Студент(ка):	Иванов Иван
Группа:	0380
Вариант:	1
Дата:	26 ноября 2020 г.

Статистический анализ
Индивидуальное домашнее задание №2

Вар. 4 (8383 2020)

1. В результате эксперимента получены данные, приведенные в таблице 1.

a) Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.

b) Вычислить выборочные аналоги следующих числовых характеристик:
(i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса,
(vi) вероятности $\mathbf{P}(X \in [a, b])$.

c) В предположении, что исходные наблюдения являются выборкой из распределения Пуассона, построить оценку максимального правдоподобия параметра λ, а также оценку λ по методу моментов. Найти смещение оценок.

d) Построить асимптотический доверительный интервал уровня значимости α_1 для параметра λ на базе оценки максимального правдоподобия.

e) Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением Пуассона с параметром λ_0. Проверить гипотезу на уровне значимости α_1. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

f) Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением Пуассона. Проверить гипотезу на уровне значимости α_1. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

g) Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром $\lambda = \lambda_0$ при альтернативе пуассоновости с параметром $\lambda = \lambda_1$. Проверить гипотезу на уровне значимости α_1. Что получится, если поменять местами основную и альтернативную гипотезы?

h) В пунктах (c)–(f) заменить семейство распределений Пуассона на семейство геометрических распределений

$$\mathbf{P}_\lambda(X = k) = \frac{\lambda^k}{(\lambda + 1)^{k+1}}, \; k = 0, 1, \ldots .$$

Таблица 1 $\alpha_1 = 0.05$; $a = 3.66$; $b = 5.45$; $\lambda_0 = 6.00$; $\lambda_1 = 5.00$.

4 5 5 2 1 5 3 5 5 6 4 5 1 7 2 5 6 10 4 10 1 8 6 6 10 2 3 4 4 10 9 7 7 9 7 2 5 1 4 2 7 4 6 7 4
2 4 5 3 6

2. В результате эксперимента получены данные, приведенные в таблице 2.

a) Построить вариационный ряд, эмпирическую функцию распределения, гистограмму и полигон частот с шагом h.

b) Вычислить выборочные аналоги следующих числовых характеристик:
(i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса,
(vi) вероятности $\mathbf{P}(X \in [c, d])$.

c) В предположении, что исходные наблюдения являются выборкой из нормального распределения, построить оценку максимального правдоподобия параметров (a, σ^2) и соответствующие оценки по методу моментов. Найти смещение оценок.

d) Построить доверительные интервалы уровня значимости α_2 для параметров (a, σ^2).

e) С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с нормальным распределением с параметрами a_0, σ_0^2. Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

f) Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с нормальным распределением с параметрами (a_0, σ_0^2). Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

g) Построить критерий проверки значимости χ^2 сложной гипотезы согласия с нормальным распределением. Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

h) Построить наиболее мощный критерий проверки простой гипотезы о нормальности с параметром $(a, \sigma^2) = (a_0, \sigma_0^2)$ при альтернативе нормальности с параметром $(a, \sigma^2) = (a_1, \sigma_1^2)$. Проверить гипотезу на уровне значимости α_2. Что получится, если поменять местами основную и альтернативную гипотезы?

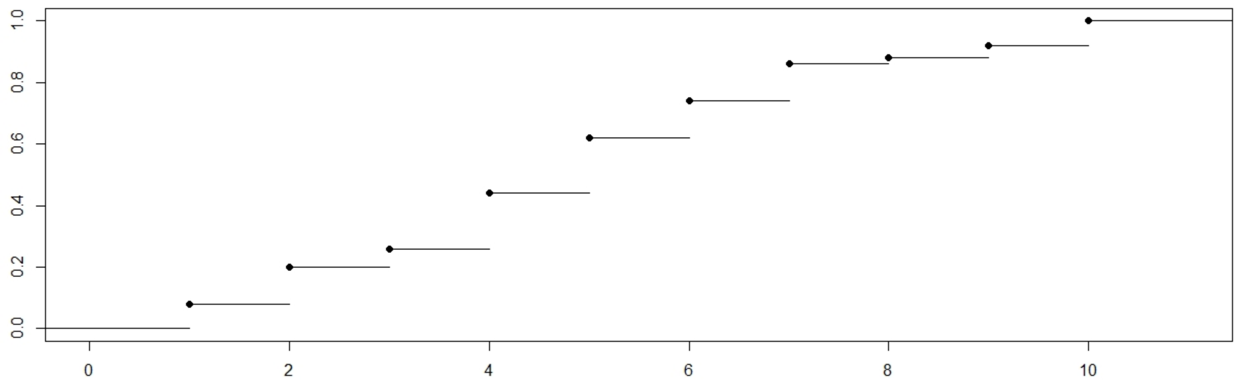
i) В пунктах (c)–(g) заменить семейство нормальных распределений на двухпараметрическое семейство распределений Лапласа с плотностями $f(x) = \frac{1}{\sqrt{2}\,\sigma} \mathrm{e}^{-\frac{\sqrt{2}}{\sigma}|x-a|}$.

