

## Лабораторная работа

### «Критерии согласия»

С помощью критерия Пирсона  $\chi^2$  выполнить процедуру проверки гипотез о виде закона распределения для дискретной или непрерывной случайной величины.

#### Ход работы:

1. По заданной выборке (варианты задания приведены в приложении Б), состоящей из значений дискретной (или непрерывной) случайной величины, сформировать дискретный (или непрерывный) вариационный ряд и построить полигон частот (или гистограмму).
2. Сформулировать гипотезы.
3. В предположении о виде закона распределения рассчитать теоретические частоты. Эмпирические и теоретические частоты представить графически.
4. Рассчитать наблюдаемое значение критерия Пирсона  $\chi^2$  и сделать выводы.
5. Ответить на поставленные к задаче вопросы в соответствии с вариантом.

#### Теоретические сведения

Критерии согласия – статистические критерии, предназначенные для проверки гипотез о виде закона распределения.

Критерий Пирсона  $\chi^2$ .

$H_0$ : Генеральная совокупность имеет закон распределения  $P(x)$  (или  $f(x)$ ),

$H_1$ : Генеральная совокупность имеет распределение отличное от  $P(x)$  (или  $f(x)$ ).

В качестве критерия принимается случайная величина, распределенная по закону  $\chi^2$ , имеющая вид:

$$\chi^2 = \sum_i \frac{(n_i^{теор} - n_i^{эмпир})^2}{n_i^{теор}},$$

где  $n_i^{теор}$  - теоретические частоты, вычисленные в предположении закона распределения  $P(x)$  (или  $f(x)$ );

$n_i^{эмпир}$  - эмпирические частоты.

*Правило:*

Если наблюдаемое значение критерия (вычисленное по выборке)  
 $\chi^2_{набл} < \chi^2_{кр}(\alpha, k - 1 - r)$ , то нет оснований отвергать гипотезу  $H_0$ .

Если наблюдаемое значение критерия (вычисленное по выборке)  
 $\chi^2_{набл} > \chi^2_{кр}(\alpha, k - 1 - r)$ , то гипотезу  $H_0$  отвергают.

где  $\chi^2_{кр}(\alpha, k - 1 - r)$  - значение функции распределения  $\chi^2$  при уровне значимости  $\alpha$  и числе степеней свободы  $k - 1 - r$ ;

$k$  - количество групп или частичных интервалов;

$r$  - количество параметров распределения.

*Дискретное распределение.*

$X$  – дискретная случайная величина, закон распределения не известен. Проведено  $n$  опытов, известны эмпирические частоты наблюдаемых значений. Предполагается некоторый закон распределения. Проверить согласуется ли это предположение с данными испытаний.

Для этого вычисляют теоретические частоты  $n'_i$  в предположении выбранного закона.

Теоретическими (выравнивающими) частотами называют частоты, вычисленные теоретически. Так для дискретной случайной величины:

$$n'_i = nP_i = nP(X = x_i), \quad (2.1)$$

где  $n$  – кол-во испытаний;

$P_i$  – вероятность наблюдаемого значения  $x_i$ , вычисленная по предполагаемому закону распределения.

*Пример.* Контролировалось количество дефектов на 20 слитках. Данные приведены в таблице. Найти теоретические частоты, предполагая, что данные распределены по закону Пуассона. Проверить гипотезу о том, что данные подчиняются распределению Пуассона.

Таблица 2.1 – Эмпирические частоты

$x_i$ – кол-во появления события в одном опыте	0	1	2	...	N
$n_i$ – частота (кол-во опытов, в которых зарегистрировано событие), $\sum_{i=0}^N n_i = n$	$n_0$	$n_1$	$n_2$	...	$n_N$

$X$  – случайная величина, значение  $x_i$  которой соответствует кол-ву дефектов.

Таблица 2.2 – Эмпирические частоты

$x_i$ – кол-во дефектов	0	1	2	3	4	5
$n_i$ – частота	3	6	5	3	2	1

$$P_n(X = m) = \frac{\lambda^m \exp(-\lambda)}{m!} - \text{ф. Пуассона.} \quad (2.2)$$

где  $\lambda = MX = DX$  – параметр распределения Пуассона.

Оценкой  $\lambda$  является среднее значение  $\bar{x}$ :

$$\lambda \approx \bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = 1.9,$$

$k = 6$  - кол-во групп.

