Задача №1 Вариант 301 Ягубов Р.Б.

Дана генеральная совокупность, состоящая из случайных величин, являющихся результатами некоторого опыта. Необходимо проанализировать данные величины, выдвинуть гипотезу о виде их распределения и подтвердить или опровергнуть её с определённой долей вероятности.

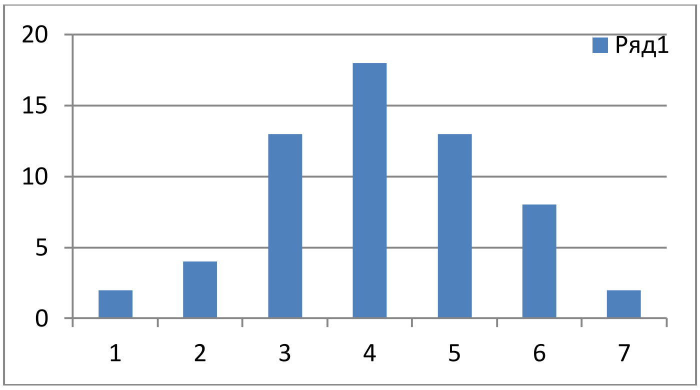
Генеральная совокупность $\vec{x}_n = (x_1, x_2, ..., x_n)$, $n = 60$, её вариационный ряд $\vec{x}_n = (x_{(1)} \le x_{(2)} \le ... \le x_{(n)})$ имеет вид:

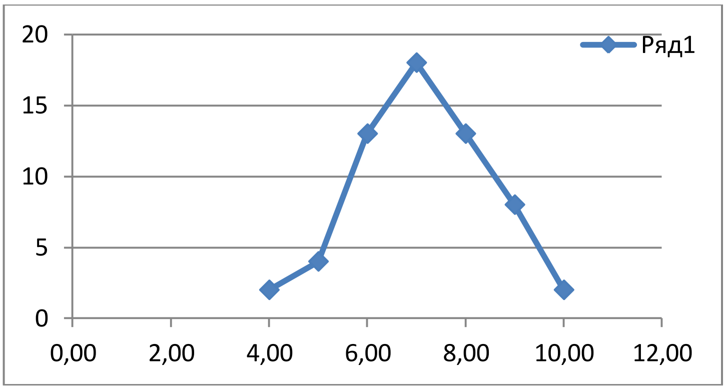
3	3,84	4,11	4,52	4,54	4,98	5,2	5,36	5,5	5,53
5,57	5,57	5,66	5,74	5,74	5,87	5,88	5,88	5,9	6,02
6,09	6,09	6,11	6,14	6,18	6,21	6,22	6,36	6,5	6,54
6,58	6,79	6,81	6,81	6,88	6,89	6,91	7,06	7,16	7,16
7,33	7,34	7,39	7,4	7,4	7,53	7,7	7,7	7,77	7,87
8,06	8,1	8,28	8,32	8,41	8,51	8,51	8,69	9,06	10

Возьмём количество разбиений равное 7-ми, тогда $h = \dfrac{x_{(n)} - x_{(1)}}{7} = 1$

N	x_i	x_{i+1}	n_i	n_i/n	x^*
1,00	3,00	4,00	2,00	0,03	3,50
2,00	4,00	5,00	4,00	0,07	4,50
3,00	5,00	6,00	13,00	0,22	5,50
4,00	6,00	7,00	18,00	0,30	6,50
5,00	7,00	8,00	13,00	0,22	7,50
6,00	8,00	9,00	8,00	0,13	8,50
7,00	9,00	10,00	2,00	0,03	9,50

Для приближённого определения плотности теоретической вероятности функции распределения, на основе данной таблицы построим гистограмму и полигон:





По виду гистограммы мы предположим, что наша генеральная совокупность имеет нормальное распределение.

$$H_0 : N(a, \sigma^2), \qquad f(x, a, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

где a – математическое ожидание, а σ^2 – дисперсия нормального распределения.

