Студент: Переверзев Дмитрий
Группа: 8383
Вариант: 18
Дата: 23 декабря 2020 г.

Статистический анализ

Индивидуальное домашнее задание №2

Часть 1

В результате эксперимента получены данные.

3	3	3	5	4	2	0	4	4	1	2	3	1	2	1	3	3	2	4	3	5	4	2	3	3
1	4	3	2	4	6	3	3	4	4	7	1	3	1	1	1	3	3	2	6	5	1	3	2	6

$\alpha_1 = 0.01$
$a = 1.27$
$b = 3.69$
$\lambda_0 = 4.00$
$\lambda_1 = 3.00$

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

Решение.

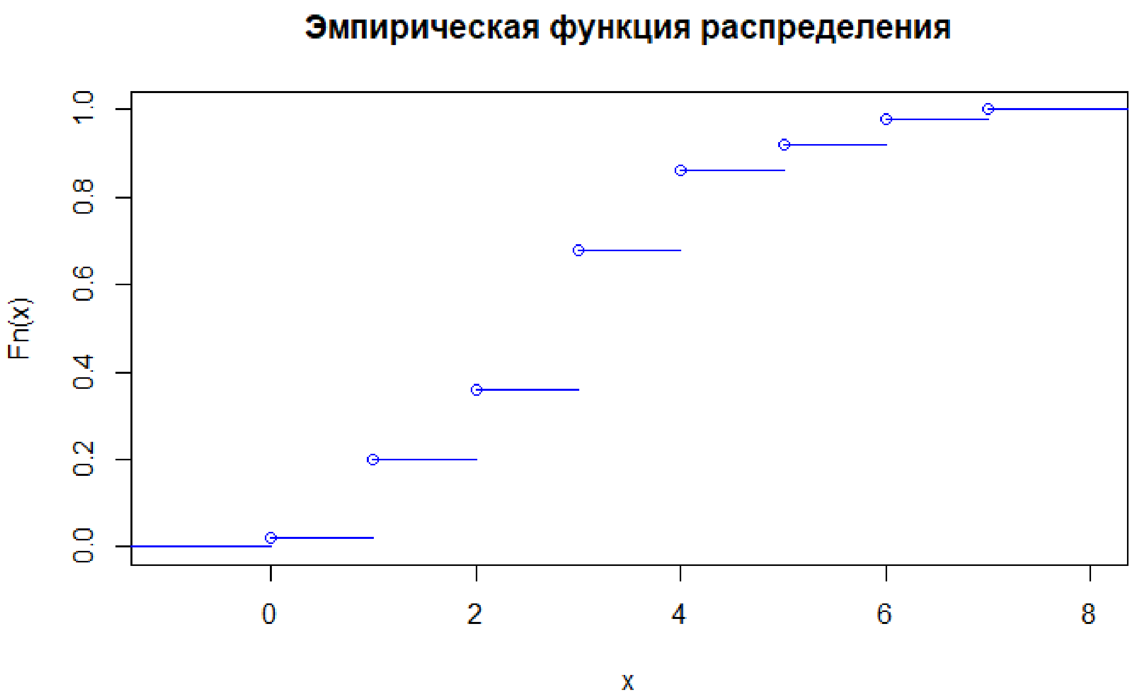
Вариационный ряд:
0 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 5 5 6 6 6 7

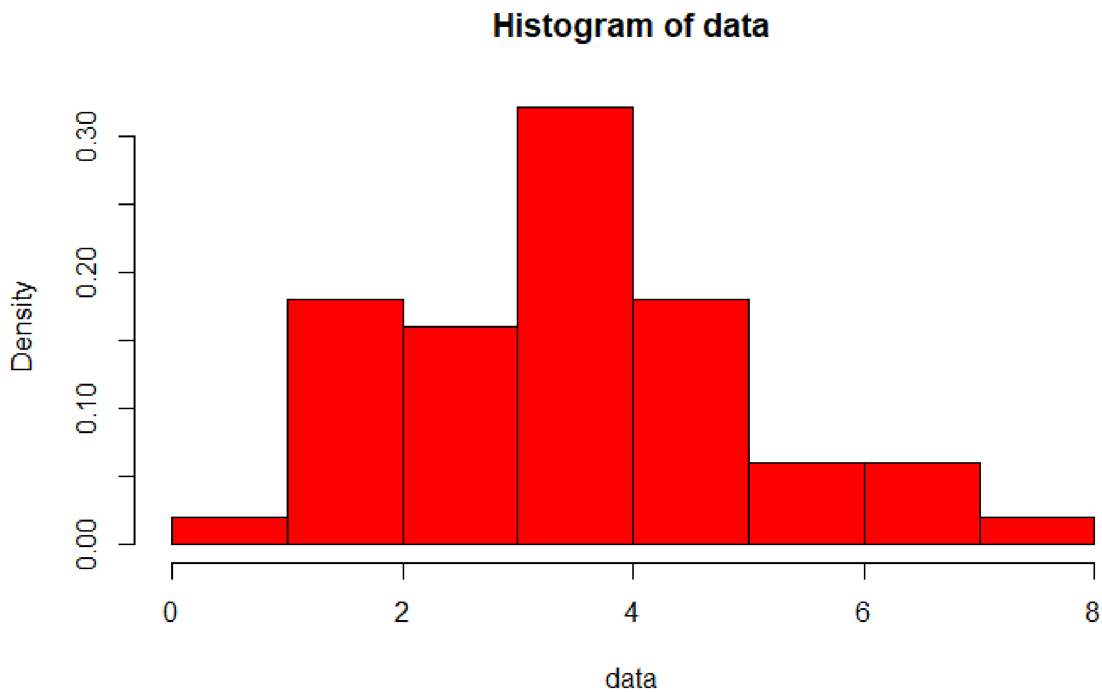
Построим таблицу частот для выборки.

x_j	0	1	2	3	4	5	6	7
m_j	1	9	8	16	9	3	3	1
p_j^*	0.02	0.18	0.16	0.32	0.18	0.06	0.06	0.02

Построим эмпирическую функцию распределения по полученным данным:

$$F(x) = \begin{cases} 0, x \leqslant 0 \\ 0.02, 0 < x \leqslant 1 \\ 0.20, 1 < x \leqslant 2 \\ 0.36, 2 < x \leqslant 3 \\ 0.68, 3 < x \leqslant 4 \\ 0.86, 4 < x \leqslant 5 \\ 0.92, 5 < x \leqslant 6 \\ 0.98, 6 < x \leqslant 7 \\ 1, x > 7 \end{cases}$$ □





Задача 2. *Вычислить выборочные аналоги характеристик:*

Решение.