Теоретические частоты (2.1) вычислены в предположении закона Пуассона (2.2) приведены в таблице 2.3.

Таблица 2.3 – Эмпирические и теоретические частоты

$x_i$ – кол-во дефектов	0	1	2	3	4	5
$n_i$ – частота случаев	3	6	5	3	2	1
$n'_i$ - теор. частота	2,99	5,68	5,40	3,42	1,62	0,62

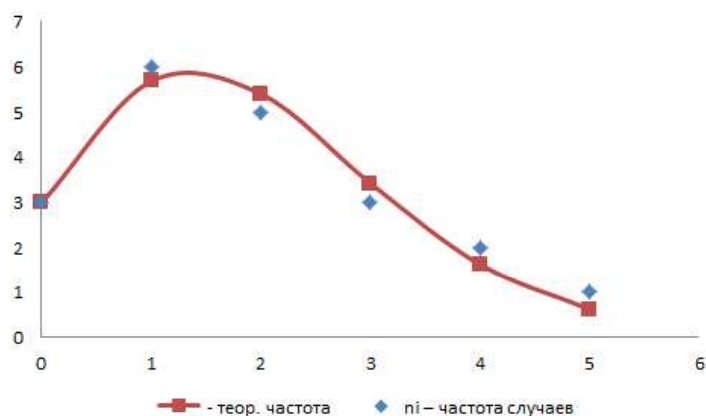


Рисунок 2.1 - Теоретические и эмпирические частоты

Графически видно согласование теоретических и эмпирических частот. К тому же, если рассчитать оценку дисперсии, то она окажется близка к мат. ожиданию, т.е  $\lambda = MX = DX$  (что характерно для распределения Пуассона):

$$S^2 = \frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = 1.99.$$

Для более точной оценки используют критерии согласия (критерий проверки гипотезы о предполагаемом законе неизвестного распределения), в частности критерий Пирсона Хи-квадрат. С помощью критерия проверяют является ли расхождение между эмпирическими и теоретическими частотами случайным (обусловленным малым количеством наблюдений, способом группировки и т.д.) или неслучайным и обусловлены неверным выбором закона распределения.

Для рассматриваемого примера в предположении распределения Пуассона при уровне значимости  $\alpha$  требуется проверить нулевую гипотезу:

$H_0$ : Случайная величина  $X$  имеет распределение Пуассона,

тогда альтернативная гипотеза может быть сформулирована так:

$H_1$ : Случайная величина  $X$  имеет распределение отличное от распределения Пуассона.

$$\chi_{набл}^2 = \sum_i \frac{(n_i^{теор} - n_i^{эмпир})^2}{n_i^{теор}} \approx 0.42$$

Наблюдаемое значение сравнивается с критическим значением, рассчитанным по закону распределения хи-квадрат с заданным уровнем значимости  $\alpha$  и количеством степеней свобод  $k - 1 - r$ :

$$\chi_{кр}^2(\alpha, k - 1 - r) = \chi_{кр}^2(0.05, 5) = 11.1$$

где  $k$  – кол-во вариант (значений случайной величины);

$r$  – кол-во параметров распределения.

Для рассматриваемого примера  $r=1$ , т.к. распределение Пуассона имеет один параметр  $\lambda$ .

Так как  $\chi_{набл}^2 < \chi_{кр}^2$ , то нет оснований отвергнуть гипотезу  $H_0$ .

*Замечание.* Малочисленные группы следует объединять, суммируя частоты.

*Непрерывное распределение.*

$X$  – непрерывная случайная величина, закон распределения не известен.

Вероятность принять отдельные значения равна нулю, поэтому весь интервал возможных значений делят на  $k$  частичных непересекающихся интервалов. Для определения количества интервалов часто используют формулу Стерджеса:

$$k = 1 + [3.32 \lg n].$$

По результатам опытов подсчитывают количество попаданий значений случайной величины в каждый интервал (эмпирические частоты). Для наглядности строят гистограмму частот (или относительных частот).

Теоретические частоты вычисляют аналогично (2.1):

$$n_i' = nP(x_i < X < x_{i+1}) = F(x_{i+1}) - F(x_i), \quad (2.5)$$

где  $F(x) = \int_{-\infty}^x f(x)dx$  - функция распределения непрерывной случайной величины.

