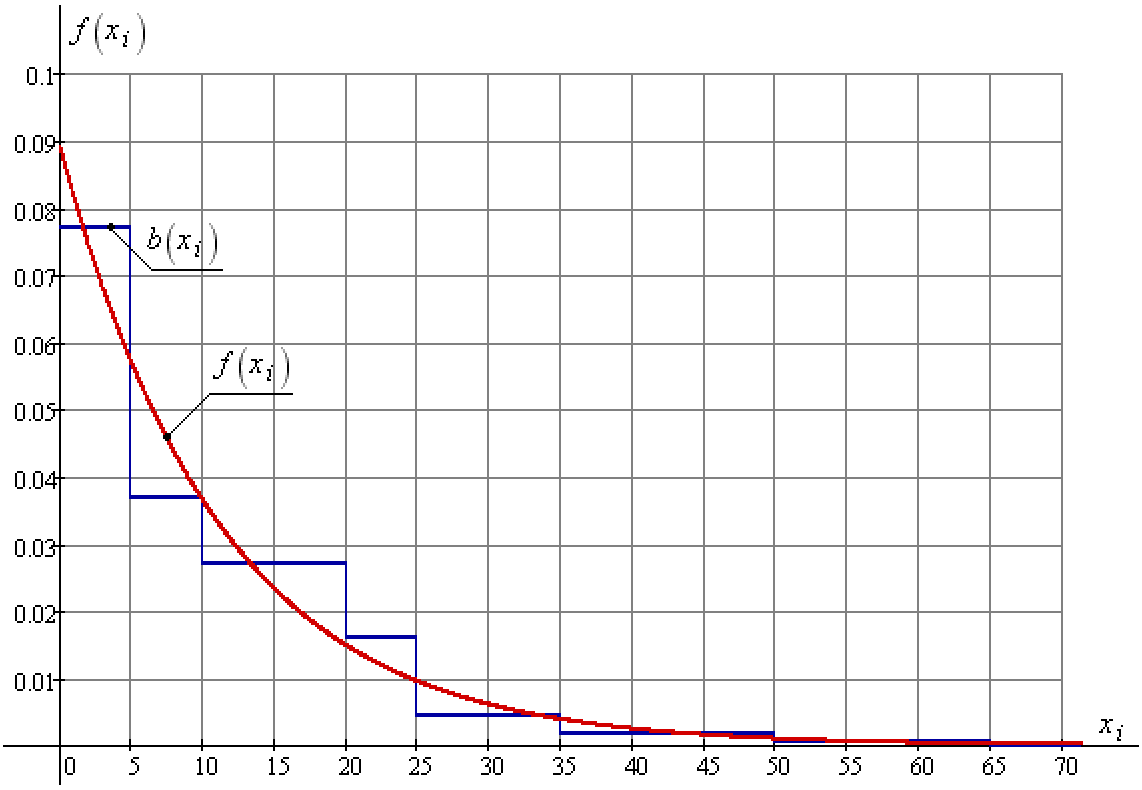
4

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
x_i	m_i	$\lambda\cdot x_i$	$e^{-\lambda x_i}$	$y_i=f(x_i)= =\lambda\cdot e^{-\lambda\cdot x_i}$	$p_i=h\cdot y_i$	$n\cdot p_i$	$m_i-n\cdot p_i$	$\dfrac{(m_i-n\cdot p_i)^2}{n\cdot p_i}$
2,5	38,5	0,223	0,800	0,07152	0,3576	35,76	2,74	0,210
7,5	18,5	0,670	0,512	0,04572	0,2286	22,86	-4,36	0,832
12,5	13,5	1,117	0,327	0,02924	0,1462	14,62	-1,12	0,086
17,5	13,5	1,564	0,209	0,01870	0,0935	9,35	4,15	1,842
22,5	8	2,011	0,134	0,01196	0,0598	5,98	2,02	0,682
30	4,5	2,682	0,068	0,00612	0,0612	6,12	-1,62	0,429
42,5	2,5	3,799	0,022	0,00200	0,0300	3,00	-0,50	0,083
57,5	1	5,140	0,006	0,00052	0,0079	0,79	0,21	0,056
								4,220

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 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 $\chi^2_{\text{р}} = \sum_{i=1}^{9} \frac{(m_i - np_i)^2}{np_i} \approx 4,22$, характеризующую меру расхождения теоретического и статистического распределений. Чем больше величина χ^2, тем больше это расхождение.

Зададимся некоторым критическим значением уровня значимости $\alpha = P\left(\chi^2 \geq \chi^2_{\text{р}}\right)$, например $\alpha_{\text{кр.}} = P\left(\chi^2 \geq \chi^2_{\text{р}}\right) = 0,05$, 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 $\alpha = 0,01 \ldots 0,05$ ($\alpha = 0,01$ 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 $\chi^2_{\text{р}} = 4,22$ линейной интерполяцией находим величину $\alpha = P\left(\chi^2 \geq \chi^2_{\text{р}}\right) = P\left(\chi^2 \geq 4,22\right) \approx 0,6$, т.е. вероятность того, что величина, распределённая по закону χ^2, превысит значение $\chi^2_{\text{р}}$.

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

$$\phi(x,r) := \frac{1}{2^{\frac{r}{2}} \cdot \Gamma\left(\frac{r}{2}\right)} \cdot x^{\frac{r}{2}-1} \cdot e^{-\frac{x}{2}} \qquad \alpha := \int_{4.22}^{\infty} \phi(x,5)\, dx = 0.5182$$

Поскольку вероятность $\alpha > \alpha_{\text{кр.}}$ $(0,6 > 0,05)$, 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) $1-0,05=0,95$.