В качестве оценки параметра a возьмём выборочное среднее, а в качестве оценки параметра σ^2 возьмём исправленную выборочную дисперсию:

$$\hat{a}(\vec{x}_n) = \bar{x} = 6.62 \qquad \hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{7}(x_i^* - \bar{x})^2 = 1.85 \Rightarrow \hat{\sigma} = S = 1.36$$

Функция распределения с учётом найденных параметров примет вид:

$$f(x) = \frac{1}{1.36\sqrt{2\pi}} e^{-\frac{(x-6.62)^2}{3.69}} \approx 0.29 e^{-\frac{(x-6.62)^2}{3.69}}$$

Критерий Пирсона

W – статистика Пирсона, имеющая распределения χ^2.

$$W = \sum_{i=1}^{n} \frac{(n_i - n_i^{\tau})^2}{n_i^{\tau}} = \sum_{i=1}^{7} \frac{n_i^2}{n_i^{\tau}} - n \qquad n_i^{\tau} = np_i$$

Вероятность попадания случайной величины в интервал для нормального распределения описывается уравнением:

$$P(x_i < \xi < x_{i+1}) = \Phi\left(\frac{x_{i+1} - a}{\sigma}\right) - \Phi\left(\frac{x_i - a}{\sigma}\right)$$

$$z_i = \frac{x_i - \overline{x}}{S_1}, \qquad z_{i+1} = \frac{x_{i+1} - \overline{x}}{S_1} \Rightarrow P(x_i < \xi < x_{i+1}) = \Phi(z_{i+1}) - \Phi(z_i)$$

z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	n_i^{τ}	n_i^2/n_i^{τ}
-2,71	-1,96	-0,50	-0,48	0,02	1,30	3,08
-1,96	-1,21	-0,48	-0,39	0,09	5,27	3,04
-1,21	-0,46	-0,39	-0,18	0,21	12,52	13,50
-0,46	0,28	-0,18	0,11	0,29	17,43	18,59
0,28	1,03	0,11	0,35	0,24	14,23	11,88
1,03	1,78	0,35	0,46	0,11	6,81	9,40
1,78	2,53	0,46	0,49	0,03	1,91	2,10

После расчётов приведённых в таблице получили, что $W_{наб} = 1.59$

Теперь выясним, является ли $W_{наб}$ допустимой. Пусть $\alpha = 0.01$ - уровень значимости (вероятность совершить ошибку), в свою очередь $1 - \alpha = 0.99$

$$P\left(W > \delta_{кр} \mid H_0\right) = \alpha; \qquad W \sim \chi^2(r)$$

где r - число степеней свободы, равное: $r = k - 1 - l = 7 - 1 - 2 = 4 \Rightarrow \qquad W \sim \chi^2(4)$

$$\delta_{кр} = \chi^2_{(1-\alpha)}(4) = \chi^2_{(0,99)}(4) = 13.28$$

Мы видим, что $W_{набл} < \delta_{кр} \Rightarrow$ <u>с надёжностью 0.99 наши экспериментальные результаты описываются нормальным распределением</u>.

Построим доверительные интервалы для параметров a и σ^2:

Так как параметр σ^2 неизвестен, то воспользуемся известной статистикой, имеющей распределение Стьюдента с $n-1$ степенями свободы:

$$T(n-1) = \frac{\overline{x} - a}{S}\sqrt{n}; \overline{x} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1) < a < \overline{x} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2}}(n-1)$$

$$t_{0,995}(59) = 2.66; \qquad 6.43 < a < 7.01$$

То есть мы получили, что $a = 6.62 \pm 0.39$ с надёжностью 0.99

Для параметра σ^2 возьмём статистику вида:

$$\chi^2(n-1) = (n-1)\frac{S^2}{\sigma^2}; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$$

$$\chi^2_{0,995}(59) = 91; \qquad \chi^2_{0,005}(59) = 36; \Rightarrow \qquad 1.2 < \sigma^2 < 3.03 \Rightarrow 1.09 < \sigma < 1.74$$

Мы получили, что $\sigma = 1.36 \pm 0.38$ с надёжностью 0.99

Критерий Колмогорова

Суть метода состоит в сравнении интегральной и эмпирической функций распределения.

$\hat{F}_n(t, \vec{x}_n) = \dfrac{\mu(t, \vec{x}_n)}{n}$ – эмпирическая функция распределения.

Статистика Колмогорова имеет вид: $D(\vec{x}_n) = \sup\limits_{x \in R} |\hat{F}(t, \vec{x}_n) - F(t)|$.