Если предполагается, что закон распределения нормальный, то

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \quad (2.6)$$

Как правило значения параметров распределения неизвестны и оцениваются через среднее выборочное и исправленное выборочное СКО:

$$\mu^* = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \sigma^* = s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}.$$

Для проверки нулевой гипотезы о нормальном законе распределения также вычисляется значение критерия (2.3) и сравнивается с критическим значением

$\chi_{кр}^2(\alpha, k-1-r)$ , где  $k$  - количество интервалов,  $r=2$  - количество оцениваемых параметров для нормального распределения.

Пример. Проведено 100 равнооточных измерений напряжения. Проверить гипотезу о нормальности распределения.

8,42	5,28	6,75	5,97	5,43	4,66	6,29	5,88	1,86	5,37
5,82	6,16	7,18	1,28	0,48	5,88	4,49	7,09	4,63	3,98
7,24	5,10	5,33	1,75	5,06	6,23	4,55	4,57	3,75	7,46
6,48	7,25	5,17	5,62	5,48	1,24	4,93	4,12	6,50	2,32
5,16	5,74	6,01	1,80	6,60	5,50	8,17	3,76	8,94	6,21
8,07	1,60	5,53	7,54	4,87	1,73	7,72	6,19	5,54	4,15
3,87	2,12	3,75	3,71	4,21	6,74	2,39	7,47	3,33	6,01
5,16	8,66	5,31	6,55	7,46	6,32	5,18	3,96	4,14	6,02
2,54	3,05	3,38	2,69	4,48	4,99	4,59	5,67	2,72	8,24
5,40	4,75	4,63	4,96	3,04	4,00	7,07	3,87	3,82	5,92

Для построения гистограммы частот можно придерживаться следующего алгоритма:

- определить размах выборки:

$$R = X_{\max} - X_{\min} = 8.94 - 0.48 = 8.46.$$

- определить по формуле Стерджеса количество интервалов разбиения  $k$ :

$$k = 1 + [3.32 \lg n] = 7.$$

- определить длину и границы частичных интервалов:

$$\Delta x = \frac{R}{k} = \frac{8.46}{7} \approx 1.21. \quad x_1 = x_{\min}, \quad x_{i+1} = x_i + \Delta x, \quad i = \overline{1, k}.$$

- подсчитать эмпирические частоты для каждого интервала:

интервал	[0.48;1.69)	[1.69;2.9)	[2.9;4.11)	[4.11;5.31)	[5.31;6.52)	[6.52;7.73)	[7.73;8.94]
эмп.частота	4	10	14	25	27	14	6

Например, в пакете в некоторых математических пакетах программ можно воспользоваться стандартной функцией подсчета частот с параметрами  $k$  – количество интервалов и  $X$  – массив вариантов, например:

$$\text{V} := \text{histogram}(k, X) = \begin{pmatrix} 1.085 & 4 \\ 2.293 & 10 \\ 3.501 & 14 \\ 4.709 & 25 \\ 5.917 & 27 \\ 7.126 & 14 \\ 8.334 & 6 \end{pmatrix}$$

функция возвращает двумерный массив, первый столбец – середины интервалов, второй - соответствующие частоты.

- построить гистограмму частот как столбчатую диаграмму. По оси абсцисс откладываются интервалы, а по оси ординат соответствующие им частоты.

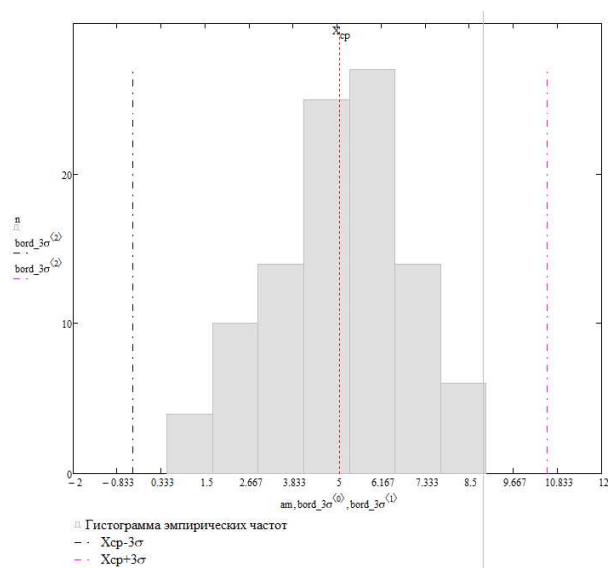


Рисунок 2.2 - Гистограмма эмпирических частот

По виду гистограммы (рис.2.2) можно предположить, что случайная величина подчиняется нормальному закону распределения. В связи с чем, можно рассчитать теоретические частоты по ф. (2.5)-(2.6).