(1) Математическое ожидание:

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n} x_i m_i = 2.98$$

(2) Дисперсия:

$$D_{\text{в}} = \bar{x^2} - \bar{x}^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = 2.3396$$

(3) Медиана:

$$Me = 3$$

(4) Ассиметрия:

$$As = \frac{\mu_3^*}{\sigma^3} = 0.4230174$$

(5) Эксцесс:

$$Ex = \frac{\mu_4^*}{\sigma^4} - 3 = -0.09535895$$

(6) Вероятность:

$$P(x \in [a,b]) = P(x \in [1.27, 3.69]) = F(3.69) - F(1.27) = 0.48$$

$\square$

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.

Распределение Пуассона: $P_\lambda = (x = k) = \dfrac{\lambda^k}{k!} \cdot \exp(-\lambda)$

(1) Метод максимального правдоподобия

$$l(x,\lambda) = \frac{\lambda^{\sum\limits_1^n x_i} \cdot \exp(-\lambda n)}{\prod\limits_1^n x_i!} \Rightarrow ll(\bar{x},\lambda) = \sum_1^n x_i ln\lambda - n\lambda - \sum_1^n ln x_i! \Rightarrow$$

$$\frac{\partial ll(\bar{x},\lambda)}{\partial \lambda} = \frac{1}{\lambda}\sum_1^n x_i - n = 0 \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_1^n x_i = \bar{x_в} = 2.98$$

(2) Метод моментов

$P(x,\theta)$

$M_1^* = \bar{x_в}; \mathbb{E}X = \bar{x_в};$

$\mathbb{E}X = \int_{\mathbb{R}} xp(x,\theta)dx = \varphi(\theta)$

$M_1 = \mathbb{E}X = \lambda; M_1^* = \bar{x_в} \Rightarrow \hat{\lambda} = \bar{x_в} = 2.98$

Чтобы найти смещение оценки, найдем:

$\mathbb{E}\hat{\lambda} = \dfrac{1}{n}\sum\limits_1^n \mathbb{E}x_i = \dfrac{n\lambda}{n} = \lambda \Rightarrow$ оценки несмещенные $\square$

Задача 4. *Построить асимптотический доверительный интервал уровня значимости α_1 для параметра λ на базе оценки максимального правдоподобия.*

Решение.

$\hat{\lambda} = \bar{x_в} = 2.98; \alpha_1 = 0.01; \gamma = 1 - \alpha_1 = 0.99;$

$\dfrac{\sqrt{n}(\bar{x} - \lambda)}{\sqrt{\lambda}} \underset{n\to\infty}{\longrightarrow} N(0,1)$

$t_\gamma : \phi(t_\gamma) = 1 - \dfrac{\alpha_1}{2} \Rightarrow t_\gamma = 2.575829$

$P(-t_\gamma \leqslant \dfrac{\sqrt{n}(\bar{x_в} - \lambda)}{\sqrt{\lambda}} \leqslant t_\gamma) \longrightarrow 1 - \alpha$

$n(\bar{x_в} - \lambda)^2 = t_\gamma^2 \lambda$

$\lambda^2 - 2\lambda(\bar{x_в} + \dfrac{t_\gamma^2}{2n}) + \bar{x_в} = 0$

$\lambda_{1,2} = \bar{x_в} + \dfrac{t_\gamma^2}{2n} \pm t_\gamma \sqrt{\dfrac{1}{n}(\bar{x_в} + \dfrac{t_\gamma^2}{4n})} \Rightarrow [2.35116; 3.60884]$ $\square$

Задача 5. *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.*

Решение. $\hat{\lambda_0} = 4.00; \alpha_1 = 0.01;$
Простая гипотеза H_0 имеет вид:
$H_0 : p(x) = \dfrac{\lambda_0^x}{x!}\exp(-\lambda_0)$
$p_k = P(X = k) = \dfrac{0.7^k}{k!}\exp(-0.7)$
Построим таблицу оценки методом χ^2.

x_i	0	1	2	3	4	5	6	7	$\sum$
m_i	1	9	8	16	9	3	3	1	50
p_i	0.018	0.073	0.147	0.195	0.195	0.156	0.104	0.111	1
$m_i^{'}$	0.916	3.663	7.326	9.768	9.768	7.815	5.210	5.534	50
$m_i - m_i^{'}$	0.084	5.337	0.674	6.232	-0.768	-4.815	-2.210	-4.534	0
$\dfrac{(m_i - m_i^{'})^2}{m_i^{'}}$	0.008	7.775	0.062	3.975	0.060	2.966	0.937	3.714	$\chi^2_{\text{набл}}$

$$\chi^2_{\text{набл}} = \sum_1^k \frac{(m_i - m_i^{'})^2}{m_i^{'}} = 19.49903$$
$$l = k - r - 1 = 8 - 1 - 1 = 6$$
$$\chi^2_{\text{кр}} = \chi^2_{\alpha;l} = \chi^2_6 = 16.81189$$

$\chi^2_{\text{набл}} > \chi^2_{\text{кр}} \Rightarrow$ гипотеза H_0 отвергается.
Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = 0.003398824 □

Задача 6. *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.*

Решение.

Сложная гипотеза H_0 имеет вид:

$$H_0 : x_1,...,x_n \sim P_{ois}(\lambda)$$
$$\sum_1^k \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)} \longrightarrow \chi^2_{k-r-1}$$

Метод минимизации хи-квадрат:

$$\underset{\lambda}{argmin} \sum_{1}^{r} \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)}$$

Задача реализована в R следующим скриптом:

$P < -function(a)\{$

$p < -0$

$p[1] < -ppois(0, a)$

$p[2] < -ppois(1, a) - sum(p)$

$p[3] < -ppois(2, a) - sum(p)$

$p[4] < -ppois(3, a) - sum(p)$

$p[5] < -ppois(4, a) - sum(p)$

$p[6] < -ppois(5, a) - sum(p)$

$p[7] < -ppois(6, a) - sum(p)$

$p[8] < -1 - sum(p)$

$p\}; X2 < -function(a)\{g < -n * P(a); f < -(nu - g)^2/g; sum(f)\}$

$nu < -c(1{,}9{,}8{,}16{,}9{,}3{,}3{,}1); XM < -nlm(X2, lambda0)$

В результате вычислений получим, что $\chi^2_{\text{набл}} = 5.400952 < \chi^2_{\text{крит}} = 15.08627$ Таким образом, гипотеза принимается.

Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = 0.3689291 $\square$

Задача 7. *Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром $\lambda = \lambda_0 = 4.00$ при альтернативе пуассоновости с параметром $\lambda = \lambda_1 = 3.00$. Проверить гипотезу на уровне значимости α_1. Что получится, если поменять местами основную и альтернативную гипотезы?*

Решение.

Сформулируем гипотезы.

$H_0 : \lambda = \lambda_0 = 4.00$

$H_1 : \lambda = \lambda_1 = 3.00$

По лемме Неймана-Пирсона:

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

$$l(\bar{x}, 4, 3) = \frac{L(\bar{x}, 3)}{L(\bar{x}, 4)} = 0.75^{\sum x_i} \cdot \exp(n * (\lambda_0 - \lambda_1)) = 0.75^{\sum x_i} \cdot \exp(n)$$

$$ll(\bar{x}, \lambda_0, \lambda_1) = \sum x_i \cdot ln0.75 + n < lnC; \sum x_i < \frac{lnC - n}{ln0.75}; \hat{C} = \frac{-lnC - n}{ln0.75}$$

Критерий принимает вид:

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

Вычислим $\hat{C}$, p и α_0:

$P_{\lambda_0}(l(\bar{x}, \lambda_0, \lambda_1) > C) + p \cdot P_{\lambda_0}(l(\bar{x}, \lambda_0, \lambda_1) = C) =$

$= P_{\lambda_0}(\sum\limits_1^n x_i > \hat{C}) + p \cdot P_{\lambda_0}(\sum\limits_1^n x_i = \hat{C}) = \alpha_1 = 0.01$

$x_i \to P_{ois}(\lambda_0); \sum x_i \to P_{ois}(n\lambda_0)$

$\alpha_0 = P_{\lambda_0}(\sum\limits_1^n x_i > \hat{C}) = 1 - P_{n\lambda_0}(\hat{C}) - p_{n\lambda_0}(\hat{C}) < \alpha_1$

$$p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_0}(\sum\limits_1^n x_i = A)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_0}(A)}$$

В результате расчета получим: $\alpha_0 = 0.009888703$; $\hat{C} = 232$; $p = 0.04845126$

$\sum\limits_1^n x_i = 149$

$149 < 232 \Rightarrow$ Таким образом, отвергаем гипотезуH_0

Теперь поменяем местами основную и альтернативную гипотезы.

$H_0 : \lambda = \lambda_1 = 3.00$

$H_1 : \lambda = \lambda_0 = 4.00$

$l(\bar{x}, 3, 4) = \dfrac{L(\bar{x}, 4)}{L(\bar{x}, 3)} = (\frac{4}{3})^{\sum x_i} \cdot \text{exp}(n * (\lambda_1 - \lambda_0)) = (\frac{4}{3})^{\sum x_i} \cdot \text{exp}(-n)$

$$ll(\bar{x}, \lambda_0, \lambda_1) = -\sum x_i \cdot ln(\tfrac{4}{3}) - n < lnC; \sum x_i > \frac{-lnC - n}{ln(\frac{4}{3})}; \hat{C} = \frac{-lnC - n}{ln(\frac{4}{3})}$$

Тогда критерий примет вид:

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

Вычислим $\hat{C}$ и p.

В результате расчета получим: $\alpha_0 = 0.008352359$; $\hat{C} = 121$; $p = 0.9165829$

$\sum\limits_1^n x_i = 149$

$149 > 121 \Rightarrow$ Таким образом, принимаем гипотезуH_0

При замене меняется гипотеза, которая принимается, но так как изменение происходит со сменой гипотез местами, решение не меняется. □

Задача 8. *В пунктах (c) - (f) заменить семейство распределений Пуассона на семейство геометрических распределений.*

Решение.

$$P_\lambda(X=k) = \frac{\lambda^k}{(\lambda+1)^{k+1}}, k = 0, 1, ...$$

□

Задача 9. *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:

$$P_\lambda(X=k) = \frac{\lambda^k}{(\lambda+1)^{k+1}}$$

(1) Метод максимального правдоподобия

$$l(\bar{x},\lambda) = \prod_1^n \frac{\lambda^{x_i}}{(\lambda+1)^{x_i+1}} = \frac{\lambda^{\sum_1^n x_i}}{(\lambda+1)^{\sum_1^n x_i + n}}$$

$$ll(\bar{x},\lambda) = ln\lambda \cdot \sum_1^n x_i - ln(\lambda+1)\sum_1^n x_i - nln(\lambda+1)$$

$$\frac{\partial ll}{\partial \lambda} = \frac{1}{\lambda}\sum_1^n x_i - \frac{1}{\lambda+1}\sum_1^n x_i - \frac{n}{\lambda+1}$$

$$\frac{\partial ll}{\partial \lambda} = 0 \to \hat{\lambda} = \frac{1}{n}\sum_1^n x_i = \bar{x} = 2.98$$

(2) Метод моментов

$$M_1 = \mathbb{X} = \lambda; M_1^* = \hat{X}; \hat{\lambda} = \bar{X}$$

Чтобы найти смещение оценки, найдем:

$\mathbb{E}\hat{\lambda} = \mathbb{E}\hat{X} = \frac{1}{n}\sum_1^n \mathbb{E}(x_i) = \frac{n\lambda}{n} = \lambda \Rightarrow$ оценки несмещенные □

Задача 10. *Построить асимптотический доверительный интервал уровня значимости $\alpha_1 = 0.10$ для параметра λ на базе оценки максимального правдоподобия.*

Решение.

$$\frac{\partial^2 ll}{\partial \lambda^2} = \frac{1}{\lambda^2}\sum_1^n x_i + \frac{1}{(\lambda+1)^2}\sum_1^n x_i + \frac{n}{(\lambda+1)^2}$$

$$\hat{I} = -\frac{\partial^2 ll}{\partial \lambda^2}(\hat{\lambda}) = -\frac{\partial^2 ll}{\partial \lambda^2}(\hat{X}) = n(\frac{1}{\bar{X}} - \frac{1}{\bar{X}+1}) = 4.215709$$

$\sigma^2(\hat\lambda) = \hat I^{-1} = 0.237208; \sigma = \sqrt{\hat I^{-1}} = 0.48704$
$[\hat\lambda - x_\alpha\sigma, \hat\lambda + x_\alpha\sigma]$
$x_\alpha = \phi^{-1}(1 - \frac{\alpha_1}{2}) = 2.575829$
Получен доверительный интервал $[1.725468, 4.234532]$ □