$\hat{F}(t, \overrightarrow{x_n})$	$F(x_i^*)$	$\lvert\hat{F} - F\rvert$
0,03	0,01	0,02
0,10	0,06	0,94
0,32	0,20	0,11
0,62	0,46	0,15
0,83	0,74	0,09
0,97	0,92	0,05
1,00	0,98	0,02

$$D_{набл} = 0.94; \quad P(D(\vec{x}_n) > d_{кр} \mid H_1) = 0.01; \quad d_{кр} = 0.207$$

Видно, что гипотеза H_1 не выполняется, так $D_{набл} < d_{кр} \Rightarrow$ что <u>наши экспериментальные данные описываются нормальным законом распределения с надёжностью 0,99</u>

Задача №2 Вариант 207 Ягубов Р.Б.

Дана генеральная совокупность, состоящая из случайных величин, являющихся результатами некоторого опыта. Необходимо проанализировать данные величины, выдвинуть гипотезу о виде их распределения и подтвердить или опровергнуть её с определённой долей вероятности.

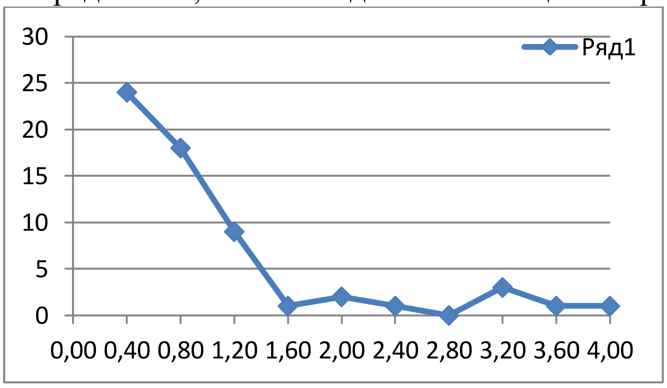
Генеральная совокупность $\vec{x}_n = (x_1, x_2, ..., x_n)$, $n = 60$, её вариационный ряд $\vec{x}_n = (x_{(1)} \le x_{(2)} \le ... \le x_{(n)})$ имеет вид:

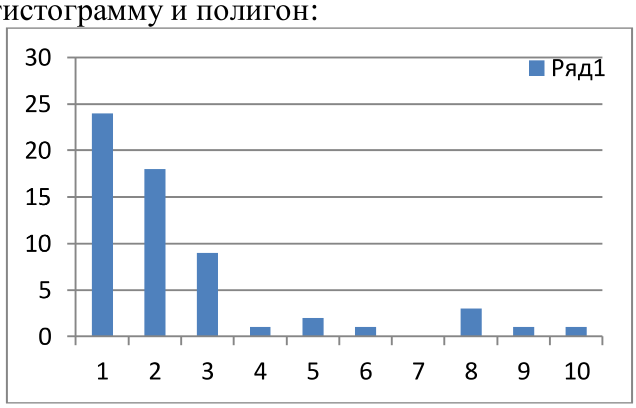
0,00	0,02	0,03	0,04	0,04	0,06	0,07	0,15	0,17	0,17
0,18	0,19	0,23	0,24	0,24	0,29	0,30	0,30	0,32	0,33
0,37	0,37	0,37	0,39	0,42	0,43	0,43	0,43	0,44	0,46
0,47	0,55	0,58	0,64	0,64	0,66	0,66	0,66	0,69	0,71
0,75	0,80	0,89	0,94	0,96	0,96	0,99	0,99	1,08	1,08
1,11	1,28	1,65	1,84	2,19	3,05	3,05	3,05	3,40	4,00

Возьмём количество разбиений равное 10-ми, тогда $h = \frac{x_{(n)} - x_{(1)}}{10} = 0.4$

N	x_i	x_{i+1}	n_i	n_i/n	x^*
1	0,00	0,40	24	2/5	0,2
2	0,40	0,80	18	1/5	0,6
3	0,80	1,20	9	1/9	1
4	1,20	1,60	1	1/60	1,4
5	1,60	2,00	2	1/20	1,8
6	2,00	2,40	1	1/20	2,2
7	2,40	2,80	0	1/15	2,6
8	2,80	3,20	3	1/60	3
9	3,20	3,60	1	1/30	3,4
10	3,60	4,00	1	1/60	3,8

Для приближённого определения плотности теоретической вероятности функции распределения, на основе данной таблицы построим гистограмму и полигон:





По виду гистограммы мы предположим, что наша генеральная совокупность имеет экспоненциальное распределение.

$$H_0 : Exp(t, \lambda), \qquad f(t, \lambda) = \begin{cases} 0, t < 0 \\ \lambda e^{-\lambda t}, t \ge 0 \end{cases}$$

