1 Задание 1

1.1 Исходные данные

Вар. 30 (513020125)

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

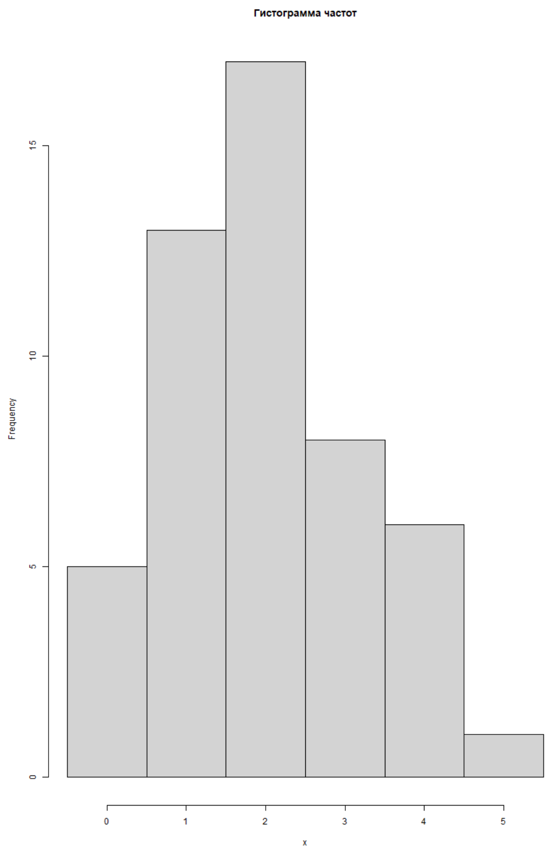
$$\mathbf{P}_\lambda(X = k) = \frac{\lambda^k}{(\lambda+1)^{k+1}},\ k = 0, 1, \ldots.$$

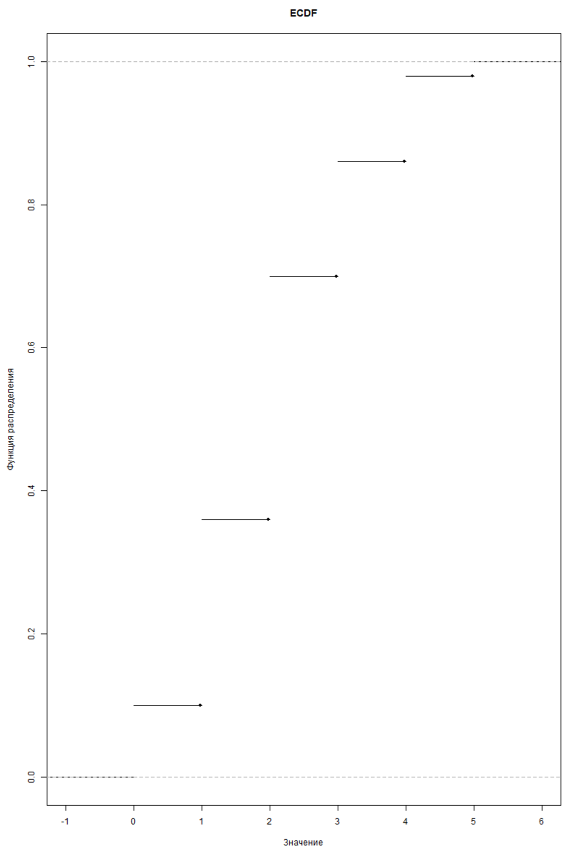
Таблица 1 $\alpha_1 = 0.10$; $a = 0.00$; $b = 3.13$; $\lambda_0 = 4.00$; $\lambda_1 = 2.00$.

5 0 3 1 4 2 0 2 2 3 1 3 0 2 2 4 3 2 0 4 1 2 2 1 1 0 2 1 2 4 1 3 1 2 1 1 2 2 2 1 1 3 3 2 4 4 1
2 3 2

1.2 A

Вариационный ряд : 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5





1.3 B

(i) Математическое ожидание:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{5+0+3+\ldots+2}{50} = 2$$

(ii) Дисперсия:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{50}\sum_{i=1}^{50}(x_i - 2)^2 = 1.48.$$

Примечание: В коде использована формула $s^2 = \frac{n-1}{n} \cdot \mathrm{var}(x)$, так как функция `var()` в R возвращает несмещённую оценку.