Таблица 2 $\alpha_2 = 0.01$; $c = -3.40$; $d = -1.00$; $h = 0.80$; $a_0 = -3.00$; $\sigma_0 = 2.00$; $a_1 = -0.00$; $\sigma_1 = 2.00$.

0.717 −1.085 −2.259 −4.178 −4.180 −3.894 −1.989 −2.612 −3.116 −3.240 −5.980 −0.722 −2.648 −5.245 −5.583 −6.895
−3.993 −4.575 −2.680 −1.620 −3.529 −3.665 −3.267 −0.875 −1.739 −4.847 −0.290 −1.618 −5.660 −3.602 −4.470 −3.298
−8.996 −6.009 −0.546 0.452 −0.360 −4.743 −3.538 −1.820 −2.589 −1.196 −0.934 −2.608 −4.815 −1.056 −2.464 −1.045
−1.425 −4.708

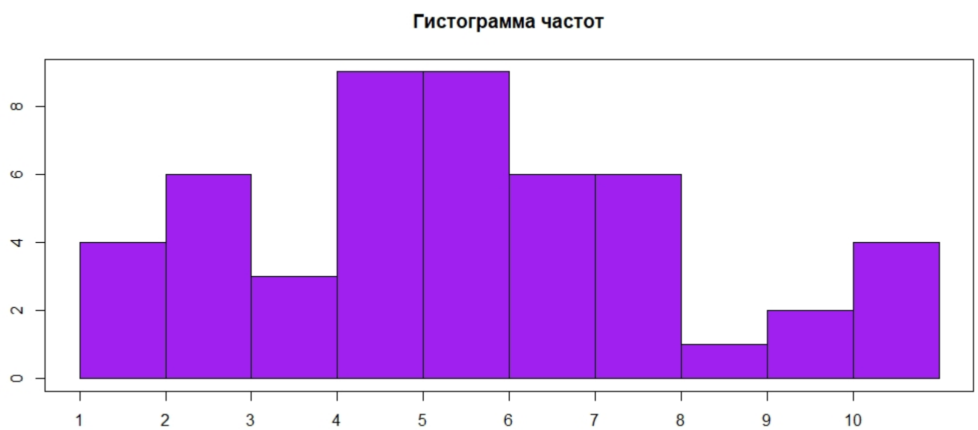
Задача 1. *Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.*

Решение. Вариационный ряд
1 1 1 1 2 2 2 2 2 2 3 3 3 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 8 9 9 10 10 10 10
Эмпирическая функция распределения



Гистограмма частот

$\square$



Задача 2. *Вычислить выборочные аналоги числовых характеристик*

Решение. Выборочное среднее
$\bar{x} = \frac{1}{n} * \sum_{i=1}^{n}(x_i) = 5$
Выборочная дисперсия
$s^2 = \frac{1}{n} * \sum_{i=1}^{n}(x_i - \bar{x})^2 = 6.33$
Медиана
$Me = 5$
Ассиметрия
$As = \frac{\mu_3^*}{s^3} = 0.324$

Эксцесс
$Ex = \frac{\mu_4^*}{s^4} = 2.38$
Вероятность попадания в заданный промежуток
$P(X \in [3.66; 5.45]) = F(b) - F(a) = 0.36$ $\square$

Задача 3. *В предположении, что исходные наблюдения являются выборкой из распределения Пуассона, построить оценку максимального правдоподобия параметра* λ*, а также оценку* λ *по методу моментов. Найти смещение оценок.*