В качестве оценки параметра λ возьмём выборочное среднее в минус первой степени:

$$\hat{\lambda} = \frac{1}{\bar{x}} = 1.25$$

Функция распределения с учётом найденных параметров примет вид:

$$f(t, \hat{\lambda}) = \begin{cases} 0, t < 0 \\ 1,25 e^{-1,25t}, t \ge 0 \end{cases}$$

Критерий Пирсона

W – статистика Пирсона, имеющая распределения χ^2.

$$W = \sum_{i=1}^{n} \frac{(n_i - n_i^{\tau})^2}{n_i^{\tau}} = \sum_{i=1}^{7} \frac{n_i^2}{n_i^{\tau}} - n; \qquad n_i^{\tau} = np_i.$$

Известно, что вероятность попадания случайной величины в интервал для показательного распределения описывается уравнением:

$$P(x_i < \xi < x_{i+1}) = F(x_{i+1}) - F(x_i)$$

$$F(t, \hat{\lambda}) = \begin{cases} 0, t < 0 \\ 1 - e^{-\hat{\lambda} t}, t \ge 0 \end{cases} = \begin{cases} 0, t < 0 \\ 1 - e^{-1{,}25t}, t \ge 0 \end{cases}$$

x_i	x_{i+1}	$F(x_i)$	$F(x_{i+1})$	p_i	n_i^{τ}	n_i^2/n_i^{τ}
0	0,4	0	0,393469	0,393469	23,60816	24,39834
0,4	0,8	0,393469	0,632121	0,238651	14,31907	22,62716
0,8	1,2	0,632121	0,77687	0,144749	8,684957	9,326471
1,2	1,6	0,77687	0,864665	0,087795	5,267693	0,189836
1,6	2	0,864665	0,917915	0,05325	3,195017	1,251949
2	2,4	0,917915	0,950213	0,032298	1,937876	0,516029
2,4	2,8	0,950213	0,969803	0,01959	1,175381	0
2,8	3,2	0,969803	0,981684	0,011882	0,712905	12,62441
3,2	3,6	0,981684	0,988891	0,007207	0,432399	2,312681
3,6	4	0,988891	0,993262	0,004371	0,262263	3,812967

После расчётов приведённых в таблице получили, что $W_{наблюд} = 17.06$

Теперь выясним, является ли $W_{наблюд}$ допустимой. Пусть $\alpha = 0.05$- уровень значимости (вероятность совершить ошибку), в свою очередь $1 - \alpha = 0.95$

$$P\left(W > \delta_{кр} \mid H\right) = \alpha; \qquad W \sim \chi^2(r)$$

где r - число степеней свободы, равное: $r = k - 1 - l = 10 - 1 - 1 = 8 \Rightarrow \quad W \sim \chi^2(8)$

$$\delta_{кр} = \chi^2_{(1-\alpha)}(8) = \chi^2_{(0,95)}(8) = 15.51$$

Мы видим, что $W_{набл} < \delta_{кр} \Rightarrow$ <u>с надёжностью 0.95 наши экспериментальные результаты не описываются экспоненциальным законом распределением</u>.

Критерий Колмогорова

Суть метода состоит в сравнении интегральной и эмпирической функций распределения.

$\hat{F}_n(t, \vec{x}_n) = \dfrac{\mu(t, \vec{x}_n)}{n}$ – эмпирическая функция распределения.

Статистика Колмогорова имеет вид: $D(\vec{x}_n) = \sup\limits_{x \in R} |\hat{F}(t, \vec{x}_n) - F(t)|$.

$\hat{F}(t, \overrightarrow{x_n})$	$F(x_i^*)$	$\lvert\hat{F} - F\rvert$
0,4	0,318869	0,081131
0,7	0,683996	0,016004
0,85	0,853393	-0,00339
0,866667	0,931983	-0,06532
0,9	0,968444	0,068444
0,916667	0,98536	0,068693
0,916667	0,993208	-0,07654
0,966667	0,996849	-0,03018
0,983333	0,998538	-0,0152
1	0,999322	0,000678

$$D_{набл} = 0.08; \qquad P(D(\vec{x}_n) > d_{кр} \mid \mathrm{H}) = 0.01; \qquad d_{кр} = 0.207$$

Видно, что гипотеза H не выполняется, так $D_{набл} < d_{кр} \Rightarrow$ <u>что наши экспериментальные данные описываются экспоненциальным законом распределения с надёжностью 0,99</u>.