(iii) Медиана:

$$\text{Медиана} = \begin{cases} x_{(n/2)}, & \text{если } n \text{ нечётно,} \\ \frac{x_{((n-1)/2)}+x_{((n+1)/2)}}{2}, & \text{если } n\text{чётно.} \end{cases}$$

Для данных ($n = 50$):

$$\text{Медиана} = \frac{x_{(25)} + x_{(26)}}{2} = \frac{2+2}{2} = 2.$$

(iv) Асимметрия:

$$\gamma_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^{23/2}} = 0.3332412$$

(v) Эксцесс:

$$\gamma_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^{22}} - 3 = -0.4616508$$

(vi) Вероятность $P(X \in [0, 3.13])$:

$$\hat{P} = \frac{\text{Число } x_i \in [0, 3.13]}{n} = 0.86.$$

1.4 C

1. Оценка максимального правдоподобия (ОМП)

Для выборки $x_1, x_2, \ldots, x_n$ из распределения Пуассона с параметром λ функция правдоподобия имеет вид:

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}.$$

Логарифмируя, получаем:

$$\ln L(\lambda) = \sum_{i=1}^{n} \left(x_i \ln \lambda - \lambda - \ln x_i! \right).$$

Дифференцируя по λ и приравнивая к нулю:

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0 \implies \hat{\lambda}_{\text{ОМП}} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}.$$

Для данных из таблицы 1 ($n = 50$, $\sum x_i = 100$):

$$\hat{\lambda}_{\text{ОМП}} = \frac{100}{50} = 2.$$

2. Оценка по методу моментов

Для распределения Пуассона математическое ожидание равно λ. Приравниваем теоретический первый момент к выборочному:

$$E[X] = \lambda \implies \hat{\lambda}_{\text{MM}} = \bar{x}.$$

Таким образом:

$$\hat{\lambda}_{\text{MM}} = 2.$$

3. Определение смещения: Смещение оценки $\hat{\theta}$ параметра θ определяется как:

$$\text{Смещение}(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

Если смещение$(\hat{\theta}) = 0$, оценка называется *несмещённой.*

Смещение ОМП: Для распределения Пуассона оценка максимального правдоподобия (ОМП) параметра λ равна выборочному среднему:

$$\hat{\lambda}_{\text{ОМП}} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Математическое ожидание $\hat{\lambda}_{\text{ОМП}}$:

$$E[\hat{\lambda}_{\text{ОМП}}] = E\left[\frac{1}{n} \sum_{i=1}^{n} x_i \right] = \frac{1}{n} \sum_{i=1}^{n} E[x_i].$$

Для Пуассона $E[x_i] = \lambda$, поэтому:

$$E[\hat{\lambda}_{\text{ОМП}}] = \frac{1}{n} \cdot n\lambda = \lambda.$$

Следовательно:

$$\text{Смещение}(\hat{\lambda}_{\text{ОМП}}) = \lambda - \lambda = 0.$$

Смещение оценки по методу моментов: Метод моментов даёт ту же оценку, что и ОМП:

$$\hat{\lambda}_{\text{ММ}} = \bar{x}.$$

Математическое ожидание:

$$E[\hat{\lambda}_{\text{ММ}}] = \lambda \quad \text{(аналогично ОМП)}.$$

Смещение:

$$\text{Смещение}(\hat{\lambda}_{\text{ММ}}) = \lambda - \lambda = 0.$$

Вывод: Обе оценки ($\hat{\lambda}_{\text{ОМП}}$ и $\hat{\lambda}_{\text{ММ}}$) являются **несмещёнными**, так как их математические ожидания совпадают с истинным значением параметра λ.

1.5 D

Шаги построения:

1. **Оценка $\hat{\lambda}$:** ОМП параметра λ равна выборочному среднему:

$$\hat{\lambda} = \bar{x} = 2.$$

2. **Стандартная ошибка:** Для распределения Пуассона дисперсия равна λ, поэтому стандартная ошибка оценки:

$$\text{SE} = \sqrt{\frac{\hat{\lambda}}{n}} = \sqrt{\frac{2}{50}} = 0.2$$

3. **Квантиль нормального распределения:** Для уровня значимости $\alpha_1 = 0.10$ двусторонний доверительный интервал требует квантиль:

$$z_{1-\alpha/2} = z_{0.95} \approx 1.6449.$$

4. **Границы интервала:**

$$\hat{\lambda} \pm z_{0.95} \cdot \text{SE} = 2 \pm 1.6449 \cdot 0.2$$

Вычисляем:

$$\text{Нижняя граница} = 1.67102, \text{Верхняя граница} = 2.32898.$$

Итоговый интервал:

$$\lambda \in (1.67102,\ 2.32898) \quad \text{с уровнем доверия } 90\%.$$

1.6 E

Критерий χ^2 используется для проверки гипотезы о согласии наблюдаемых данных с теоретическим распределением. Основные шаги:

1. Группировка данных на k интервалов.
2. Вычисление наблюдаемых частот O_i для каждого интервала.
3. Вычисление ожидаемых частот E_i для каждого интервала на основе гипотетического распределения.
4. Проверка условия χ^2: все $E_i \geq 5$. Если это условие не выполняется, интервалы объединяются.
5. Вычисление статистики χ^2:
$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$
6. Определение критического значения $\chi^2_{\text{критическое}}$ для уровня значимости α и степеней свободы $df = k - 1$.
7. Сравнение статистики χ^2 с критическим значением: если $\chi^2 > \chi^2_{\text{критическое}}$, гипотеза H_0 отвергается.