Решение. По методу максимального правдоподобия
$P_\lambda(X = k) = \frac{\lambda^k}{k!} * exp(-\lambda)$
$l(\overrightarrow{x}, \lambda) = \frac{\lambda^{\sum x_i} * exp(-\lambda * n)}{\prod^n x_i}$
$ll(\overrightarrow{x}, \lambda) = \sum^n x_i ln\lambda - n\lambda - ln \prod^n x_i!$
$\frac{dll}{d\lambda} = \frac{1}{\lambda} * \sum^n x_i - n = 0 \to \hat{\lambda} = \frac{\sum x_i}{n} = \bar{x}$
По методу моментов
$M_1^* = \bar{x}$
$M_1 = \lambda$
$\hat{\lambda} = \bar{x}$
Смещение оценки
$\mathbb{E}\hat{\lambda} = \frac{1}{n} \sum^n Ex_i = \frac{n\lambda}{n} = \lambda$
Оценка несмещенная $\square$

Задача 4. *Построить асимптотический доверительный интервал уровня значимости* $\alpha_1 = 0.05$ *для параметра* λ *на базе оценки максимального правдоподобия.*

Решение. Из предыдущего пункта $\hat{\lambda} = \bar{x}$
$\gamma = 1 - \alpha = 0.95$
$\frac{\sqrt{n}(\bar{x} - \lambda)}{\sqrt{\lambda}} \sim \mathcal{N}(0, 1)$
$\Phi(t_\gamma) = 1 - \frac{\alpha_1}{2}$
$P(-t_\gamma \leq \frac{\sqrt{n}(\bar{x} - \lambda)}{\sqrt{\lambda}} \leq t_\gamma) \to 1 - \alpha$
$n(\bar{x} - \lambda)^2 = t_\gamma^2 \lambda$
$\lambda^2 - 2\lambda(\bar{x} + \frac{t_\gamma^2}{2n}) + \bar{x}^2 = 0$
$\lambda_{1,2} = \bar{x} + \frac{t_\gamma^2}{2n} \pm t_\gamma \sqrt{\frac{1}{n}(\bar{x} + \frac{t_\gamma^2}{4n})}$
Отсюда доверительный интервал $[4.38; 5.62]$
$\square$

Задача 5. *Используя гистограмму частот, построить критерий значимости* χ^2 *проверки простой гипотезы согласия с распределением Пуассона с параметром* $\lambda_0 = 6.00$*. Проверить гипотезу на уровне значимости* $\alpha_1 = 0.05$*. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение. Составим таблицу

x_i	m_i	p_i	np_i	$(\frac{m_i - np_i}{\sqrt{m_i}})^2$
1	4	0.017	0.85	11.31
2	6	0.045	2.25	6.37
3	3	0.089	4.45	0.48
4	9	0.134	6.7	0.8
5	9	0.16	8	0.117
6	6	0.16	8	0.514
7	6	0.138	6.9	0.113
8	1	0.103	5.15	3.357
9	2	0.069	3.45	0.604
10	4	0.084	4.2	0.009

Полученное значение $\chi^2 = 23.67$
Число степеней свободы $l = k - r - 1 = 8$
$\chi^2_{critical} = 15.5$
Следовательно, H_0 отвергается. Наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу: $1 - \chi^2_{r=8}(23.67) = 0.003$

□

Задача 6. *Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением Пуассона. Проверить гипотезу на уровне значимости $\alpha_1 =$ 0.05. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение. Сложная гипотеза $H_0 : X_1,...,X_n \sim Pois(\lambda)$
Известно, что в случае регулярности эксперимента статистика$\hat{\chi}^2(\lambda) = min\chi^2(\lambda)$ сходится к χ^2_{r-d-1}
Критерий имеет вид

$$\phi(\overrightarrow{x}) = \begin{cases} 0 & \text{if } \hat{\chi}^2 \leq x_\alpha \\ 1 & \text{if } \hat{\chi}^2 > x_\alpha \end{cases}$$

В результате вычислений на R получили, что $\hat{\chi}^2 = 25.53 > x_\alpha = 15.5$
Таким образом, гипотеза отвергается. Наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу, крайне мало: $1 - \chi^2_{r=8}(25.53) = 0.001$

□

Задача 7. *Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром $\lambda = 6.00$ при альтернативе пуассоновости с параметром $\lambda = 5.00$. Проверить гипотезу на уровне значимости 0.05. Что получится, если поменять местами основную и альтернативную гипотезы?*