Например, для первого интервала:

$$n'_1 = nP(0.48 < X < 1.69) = 100(F(1.69) - F(0.48)) =$$

$$= 100 \frac{1}{\sigma\sqrt{2\pi}} \int_{0.48}^{1.69} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \approx 3.70.$$

Аналогичным образом рассчитать частоты для оставшихся интервалов:

интервал	[0.48;1.69)	[1.69;2.9)	[2.9;4.11)	[4.11;5.31)	[5.31;6.52)	[6.52;7.73)	[7.73;8.94]
эмп.частота	4	10	14	25	27	14	5
теор.частота	2,66	8,59	18,2	25,4	23,3	14	5,54

Теоретические частоты также можно изобразить графически и сопоставить с эмпирическими частотами на рисунке 2.3.

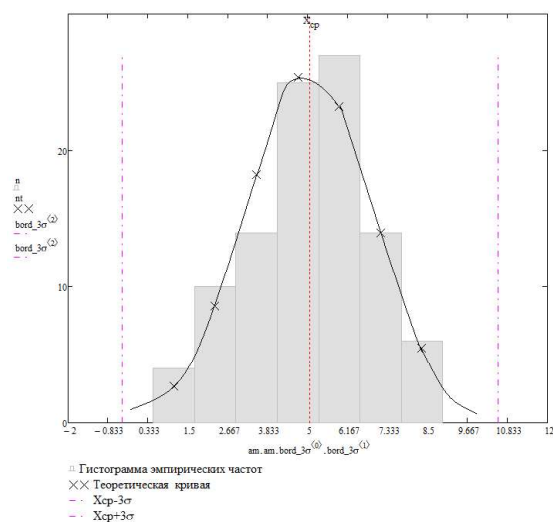


Рисунок 2.3 - Теоретические и эмпирические частоты

Качественно можно предположить, что теоретические и эмпирические частоты согласуются, однако для количественной оценки необходимо воспользоваться критерием согласия.

По ф. (3) расчетное значение критерия равно:

$$X_{набл}^2 = 2.5.$$

А критическое значение при  $\alpha=0,05$  и количестве степеней свобод  $k-1-2=4$  равно  $X_{кр}^2(0.05,4) = 9.5$ . Таким образом, нет оснований отвергнуть гипотезу о нормальности распределения.

### Примерные контрольные вопросы

1. Как проверить гипотезу о нормальном распределении, используя показатели асимметрии и эксцесса?
2. Как рассчитываются теоретические вероятности для дискретных и непрерывных случайных величин, зная закон распределения вероятностей (плотности вероятностей)?
3. Как рассчитывается по выборке значение критерия Пирсона  $\chi^2$ , каков смысл этой статистики?



## Приложение Б (обязательное)

**Вариант 1, 4, 7.** Еженедельная прибыль предприятия фиксировалась в течение года (таблица). Построить гистограмму частот и предположить вид закона распределения. При  $\alpha = 0,05$  проверить гипотезу о виде закона распределения и в случае ее принятия определить вероятность разорения этого предприятия, которая наступит, если еженедельная прибыль будет менее  $\Pi_{\min}$ .

Вариант 1 $\Pi_{\min}=12$ усл.ед.				Вариант 4 $\Pi_{\min}=21$ усл.ед.				Вариант 7 $\Pi_{\min}=5,5$ усл.ед.			
13,28	30,77	17,71	23,1	25,76	34,24	35,31	36,46	5,55	40,53	14,66	34,12
10,09	36,63	17,74	23,79	29,1	34,75	42,32	48,7	7,18	15,67	14,76	31,12
12,72	38,97	16,73	22,16	20,99	31,83	39,56	49,39	8,06	38,44	15,65	29,17
11,04	36,64	18,62	23,54	22,83	33,89	41,04	45,83	10,89	47,63	24,49	29,11
13,67	38,15	15,56	24,06	20,02	35,27	37,96	45,72	5,04	43,82	20,39	37,78
11,86	36,69	16,91	29,03	29,49	31,03	35,15	46,57	10,15	46,71	24,83	35,84
14,99	39,45	18,89	29,98	25,59	31,31	38,64	45,39	27,2	48,92	21,97	35,35
12,1	35,81	22,42	28,57	31,57	30,98	41,25	48,74	8	29,76	20,97	41,37
14,97	36,07	23,92	26,18	35,63	35,98	44,47	45,04	8,95	44,33	25,76	35,27
10,19	39,97	23,57	27,46	38,55	36,29	35,8	50,5	6,17	46,89	20,27	35,46
23,92	31,27	31,38	25,9	31,59	39,11	41,41	54	14,83	18,51	26,94	29,43
23,57	34,57	30,83	27,51	30,78	40,32	42,98	52,95	15,12	35,96	29,03	41
22,16	34,72	34,38	32,37	32,19	44,23	36,45	53,58	17,44	46,63	29,04	37,67