Задача 11. *Используя гистограмму частот, построить критерий значимости* χ^2 *проверки простой гипотезы согласия с геометрическим распределением с параметром* $\lambda_0 = 4.00$. *Проверить гипотезу на уровне значимости* $\alpha_1 = 0.01$. *Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение.

$$H_0 : X_1, ..., X_n \sim Geom\left(\frac{1}{4+1}\right) = Geom\left(\frac{1}{5}\right)$$

Построим таблицу оценки методом χ^2.

x_i	0	1	2	3	4	5	6	7	$\sum$
m_i	1	9	8	16	9	3	3	1	50
p_i	0.200	0.160	0.128	0.102	0.082	0.066	0.052	0.210	1
m_i'	10.000	8.000	6.400	5.120	4.096	3.277	2.621	10.486	50
$m_i - m_i'$	-9.000	1.000	1.600	10.880	4.904	-0.277	0.379	-9.486	0
$\frac{(m_i - m_i')^2}{m_i'}$	8.100	0.125	0.400	23.120	5.871	0.023	0.055	8.581	$\chi^2_{\text{набл}}$

$$\chi^2_{\text{набл}} = \sum_1^k \frac{(m_i - m_i')^2}{m_i'} = 46.27557$$
$$\chi^2_{\text{кр}} = \chi^2_{\alpha;l} = \chi^2_6 = 16.81189$$

$\chi^2_{\text{набл}} > \chi^2_{\text{кр}} \Rightarrow$ гипотеза H_0 отвергается. Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = $2.609142e - 08$ □

Задача 12. *Построить критерий значимости* χ^2 *проверки сложной гипотезы согласия с геометрическим распределением. Проверить гипотезу на уровне значимости* $\alpha_1 = 0.01$. *Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение.

Сложная гипотеза H_0 имеет вид:

$$H_0 : X_1, ..., X_n \sim Geom\left(\frac{1}{1+\lambda}\right)$$
$$\sum_1^r \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)} \longrightarrow \chi^2_{k-r-1}$$

Метод минимизации хи-квадрат:

$$\underset{\lambda}{argmin} \sum_{1}^{k} \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)}$$

Получили $\hat{\lambda} = \dfrac{1}{0.2423235} - 1 = 3.126715$

Построим таблицу оценки методом χ^2.

x_i	0	1	2	3	4	5	6	7	$\sum$
m_i	1	9	8	16	9	3	3	1	50
p_i	0.200	0.160	0.128	0.102	0.082	0.066	0.052	0.210	1
$m_i^{'}$	10.000	8.000	6.400	5.120	4.096	3.277	2.621	10.486	50
$m_i - m_i^{'}$	-9.000	1.000	1.600	10.880	4.904	-0.277	0.379	-9.486	0
$\frac{(m_i - m_i^{'})^2}{m_i^{'}}$	8.100	0.125	0.400	23.120	5.871	0.023	0.055	8.581	$\chi^2_{\text{набл}}$

В результате вычислений получим, что $\chi^2_{\text{набл}} = 44.00928 > \chi^2_{\text{крит}} = 15.08627$

Гипотеза отвергается.

Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = $2.306196e - 08$ $\square$

Часть 2

В результате эксперимента получены данные.

-1.14	-7.70	18.88	-10.33	9.50	9.65	-6.23	-5.60	8.63	8.04
-6.12	33.22	7.95	-5.88	4.92	5.83	-8.09	-8.29	2.65	11.10
3.83	-7.62	-3.25	2.24	-3.21	6.49	15.71	0.72	1.46	17.58
9.03	1.24	12.08	-0.01	18.55	31.56	2.87	2.81	-4.75	-13.22
-14.73	2.96	6.28	4.66	10.70	3.77	12.44	7.18	-2.04	12.55

$\alpha_2 = 0.10$
$c = 0.00$
$d = 14.00$
$h = 4.00$
$a_0 = -7.00$
$\sigma_0 = 10.00$
$a_1 = 4.00$
$\sigma_1 = 10.00$

Задача 13. *Построить вариационный ряд, эмпирическую функцию распределения, гистограмму и полигон частот с шагом h.*

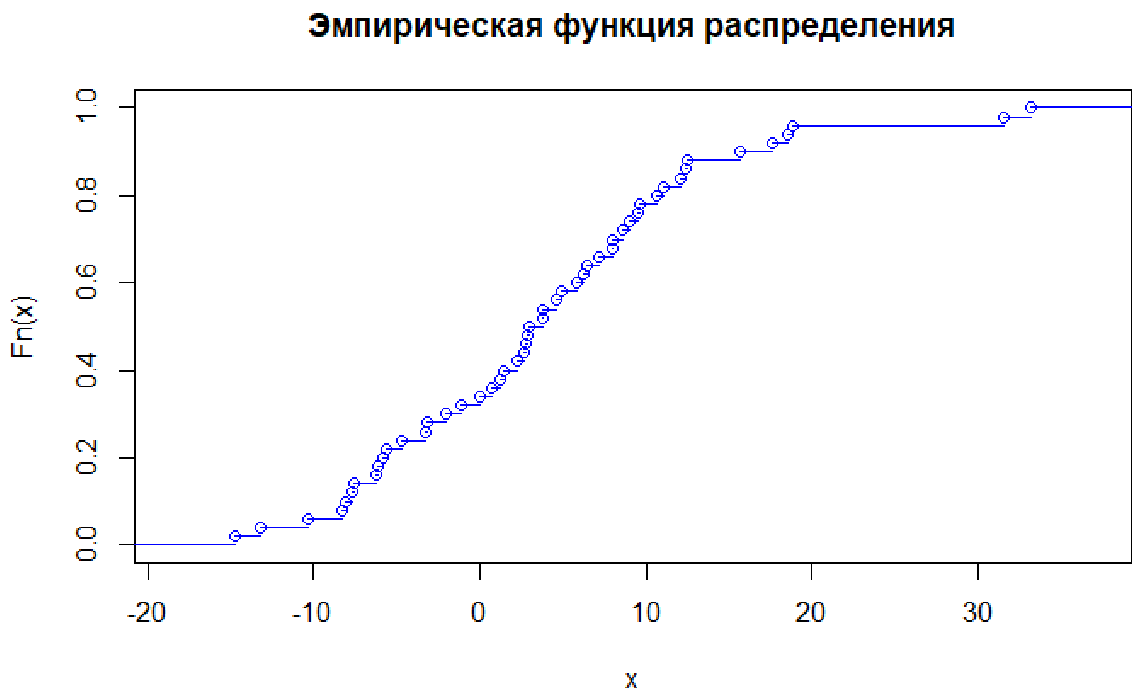
Решение.