Решение. Сформулируем гипотезы
$H_0 : \lambda = \lambda_0 = 6$
$H_1 : \lambda = \lambda_1 = 5$
По лемме Неймана-Пирсона:

$$\phi(\overrightarrow{x}) = \begin{cases} 0 & \text{if } l(\overrightarrow{x}, \lambda_0, \lambda_1) < C \\ p & \text{if } l(\overrightarrow{x}, \lambda_0, \lambda_1) = C \\ 1 & \text{if } l(\overrightarrow{x}, \lambda_0, \lambda_1) > C \end{cases}$$

$l(\overrightarrow{x}, 6, 5) = \frac{L(\overrightarrow{x},5)}{L(\overrightarrow{x},6)} = (\frac{5}{6})^{\sum x_i} * exp(n(\lambda_0 - \lambda_1)) = (\frac{5}{6})^{\sum x_i} * exp(n)$
$ll(\overrightarrow{x}, \lambda_0, \lambda_1) = \sum x_i * ln\frac{5}{6} + n < ln(C)$
$\sum x_i < \frac{ln(C) - n}{ln\frac{5}{6}}$
$\hat{C} = \frac{ln(C) - n}{ln\frac{5}{6}}$

Критерий принимает вид

$$\phi(\overrightarrow{x}) = \begin{cases} 0 & \text{if } \sum x_i < \hat{C} \\ p & \text{if } \sum x_i = \hat{C} \\ 1 & \text{if } \sum x_i > \hat{C} \end{cases}$$

Вычислим $\hat{C}$ и p из уравнения:
$P_{\lambda_0}(l(\overrightarrow{x}, \lambda_0, \lambda_1) > C) + p * P_{\lambda_0}(l(\overrightarrow{x}, \lambda_0, \lambda_1) = C) = P_{\lambda_0}(\sum_{i=1}^n x_i > \hat{C}) + p * P_{\lambda_0}(\sum_{i=1}^n x_i = \hat{C}) = \alpha_1 = 0.05$
$x_i \to Pois(\lambda_0)$
$\sum x_i \to Pois(n\lambda_0)$
Подбором среди целых чисел можно найти такое наибольшее $\hat{C}$ и α_0, что
$\alpha_0 = P_{\lambda_0}(\sum_{i=1}^n x_i > \hat{C}) = 1 - P_{n\lambda_0}(\hat{C}) - p_{n\lambda_0}(\hat{C}) < \alpha_1$
$p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_0}(\sum_{i=1}^n x_i = A)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_0}(A)}$
В результате расчета получили $\alpha_0 = 0.04536796, \hat{C} = 328, p = 0.747331$
Таким образом, принимаем гипотезу H_0
Теперь поменяем местами основную и альтернативную гипотезы
$H_0 : \lambda = \lambda_1 = 5$
$H_1 : \lambda = \lambda_0 = 6$
$l(\overrightarrow{x}, 5, 6) = \frac{L(\overrightarrow{x}, 6)}{L(\overrightarrow{x}, 5)} = (\frac{6}{5})^{\sum x_i} * exp(n(\lambda_1 - \lambda_0)) = (\frac{6}{5})^{\sum x_i} * exp(-n)$
$ll(\overrightarrow{x}, \lambda_0, \lambda_1) = -\sum x_i * ln\frac{5}{6} - n < ln(C)$
$\sum x_i > \frac{-ln(C) - n}{ln\frac{5}{6}}$
$\hat{C} = \frac{-ln(C) - n}{ln\frac{5}{6}}$
Тогда критерий примет вид

$$\phi(\overrightarrow{x}) = \begin{cases} 0 & \text{if } \sum x_i > \hat{C} \\ p & \text{if } \sum x_i = \hat{C} \\ 1 & \text{if } \sum x_i < \hat{C} \end{cases}$$

Вычислим $\hat{C}$ и p из уравнения: $P_{\lambda_1}(\sum_{i=1}^n x_i > \hat{C}) + p * P_{\lambda_1}(\sum_{i=1}^n x_i = \hat{C}) = \alpha_1 = 0.05$
$p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_1}(\sum_{i=1}^n x_i = A)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_1}(A)}$
В результате расчета получили $\alpha_0 = 0.0449179, \hat{C} = 223, p = 0.8644936$
Таким образом, принимаем гипотезу H_0
При замене основной и альтернативной гипотезы меняется также гипотеза, которая принимается. Но т. к. изменение происходит со сменой гипотез местами, решение не меняется ☐