Расчет ожидаемых частот Для распределения Пуассона с параметром λ_0 вероятность $P(X = x)$ рассчитывается по формуле:

$$P(X = x) = \frac{\lambda_0^x e^{-\lambda_0}}{x!}.$$

Ожидаемые частоты для каждого интервала вычисляются как:

$$E_i = n \cdot P(X \in \text{интервал}),$$

где n — общее количество наблюдений.

Объединение интервалов Если для какого-либо интервала $O_i < 5$, этот интервал объединяется с соседним. После объединения ожидаемые частоты пересчитываются.

Расчет статистики χ^2 После объединения интервалов статистика χ^2 вычисляется по формуле выше. Количество степеней свободы df уменьшается на количество объединений.

Проверка гипотезы Критическое значение $\chi^2_{\text{критическое}}$ определяется как квантиль распределения χ^2 с df степенями свободы и уровнем значимости α. Если:

$$\chi^2 > \chi^2_{\text{критическое}},$$

гипотеза H_0 отвергается. В противном случае гипотеза не отвергается.

Группировка данных Для упрощения анализа данные группируются в интервалы:

$$(-\infty, 0], \quad (0, 1], \quad (1, 2], \quad (2, 3], \quad (3, \infty).$$

Для каждого интервала вычисляются:

- Наблюдаемые частоты O_i.
- Теоретические вероятности $P(X \in \text{интервал})$.

- Ожидаемые частоты E_i.
- Вклады в статистику χ^2:

$$\frac{(O_i - E_i)^2}{E_i}.$$

Group	lower	upper	O_i	prob	E_i	O_i-E_i	(O_i-E_i)^2/E_i
1	-Inf	0	5	0.0183	0.9158	4.0842	18.2149
2	1	1	13	0.0733	3.6631	9.3369	23.7986
3	2	2	17	0.1465	7.3263	9.6737	12.7734
4	3	3	8	0.1954	9.7683	-1.7683	0.3201
5	4	Inf	7	0.5665	28.3265	-21.3265	16.0563

Статистика: 71.16328
Критическое значение: 7.77944
p-value: 1.289296e-14
Интерпретация результата $\chi^2 > \chi^2_{\text{критическое}}$, гипотеза H_0 отвергается, что означает, что данные не согласуются с распределением Пуассона с $\lambda_0 = 4$.

1.7 F

1. Проверка гипотезы согласия с распределением Пуассона
Мы проверяем, следуют ли данные распределению Пуассона. Для этого используем критерий хи-квадрат.
2. Формулировка гипотез

- **Нулевая гипотеза H_0:** Данные следуют распределению Пуассона с параметром λ.
- **Альтернативная гипотеза H_1:** Данные не следуют распределению Пуассона.

3. Оценка параметра λ
Если параметр λ неизвестен, его оценивают по выборке:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

где x_i — значения выборки, n — объем выборки.
4. Разбиение на интервалы
Данные разбиваются на k интервалов. Для каждого интервала вычисляются:

- **Наблюдаемая частота O_i:** Количество элементов выборки, попавших в i-й интервал.
- **Ожидаемая частота E_i:** Количество элементов, которое ожидается в i-м интервале при условии, что данные следуют распределению Пуассона:

  $$E_i = n \cdot P(\text{интервал } i),$$

  где $P(\text{интервал } i)$ — вероятность попадания в i-й интервал для распределения Пуассона.

5. Вероятности для интервалов
Для распределения Пуассона вероятности вычисляются следующим образом:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Для интервалов:

$$P(\text{интервал } i) = \begin{cases} P(X \leq 0), & \text{если } i = 1, \\ P(X = x), & \text{если } i = 2, \ldots, k-1, \\ P(X \geq x), & \text{если } i = k. \end{cases}$$

6. Статистика хи-квадрат
Статистика хи-квадрат вычисляется по формуле:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

7. Степени свободы
Число степеней свободы для критерия хи-квадрат:

$$\mathrm{df} = k - 1 - m,$$

где:

- k — количество интервалов,
- m — количество оцененных параметров (в данном случае $m = 1$, так как оценивается λ).