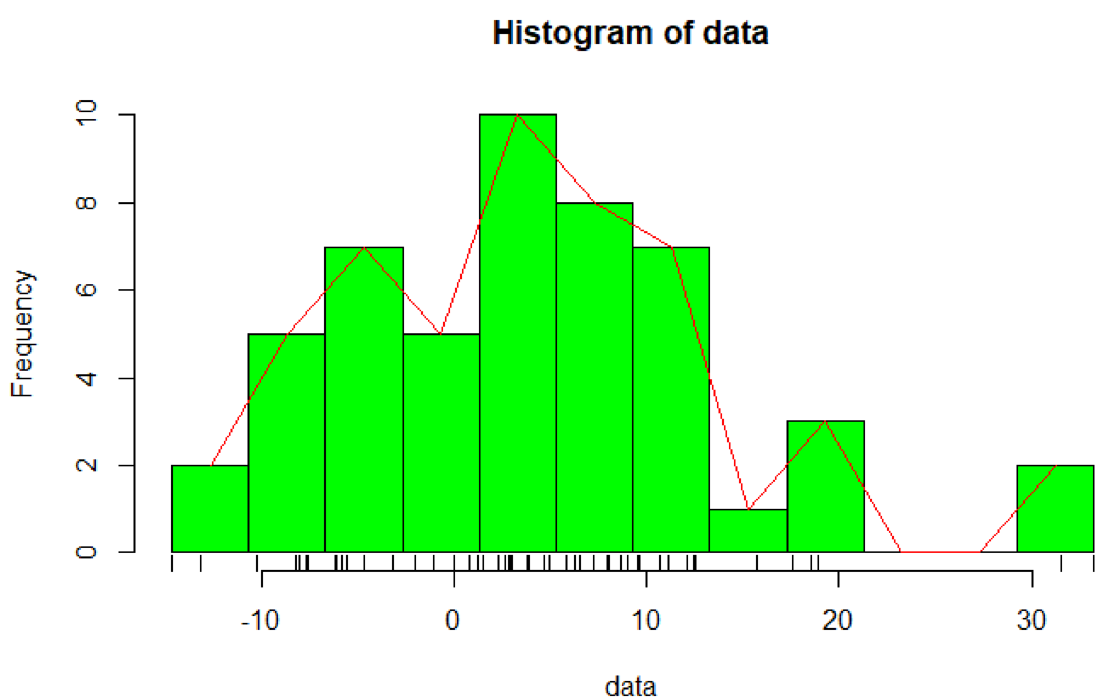
Вариационный ряд:

-14.73 -13.22 -10.33 -8.29 -8.09 -7.70 -7.62 -6.23 -6.12 -5.88 -5.60 -4.75 -3.25 -3.21 -2.04 -1.14 -0.01 0.72 1.24 1.46 2.24 2.65 2.81 2.87 2.96 3.77 3.83 4.66 4.92 5.83 6.28 6.49 7.18 7.95 8.04 8.63 9.03 9.50 9.65 10.70 11.10 12.08 12.44 12.55 15.71 17.58 18.55 18.88 31.56 33.22

На следующих рисунках представлены эмпирическая функция распределения, гистограмма и полигон частот.

□





Задача 14. *Вычислить выборочные аналоги характеристик:*

Решение.

(1) Математическое ожидание:

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n}(x_i) = 3.9774$$

(2) Дисперсия:

$$D_{\text{в}} = \bar{x^2} - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 99.04491$$

(3) Медиана:

$$Me = 3.365$$

(4) Ассиметрия:

$$As = \frac{\mu_3^*}{\sigma^3} = 0.6413363$$

(5) Эксцесс:

$$Ex = \frac{\mu_4^*}{\sigma^4} - 3 = 0.8044914$$

(6) Вероятность:

$$P(x \in [c, d]) = P(x \in [0.00, 14.00]) = F(14.00) - F(0.00) = 0.54$$

□

Задача 15. *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.*

Решение.

Плотность нормального распределения:

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

(1) Метод максимального правдоподобия

$$L(\vec{x}, a, \sigma^2) = \frac{1}{\sigma^n\sqrt{(2\pi)^n}} \cdot \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2)$$

$$LL(\vec{x}, a, \sigma^2) = \frac{n}{2}log2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2$$

$$\begin{cases} \dfrac{\partial LL(\vec{x}, a, \sigma^2)}{\partial a} = \dfrac{1}{\sigma^2}\left(\sum\limits_{i=1}^{n} x_i - na\right) = 0 \\ \dfrac{\partial LL(\vec{x}, a, \sigma^2)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\sum\limits_{i=1}^{n}(x_i - a)^2 = 0 \end{cases}$$

$$\begin{cases} \hat{a} = \dfrac{\sum\limits_{i=1}^{n} x_i}{n} = \bar{x} \\ \hat{\sigma^2} = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 = S^2 \end{cases}$$

Найдем значения $\hat{a} = \bar{x} = 3.9774$ - выборочное среднее и $\hat{\sigma^2} = S^2 = 99.04491$ - выборочная дисперсия.

(2) Метод моментов

В случае нормального распределения имеем $a_1' = \mathbb{E}(x) = a$ и $a_2' = \mathbb{E}(x^2) = \sigma^2 + a^2$

$\tilde{a} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \bar{x} = 3.9774; \tilde{a^2} + \tilde{\sigma^2} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 = \bar{x}^2$
$\tilde{a} = \bar{x} = 3.9774$
$\tilde{\sigma^2} = \bar{x}^2 - \bar{x^2} = \dfrac{(\sum\limits_{i=1}^{n} x_i^2 - \bar{x}\sum\limits_{i=1}^{n} x_i)}{n} = S^2 = 99.04491$
Оценки являются несмещенными.

$\square$

Задача 16. *Построить доверительные интервалы уровня значимости* α_2 *для параметров* (a, σ^2)

Решение.

a:
$\sqrt{n-1}\left(\dfrac{\bar{x}-n}{s}\right) \sim S_{n-1}$

$x_\alpha : S_{n-1}(x_\alpha) = 1 - \dfrac{\alpha_2}{2}$
Получаем:

$$P\left(-x_\alpha \leqslant \sqrt{n-1}\left(\frac{\bar{x}-n}{s}\right) \leqslant x_\alpha\right) = 1 - \alpha_2 = P\left(\bar{x} - \frac{x_\alpha s}{\sqrt{n-1}} \leqslant a \leqslant \bar{x} + \frac{x_\alpha s}{\sqrt{n-1}}\right)$$

ДИ уровня значимости $\alpha_2 = 0.10$ для a:

$$[1.638857; 6.315943]$$

σ^2:
$\dfrac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$

Выберем $x_{1\alpha}, x_{2\alpha}$ - квантили распределения χ^2_{n-1} уровня $\dfrac{\alpha_2}{2}$ и $1 - \dfrac{\alpha_2}{2}$
$p\left(\dfrac{nS^2}{x_{1\alpha}} \leqslant \sigma^2 \leqslant \dfrac{nS^2}{x_{2\alpha}}\right) = 1 - \alpha_2$

ДИ уровня значимости $\alpha_2 = 0.10$ для σ^2:

$$[74.65098; 145.9534]$$

□

Задача 17. *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.*

Решение.