Задача 8. *В предположении, что исходные наблюдения являются выборкой из геометрического распределения, построить оценку максимального правдоподобия параметра λ, а также оценку λ по методу моментов. Найти смещение оценок.*

Решение. Плотность геометрического распределения имеет вид $P_\lambda(X = k) = \frac{\lambda^k}{(\lambda+1)^{k+1}}$
$l(\overrightarrow{x}, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{(\lambda+1)^{x_i+1}} = \frac{\lambda^{\sum_{i=1}^n x_i}}{(\lambda+1)^{\sum_{i=1}^n x_i + n}}$
$ll(\overrightarrow{x}, \lambda) = ln\lambda * \sum_{i=1}^n x_i - ln(\lambda + 1) \sum_{i=1}^n x_i - nln(\lambda + 1)$
$\frac{dll}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - \frac{1}{\lambda+1} \sum_{i=1}^n x_i - \frac{n}{\lambda+1}$
$\frac{dll}{d\lambda} = 0 \to \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} = 5$
По методу моментов
$M_1 = \mathbb{X} = \lambda$
$M_1^* = \bar{X}$
$\hat{\lambda} = \bar{X}$
Смещение оценки
$\mathbb{E}(\hat{\lambda}) = \mathbb{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(x_i) = \frac{1}{n} * n * \lambda = \lambda$
Оценка несмещенная ☐

Задача 9. *Построить асимптотический доверительный интервал уровня значимости α_1 = 0.05 для параметра λ на базе оценки максимального правдоподобия.*

Решение. $\frac{d^2ll}{d\lambda^2} = \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i + \frac{1}{(\lambda+1)^2}\sum_{i=1}^{n} x_i + \frac{n}{(\lambda+1)^2}$
$\hat{I} = -\frac{d^2ll}{d\lambda^2}(\hat{\lambda}) = -\frac{d^2ll}{d\lambda^2}(\bar{X}) = n(\frac{1}{\bar{X}} - \frac{1}{\bar{X}+1}) = 1.67$
$\sigma^2(\hat{\lambda}) = \hat{I}^{-1} = 0.6$
$\sigma = \sqrt{\hat{I}^{-1}} = 0.775$
Доверительный интервал будет иметь вид
$[\hat{\lambda} - x_\alpha\sigma, \hat{\lambda} + x_\alpha\sigma]$
$x_\alpha = \Phi^{-1}(1 - \frac{\alpha_1}{2}) = 1.98$
Получен доверительный интервал [3.467, 6.535] □

Задача 10. *Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением Пуассона с параметром λ_0 = 6.00. Проверить гипотезу на уровне значимости α_1= 0.05. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение. Составим таблицу

x_i	m_i	p_i	np_i	$(\frac{m_i-np_i}{\sqrt{m_i}})^2$
1	4	0.12244898	6.122449	0.7357823
2	6	0.10495627	5.247813	0.1078134
3	3	0.08996252	4.498126	0.4989591
4	9	0.07711073	3.855536	6.8642864
5	9	0.06609491	3.304745	9.8149538
6	6	0.05665278	2.832639	3.5416359
7	6	0.04855953	2.427976	5.2551393
8	1	0.04162245	2.081123	0.5616324
9	2	0.03567639	1.783819	0.0261989
10	4	0.03057976	1.528988	3.9934261

Полученное значение $\chi^2 = 31.3998$
Число степеней свободы $l = k - r - 1 = 8$
$\chi^2_{critical} = 15.5$
Следовательно, H_0 отвергается. Наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу: $1 - \chi^2_{r=8}(31.3998) = 0.000119$

□

Задача 11. *Построить критерий значимости χ^2 проверки сложной гипотезы согласия с геометрическим распределением. Проверить гипотезу на уровне значимости α_1= 0.05. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение. Оценим неизвестный параметр λ как $\mathbb{E}(X_1) = \bar{X} = 5$
Построим таблицу