**Вариант 2, 5, 9, 11.** На испытание отобрана выборка в  $km$  болтов, измерение  $D$  диаметра (мм) которых производилось по  $m$  болтов через каждый час, всего проведено  $k$  серий измерений. Построить гистограмму частот и предположить вид закона распределения. При  $\alpha = 0,05$  проверить гипотезу о виде закона распределения и в случае ее принятия определить:

**Вариант 2, 5.** Долю брака, если поля допуска составляют соответственно:  $D_{\text{нижн}}$  и  $D_{\text{верх}}$ .

**Вариант 9, 11.** Будет ли принята партия, если доля брака не должна превышать 0,3%?

Вариант 2					Вариант 5					
$m=5, k=10, D_{\text{нижн}}=25,8$ и $D_{\text{верх}}=26,1$					$m=6, k=10, D_{\text{нижн}}=9,95$ и $D_{\text{верх}}=10,5$					
26,0	26,0	26,0	26,0	26,0	9,93	10,0	9,96	9,96	10,0	9,99
25,9	26,0	25,9	26,0	25,9	9,95	9,98	10,0	10,0	10,0	10
25,9	26,0	25,9	26,0	25,9	9,98	10,0	10,0	10,1	9,99	10,1
26,0	25,9	25,9	25,9	26,0	9,97	9,97	10	9,95	10,0	10,0
26,0	25,9	26,1	26,0	25,9	10,1	10,0	10,06	9,95	10,0	9,93
26,0	26,0	26,0	26,0	26,0	10,1	10,0	10	10,0	10,0	10
26,0	26,1	26,0	26,0	25,9	9,95	10,1	9,92	10,0	9,93	10,1
26,1	26,1	25,9	25,9	25,9	10,0	9,99	10	9,92	9,93	9,96
26,1	26,2	26,0	26,0	25,9	9,97	9,91	9,96	9,96	10,0	10,1
26,0	26,0	26,0	26,1	25,9	10,0	10,1	10,1	10,0	10,0	10,0

Вариант 9						Вариант 11					
m=6, k=9, D <sub>нижн</sub> = 15,8 и D <sub>верх</sub> =16,1						m=6, k=10, D <sub>нижн</sub> = 19,9 и D <sub>верх</sub> =20,1					
16,1	16,1	16,0	16,0	16,1	16,1	20,1	20,0	20,1	20,1	19,9	20,0
16,0	16,0	16,1	16,0	16,0	16,0	20,0	19,9	20,0	20,0	20,1	20,0
16,0	16,1	16,0	15,9	16,0	16,0	20,0	19,9	20,1	19,9	20,0	20,1
16,0	15,9	16,0	15,9	16,0	15,9	19,9	20,0	20,0	20,0	20,0	20,0
16,1	16,0	16,1	16,0	16,0	15,9	20,1	20,0	20,1	20,1	20,0	20,0
15,9	16,0	16,0	16,0	16,0	16,0	20,0	20,0	19,9	19,9	20,0	20,1
16,0	15,9	15,9	16,0	16,0	16,0	20,0	20,1	20,0	20,0	20,0	20,0
16,0	16,0	16,0	16,0	16,0	15,9	20,0	20,0	20,0	19,9	20,0	20,1
16,0	16,1	15,9	16,0	15,9	16,0	19,9	20,0	19,9	20,0	20,0	20,0
						20,0	20,1	19,9	19,9	20,0	20,0

**Вариант 3, 6, 12.** Значение срока службы (Т, мес) аккумуляторов смартфона некоторой модели Х приведены в таблице. Построить гистограмму частот и предположить вид закона распределения срока службы аккумулятора. При  $\alpha = 0,05$  проверить гипотезу о виде закона распределения и в случае ее принятия определить:

**Вариант 3.** Вероятность того, что аккумулятор купленного телефона указанной модели прослужит более 4 лет?

**Вариант 6.** У скольких из 200 проданных телефонов аккумулятор откажет в первый год службы?

**Вариант 12.** Сколько аккумуляторов из партии телефонов 1000 штук прослужит не менее 5 лет?

Вариант 3				Вариант 6				Вариант 12			
37	42	28	43	29	47	26	37	53	59	62	56
41	52	31	30	59	33	32	33	59	59	59	62
31	38	33	49	50	38	24	33	58	59	58	61
43	47	25	40	31	36	24	29	58	61	56	62
35	42	36	39	33	45	13	28	62	57	64	58
33	44	41	38	46	43	40	34	63	55	58	60
30	45	35	39	54	48	22	36	63	60	64	55
42	33	43	40	36	53	37	39	54	60	63	59
36	44	31	32	12	45	44	35	60	57	57	59
53	47	33	32	55	43	32	31	60	62	63	58
45	48	31	39	47	46	35	32	58	59	60	63
30	34	37	41	58	45	22	38	58	60	59	57
54	47	32	38	19	52	27	33	63	60	57	58
33	37	34	37	39	30	24	37	58	60	57	62
29	43	31	41	27	47	31	33	59	60	60	62
26	45	27	36	28	47	28	36	60	55	60	60
44	42	31	42	40	52	27	31	58	59	54	61
37	46	33	37	44	43	22	40	62	63	63	60
40	39	36	46	19	45	33	34	63	54	65	62
37	46	34	38	32	39	16	31	58	60	60	63

**Вариант 8, 10, 13.** Для оценки качества мороженого марки «А» («В», «С» по вариантам) из партии отобрали 30 образцов и измерили их жирность. В таблице представлены результаты измерений. Построить гистограмму частот и предположить вид закона распределения. При  $\alpha = 0,05$  проверить гипотезу о виде закона распределения и в случае ее принятия определить процент брака, если нижняя граница поля допуска составляет  $D_{\text{нижн}}$ .

Вариант 8			Вариант 10			Вариант 13		
Пломбир «А», $D_{\text{нижн}} = 12\%$			Сливочное «В», $D_{\text{нижн}} = 8\%$			Молочное «С», $D_{\text{нижн}} = 3\%$		
12,6	12,4	12,5	8,38	8,27	8,23	3,26	3,17	3,22
12,7	12,5	12,4	8,27	8,23	8,15	3,31	3,04	3,12
12,3	12,1	12,4	8,27	8,31	8,25	3,11	3,25	3,16
12,5	12,5	12,4	8,21	8,48	8,21	3,28	3,33	3,22
12,7	12,4	12,2	8,31	8,33	8,24	3,21	3,23	3,12
12,6	12,3	12,4	8,29	8,22	8,30	3,21	3,21	3,16
12,5	12,5	12,5	8,26	8,35	8,34	3,21	3,31	2,95
12,3	12,8	12,5	8,55	8,29	8,51	3,21	3,11	3,26
12,6	12,5	12,4	8,33	8,15	8,30	3,22	3,21	3,26
12,3	12,3	12,4	8,33	8,28	8,47	3,29	3,24	3,28

**Вариант 14.** Произведено 100 измерений времени обслуживания (в мин) клиентов в Центральном банке города N. При  $\alpha = 0,05$  проверить гипотезу о виде закона распределения и в случае ее принятия определить вероятность того, что время обслуживания клиента составит не более 15 мин ?

20,4	6,3	29,7	25,1	19,9	13	16,1	14,3	22,9	19,7
8	16,1	11,3	22,7	14,1	19,5	14,4	19,4	18,4	18,1
24,1	24,3	18,2	18,3	19,1	18,8	17,5	13,4	16,5	10,2
16,1	18,6	11,6	14	7,5	19,1	11,1	12,2	15	10,7
10,3	16,7	24,6	21,9	15,6	16,6	17,4	21,1	13,4	13,1
26,7	11,1	5,2	21	7,7	15,6	15,1	18,3	22,2	19,2
16,7	23,4	25,2	14,7	11,3	12,2	18,3	21,8	22,3	18,5
21,5	24,1	24	9,8	15,4	29,6	18,1	18,9	19,7	14,7
21,2	13,2	9,2	18,8	19,4	12	16,1	21,3	22,2	18,6
20,7	14,7	24,4	13,6	12,3	27	25,8	17,9	9,5	20,9