Простая гипотеза $H_0 : a = a_0 = -7.00, \sigma^2 = \sigma_0^2 = 10.00$

Согласно теореме Колмогорова:

$\sqrt{n} sup|F_n(x) - F_0(x)| \to K(x)$, где $K(x)$ - функция Колмогорова

$K(C_{\alpha_2}) = 1 - \alpha_2 = 1 - 0.1 = 0.9$
$C_{\alpha_2} = C_{0.2} = 1.1992$
$H_0 : X_1, ..., X_n \sim N(-7.00, 10.00)$

С помощью R вычислим:

$sup|F_n(x) - F_0(x)| = 0.4199428$
$\sqrt{n} sup|F_n(x) - F_0(x)| = 2.969444 > 1.1992$

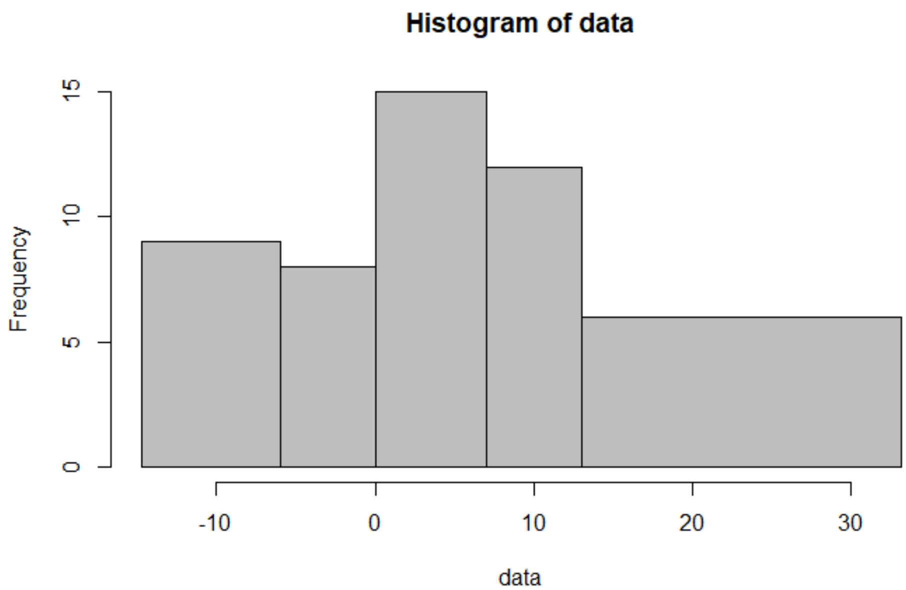
Отвергаем гипотезу H_0

$p - value = 1.8 \cdot 10^{-14}$
Наибольшее значение уровня значимости, на котором нет оснований отвергнуть гипотезу □

Задача 18. *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.*

Решение.

$H_0 : X_1, ..., X_n \sim N(-7.00, 10.00)$
$\sum_{i=1}^{k} \frac{(m_i - m_i')^2}{m_i} \to \chi^2_{\text{кр}}$
$\chi^2_{\text{кр}} = \chi^2_4 = 7.77944$
Перестроим гистограмму частот, выбрав следующие точки: $(-Inf, -6, 0, 7, 13, Inf)$



Интервал	$(-\infty;-6]$	$(-6;0]$	$(0;7]$	$(7;13]$	$(13;\infty)$	$\sum$
m_i	7	10	15	12	6	50
p_i	0.53982784	0.21820851	0.16120699	0.05800653	0.02275013	1
$\dfrac{(m_i-m_i^{'})^2}{m_i^{'}}$	14.80678546	0.07597088	5.97477145	28.54991124	20.78567469	$\chi^2_{\text{набл}}$

$\chi^2_{\text{набл}} = 70.19311 > \chi^2_{\text{кр}} = 7.77944$

$P(\chi^2_{\text{набл}} > 70.19311) = 2.065015e - 14$

Гипотеза отвергается, наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть гипотезу равно 2.065015e-14. $\square$

Задача 19. *Построить критерий значимости χ^2 проверки 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.*

Решение.

$H_0 : X_1, ..., X_n \sim N(a\sigma^2)$

$$\sum_{i=1}^{k} \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \to \chi^2_{\text{кр}}$$

Метод минимизации хи-квадрат:

$$\underset{a,\sigma^2}{argmin} \sum_{i=1}^{k} \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \to \chi^2_{\text{кр}}$$

Задача реализована в R с помощью скрипта:

$P < -function(a)\{$

$p < -0$

$p[1] < -pnorm(-6, a[1], a[2])$

$p[2] < -pnorm(0, a[1], a[2]) - sum(p)$

$p[3] < -pnorm(7, a[1], a[2]) - sum(p)$

$p[4] < -pnorm(13, a[1], a[2]) - sum(p)$

$p[5] < -1 - sum(p)$

$p\}$

$X2 < -function(a)\{g < -n * P(a); f < -(nu - g)^2/g; sum(f)\}$

$nu < -c(7,10,15,12,6)$

$a < -c(mean(data), sqrt(var(data)))$

$XM < -nlm(X2, a)$

$\chi^2_{\text{набл}} = 0.3942813, a = 3.491842, \sigma = 8.511311$

$\chi^2_{\text{набл}} < \chi^2_{\text{кр}} = 4.60517$

Сложная гипотеза согласия с нормальным распределением принимается.

$P(x > 0.3942813) = 0.8210752$

Наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть гипотезу равно 0.8210752 □

Задача 20. *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?

Решение.