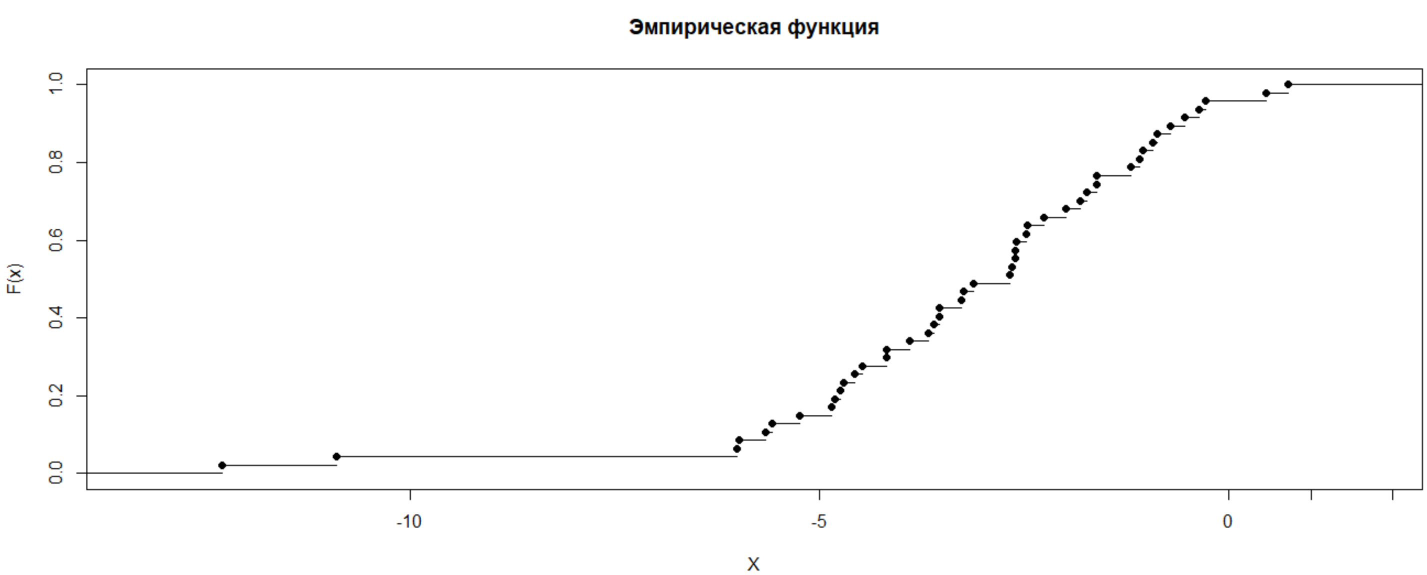
x_i	m_i	p_i	np_i	$(\frac{m_i - np_i}{\sqrt{m_i}})^2$
1	4	0.13888889	6.944444	1.248444444
2	6	0.11574074	5.787037	0.007837037
3	3	0.09645062	4.822531	0.688770864
4	9	0.08037551	4.018776	6.174167720
5	9	0.06697960	3.348980	9.535450167
6	6	0.05581633	2.790816	3.690267352
7	6	0.04651361	2.325680	5.805021450
8	1	0.03876134	1.938067	0.454045030
9	2	0.03230112	1.615056	0.091750398
10	4	0.02691760	1.345880	5.234013788

Полученное значение $\chi^2 = 32.92977$
Число степеней свободы $l = k - r - 1 = 8$
$\chi^2_{critical} = 15.5$
Следовательно, H_0 отвергается. Наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу: $1 - \chi^2_{r=8}(32.92977) = 0.00006$

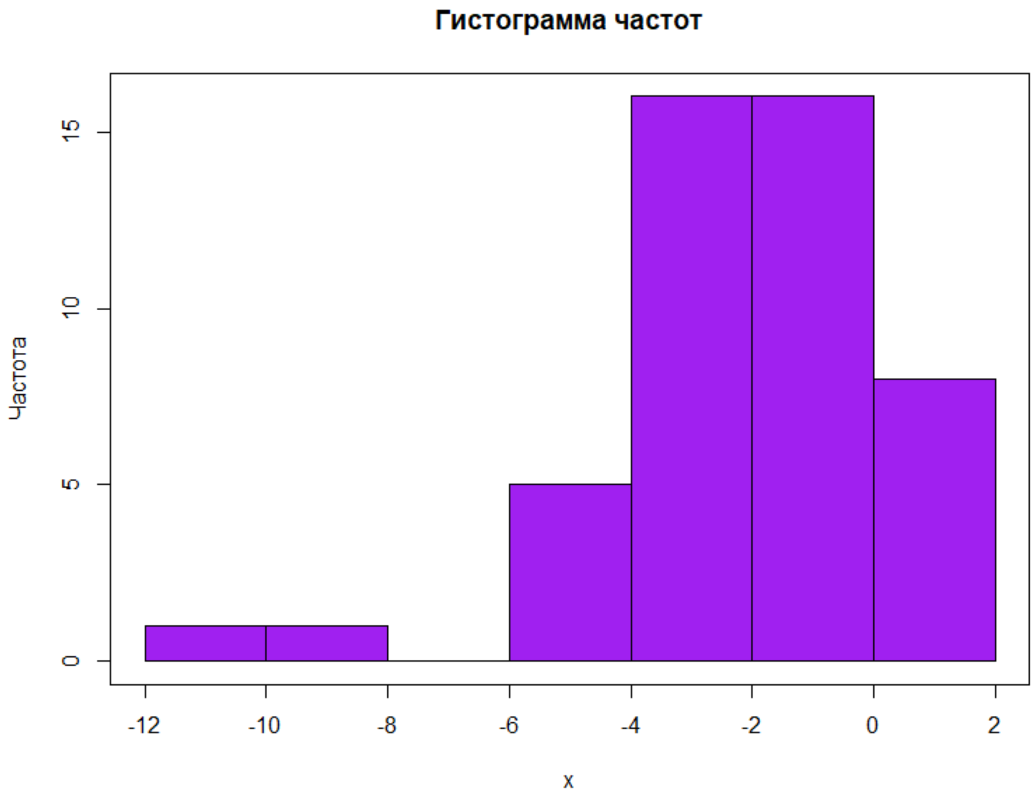
□

Задача 12. *Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.*

Решение. Вариационный ряд
-12.294 -10.888 -6.009 -5.980 -5.660 -5.583 -5.245 -4.847 -4.815 -4.743 -4.708 -4.575 -4.470 -4.180 -4.178 -3.894 -3.665 -3.602 -3.538 -3.529 -3.267 -3.240 -3.116 -2.680 -2.648 -2.612 -2.608 -2.589 -2.470 -2.464 -2.259 -1.989 -1.820 -1.739 -1.620 -1.618 -1.196 -1.085 -1.056 -0.934 -0.875 -0.722 -0.546 -0.360 -0.290 0.452 0.717
Эмпирическая функция распределения



Гистограмма частот



$\square$

Задача 13. *Вычислить выборочные аналоги числовых характеристик*

Решение. Выборочное среднее
$\bar{x} = \frac{1}{n} * \sum_{i=1}^{n}(x_i) = -3.2136$
Выборочная дисперсия
$s^2 = \frac{1}{n} * \sum_{i=1}^{n}(x_i - \bar{x})^2 = 6.2164$
Медиана
$Me = -2.68$
Ассиметрия
$As = \frac{\mu_3^*}{s^3} = -1.4257$
Эксцесс
$Ex = \frac{\mu_4^*}{s^4} = 6.2177$
Вероятность попадания в заданный промежуток
$P(X \in [3.66; 5.45]) = F(b) - F(a) = 0.4043$ $\square$

Задача 14. *В предположении, что исходные наблюдения являются выборкой из нормального распределения, построить оценку максимального правдоподобия параметров* (a, σ^2), *и соответствующие оценки по методу моментов.*

Решение. $l(\overrightarrow{x}, a, \sigma^2) = \prod_{i=1}^{n}(\frac{1}{\sigma\sqrt{2\pi}} exp(\frac{-(x_i-a)^2}{2\sigma^2})) = \sigma^{-n}(2\pi)^{-\frac{n}{2}} exp(-\sum_{i=1}^{n} \frac{-(x_i-a)^2}{2\sigma^2})$
$ll(\overrightarrow{x}, a, \sigma^2) = -nln(\sigma) - \frac{n}{2}ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2$
$\begin{cases} \frac{dll(\overrightarrow{x},a,\sigma^2)}{da} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = 0 \\ \frac{dll(\overrightarrow{x},a,\sigma^2)}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - a)^2 = 0 \end{cases}$

□