Отношение правдоподобия ($\sigma_0 = \sigma_1$):

$l(x) = \dfrac{L(x, a_1, \sigma_1^2)}{L(x, a_0, \sigma_0^2)} = \dfrac{\sigma_0^n}{\sigma_1^n} \exp\left(\dfrac{1}{2\sigma_0^2}\left(\sum\limits_{i=1}^n (x_i - a_0)^2 - \sum\limits_{i=1}^n (x_i - a_1)^2\right)\right) =$

$= \exp\left(\dfrac{1}{2\sigma_0^2}\sum\limits_{i=1}^n -2x_i a_0 + a_0^2 + 2x_i a_1 - a_1^2\right)$

Логарифмируем:

$$\exp\left(\frac{1}{2\sigma_0^2}\sum_{i=1}^n -2x_i a_0 + a_0^2 + 2x_i a_1 - a_1^2\right) > c \to \frac{1}{2\sigma_0^2}\sum_{i=1}^n (a_0^2 - a_1^2) + 2x_i(a_1 - a_0) > logc$$

$\dfrac{n(a_0^2 - a_1^2)}{2\sigma_0^2} + \dfrac{a_1 - a_0}{\sigma_0^2}\sum\limits_{i=1}^n x_i > logc$

$8.25 + 0.11\sum\limits_{i=1}^n x_i > logc \to \sum\limits_{i=1}^n x_i > \dfrac{logc - 8.25}{0.11} = c^*$

Тогда критерий примет вид:

$$\phi(x) = \begin{cases} 1, \sum\limits_{i=1}^n x_i < c^* \\ p, \sum\limits_{i=1}^n x_i = c^* \\ 0, \sum\limits_{i=1}^n x_i > c^* \end{cases} \Rightarrow \phi(x) = \begin{cases} 1, \sum\limits_{i=1}^n x_i < -440 \\ p, \sum\limits_{i=1}^n x_i = -440 \\ 0, \sum\limits_{i=1}^n x_i > -440 \end{cases}$$

$\sum\limits_{i=1}^n x_i = 198 > -440$, принимается альтернативная гипотеза о нормальности с параметром (a_1, σ_1^2).

Если поменять местами основную и альтернативную гипотезы, то будет принята основная гипотеза о нормальности с параметром (a_0, σ_0^2). □

Задача 21. *В пунктах (c) - (g) заменить семейство нормальных распределений на двухпараметрическое семейство распределений Лапласа.*

Решение.

Плотность распределения Лапласа:

$$f(x) = \frac{1}{\sqrt{2}\sigma}\exp(-\frac{\sqrt{2}}{\sigma}|x - a|)$$

□

Задача 22. *В предположении, что исходные наблюдения являются выборкой из распределения семейства Лапласа, построить оценку максимального правдоподобия параметров* (a, σ^2)*, и соответствующие оценки по методу моментов. Найти смещение оценок.*

Решение.

(1) Метод максимального правдоподобия

$$l(\vec{x}, a, \sigma^2) = \frac{1}{\sigma^n 2^{n/2}} \exp(-\frac{\sqrt{2}}{\sigma}\sum_{i=1}^{n}|x_i - a|)$$

$$ll(\vec{x}, a, \sigma^2) = -nlog(\sigma) - \frac{n}{2}log(2) - \frac{\sqrt{2}}{\sigma\sum_{i=1}^{n}}|x_i - a| =$$

$$= -nlog(\sigma) - \frac{n}{2}log(2) - \frac{\sqrt{2}}{\sigma}\sum_{i=1}^{k}(a - x_{(i)}) - \frac{\sqrt{2}}{\sigma}\sum_{i=k+1}^{n}(x_{(i)} - a) =$$

$$= -nlog(\sigma) - \frac{n}{2}log(2) - \frac{\sqrt{2}}{\sigma}ka + \frac{\sqrt{2}}{\sigma}\sum_{i=1}^{k}x_{(i)} - \frac{\sqrt{2}}{\sigma}\sum_{i=k+1}^{n}x_{(i)} + \frac{\sqrt{2}}{\sigma}(n-k-1)a =$$

$$= -nlog(\sigma) - \frac{n}{2}log(2) + \frac{\sqrt{2}}{\sigma}\sum_{i=1}^{k}x_{(i)} - \frac{\sqrt{2}}{\sigma}\sum_{i=k+1}^{n}x_{(i)} + \frac{\sqrt{2}}{\sigma}(n-k-1)a + \frac{\sqrt{2}}{\sigma}(n-2k-1)a$$

$$\frac{\partial ll(\vec{x}, a, \sigma^2)}{\partial a} = \frac{\sqrt{2}}{\sigma}(n-2k-1)$$

$$\frac{\sqrt{2}}{\tilde{\sigma}}(n-2k-1) = 0; k = \frac{n-1}{2} - \text{выборочная медиана}$$

$$\tilde{a} \in (x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}) = (x_{(25)}, x_{(26)}) = (2.96, 3.77) = 3.365$$

$$\frac{\partial ll(\vec{x}, a, \sigma^2)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sqrt{2}}{\sigma^2}\sum_{i=1}^{n}|x_i - a|$$

$$-\frac{n}{\tilde{\sigma}} + \frac{\sqrt{2}}{\tilde{\sigma}^2}\sum_{i=1}^{n}|x_i - a| = 0$$

$$\tilde{\sigma} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{a}|$$

Оценка максимального правдоподобия:

$$(\tilde{a}, \tilde{\sigma}) = (3.365, 58.18333)$$

(2) Метод моментов

$$\mathbb{E}(x) = a \to= \hat{a} = \bar{x} = 3.9774$$

$$\mathbb{D}(x) = \frac{2}{\frac{2}{\sigma^2}} = \sigma^2 \to \hat{\sigma^2} = s^2 = 99.04491$$

$$(\hat{a}, \hat{\sigma^2}) = (\tilde{a}, \tilde{\sigma^2}) = (3.9774, 99.04491)$$

$\mathbb{E}(\hat{a}) = \mathbb{E}(\bar{x}) = a \to$ несмещенная оценка

$\mathbb{E}(\hat{\sigma^2}) = \mathbb{E}(s^2) = \dfrac{n-1}{n}\sigma^2 \to \mathbb{E}(\hat{\sigma^2}) - \sigma^2 = \dfrac{n-1}{n}\sigma^2 - \sigma^2 = -\dfrac{\sigma^2}{n} \to$ смещенная оценка

□

Задача 23. *Построить доверительные интервалы уровня значимости* α_2 *для параметров* (a, σ^2)

Решение.

a:

$$\frac{\sum_{i=1}^{n} x_i - na}{\sqrt{n\sigma^2}} \to N(0,1); \frac{\sum_{i=1}^{n} x_i - na}{\sqrt{ns^2}} \to N(0,1)$$

Выберем $t_\gamma : \phi(t_\gamma) = 1 - \dfrac{\alpha_2}{2} \Rightarrow t_\gamma = 1.644854$

$$P\left(-t_\gamma \leqslant \frac{\sqrt{n}(\bar{x}-a)}{s} \leqslant t_\gamma\right) = 1 - \alpha_2 = P\left(\bar{x} - \frac{t_\gamma s}{\sqrt{n}} \leqslant x \leqslant \bar{x} + \frac{t_\gamma s}{\sqrt{n}}\right)$$

Отсюда асимптотический ДИ:

$$[1.662361, 6.292439]$$

σ^2:

$$\sqrt{n}(\tilde{\sigma} - \sigma) \sim N(0, \frac{\sigma^2}{2})$$

$$\frac{\sqrt{2n}}{\frac{\sqrt{2}}{n}\sum_{i=1}^{n}|x_i - \tilde{a}|}\left(\frac{\sqrt{2}}{n}\sum_{i=1}^{n}|x_i - \tilde{a}| - \sigma\right) \sim N(0,1)$$

$$P\left(-t_\gamma \leqslant \frac{\sqrt{2n}}{\frac{\sqrt{2}}{n}\sum_{i=1}^{n}|x_i - \tilde{a}|}\left(\frac{\sqrt{2}}{n}\sum_{i=1}^{n}|x_i - \tilde{a}| - \sigma\right) \leqslant t_\gamma\right) = 1 - \alpha_2 =$$
$$= P\left(\frac{\sqrt{2}}{n}\sum_{i=1}^{n}|x_i - \tilde{a}| - \frac{t_\gamma \sum_{i=1}^{n}|x_i - \tilde{a}|}{n^{\frac{3}{2}}} \leqslant \sigma^2 \leqslant \frac{\sqrt{2}}{n}\sum_{i=1}^{n}|x_i - \tilde{a}| + \frac{t_\gamma \sum_{i=1}^{n}|x_i - \tilde{a}|}{n^{\frac{3}{2}}}\right)$$

Вычислим ДИ:

$$[81.23379, 157.79624]$$

□

Задача 24. *С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с распределением из семейства Лапласса 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.*

Решение.

Простая гипотеза $H_0 : a = a_0 = -7.00, \sigma^2 = \sigma_0^2 = 10.00$

Согласно теореме Колмогорова:

$\sqrt{n} sup|F_n(x) - F_0(x)| \to K(x)$, где $K(x)$ - функция Колмогорова

$K(C_{\alpha_2}) = 1 - \alpha_2 = 1 - 0.1 = 0.9$
$C_{\alpha_2} = C_{0.2} = 1.1992$
$H_0 : X_1, ..., X_n \sim Laplace(-7.00, 10.00)$

С помощью R вычислим:

$sup|F_n(x) - F_0(x)| = 0.5143445$
$\sqrt{n} sup|F_n(x) - F_0(x)| = 3.636965 > 1.1992$

Отвергаем гипотезу H_0
$p - value = 1.8 \cdot 10^{-14}$
Наибольшее значение уровня значимости, на котором нет оснований отвергнуть гипотезу □

Задача 25. *Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением из семейства Лапласса 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.*

Решение.

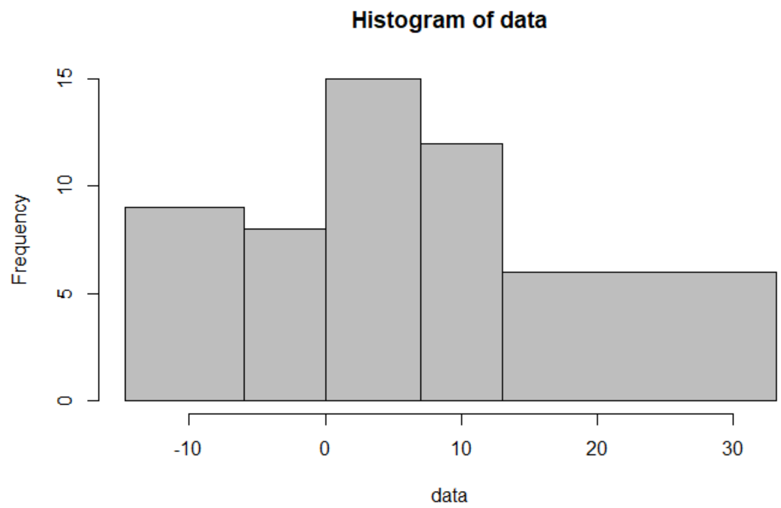
$H_0 : X_1, ..., X_n \sim Laplace(-7.00, 10.00)$

$$\sum_{i=1}^{k} \frac{(m_i - m_i')^2}{m_i} \to \chi^2_{\text{кр}}$$

$\chi^2_{\text{кр}} = \chi^2_4 = 7.77944$

Перестроим гистограмму частот, выбрав следующие точки: $(-Inf, -6, 0, 7, 13, Inf)$

Интервал	$(-\infty; -6]$	$(-6; 0]$	$(0; 7]$	$(7; 13]$	$(13; \infty)$	$\sum$
m_i	7	10	15	12	6	50
p_i	0.56593828	0.24826399	0.11675614	0.03948872	0.02955287	1
$\frac{(m_i - m_i^{'})^2}{m_i^{'}}$	16.0285515	0.4691403	14.3796781	50.9066573	13.8407570	$\chi^2_{\text{набл}}$



$\chi^2_{\text{набл}} = 95.62478 > \chi^2_{\text{кр}} = 7.77944$

$P(\chi^2_{\text{набл}} > 95.62478) \to 0$

Гипотеза отвергается, а точность чисел не позволяет вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть гипотезу (оно крайне близко к 0). □

Задача 26. *Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением из семейства Лапласса. Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение.

$H_0 : X_1, ..., X_n \sim Laplace(a\sigma^2)$

$$\sum_{i=1}^{k} \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \to \chi^2_{\text{кр}}$$

Метод минимизации хи-квадрат:

$$\underset{a,\sigma^2}{argmin} \sum_{i=1}^{k} \frac{(m_i - np_i(a, \sigma^2))^2}{np_i(a, \sigma^2)} \to \chi^2_{\text{кр}}$$

$\chi^2_{\text{набл}} = 2.306802, a = 3.908089, \sigma = 10.716147$

$\chi^2_{\text{набл}} < \chi^2_{\text{кр}} = 4.60517$

Сложная гипотеза согласия с нормальным распределением отвергается.

$P(x > 2.306802) = 0.3155618$

Наибольшее значение уровня значимости, на котором нет оснований отвергнуть гипотезу равно 0.3155618 $\square$