8. Критическая область и p-значение

- Критическая область для уровня значимости α определяется как:

$$\chi^2 > \chi^2_{\text{крит}}(\mathrm{df}, \alpha).$$

- p-значение вычисляется как:

$$p = P(\chi^2 > \chi^2_{\text{набл}}).$$

Group	lower	upper	O_i	prob	E_i	(O_i-E_i)	(O_i-E_i)^2/E_i
1	-Inf	0	5	0.1353	6.7668	-1.7668	0.4613
2	1	1	13	0.2707	13.5335	-0.5335	0.0210
3	2	2	17	0.2707	13.5335	3.4665	0.8879
4	3	3	8	0.1804	9.0224	-1.0224	0.1158
5	4	Inf	7	0.1429	7.1438	-0.1438	0.0029

Наблюдаемое значение хи-квадрат: 1.488968
Критическое значение ($\alpha = 0.1$): 6.251389
p-значение: 0.6848189
Нулевая гипотеза не отвергается на уровне значимости 0.1.
Наибольший уровень значимости, при котором гипотеза не отвергается: 0.68482

1.8 G

Логарифм отношения правдоподобия

Функция правдоподобия для распределения Пуассона имеет вид:

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}.$$

Логарифм отношения правдоподобия:

$$\ln\left(\frac{L(\lambda_1)}{L(\lambda_0)}\right) = \sum_{i=1}^{n} \left(X_i \ln\left(\frac{\lambda_1}{\lambda_0}\right) - (\lambda_1 - \lambda_0)\right).$$

Критерий отношения правдоподобия

Наиболее мощный критерий для проверки H_0 против H_1 основан на сравнении суммы наблюдений $T = \sum_{i=1}^{n} X_i$ с порогом k. Критерий принимает H_1, если:

$$T > k,$$

где k определяется из условия:

$$P(T > k \mid H_0) \leq \alpha.$$

При H_0 сумма T имеет распределение Пуассона с параметром $n\lambda_0$. Таким образом, порог k находится как:

$$k = qpois(1 - \alpha, n\lambda_0),$$

где $qpois(p, \lambda)$ — квантиль распределения Пуассона.

Поменяем местами гипотезы

Если поменять местами H_0 и H_1, то новая нулевая гипотеза $H_0 : \lambda = \lambda_1$, а альтернатива $H_1 : \lambda = \lambda_0$. В этом случае критерий принимает H_0, если:

$$T < k',$$

где k' определяется из условия:

$$P(T < k' \mid H_0) \leq \alpha.$$

При H_0 сумма T имеет распределение Пуассона с параметром $n\lambda_1$. Таким образом, порог k' находится как:

$$k' = qpois(\alpha, n\lambda_1).$$

Результаты проверки гипотез

Отклоняем H0 в пользу H1: T_obs = 100 <= k = 182

Не отклоняем H0: T_obs = 100 < k’ = 113

2 Задание 2

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.

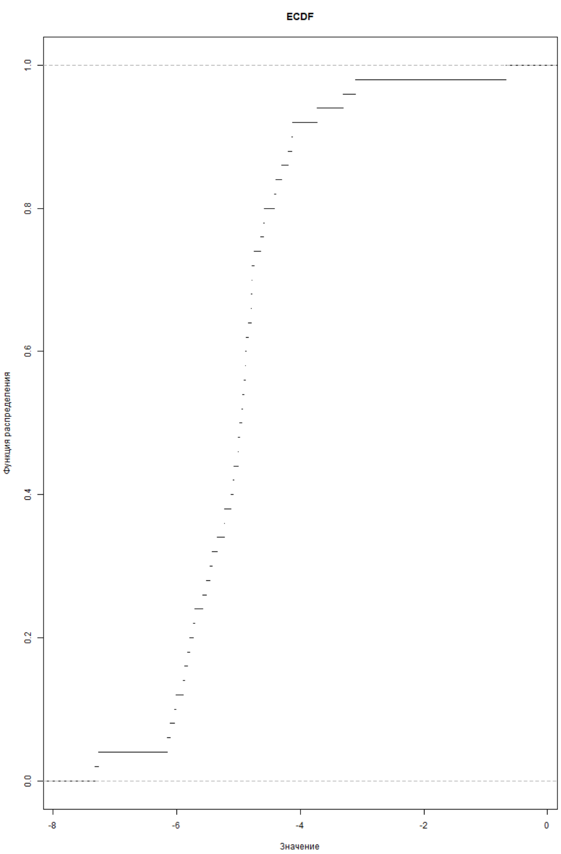
h) Построить наиболее мощный критерий проверки простой гипотезы о нормальности с параметром $(a, \sigma^2) = (a_0, \sigma_0^2)$ при альтернативе нормальности с параметром $(a, \sigma^2) = (a_1, \sigma_1^2)$. Проверить гипотезу на уровне значимости α_2. Что получится, если поменять местами основную и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

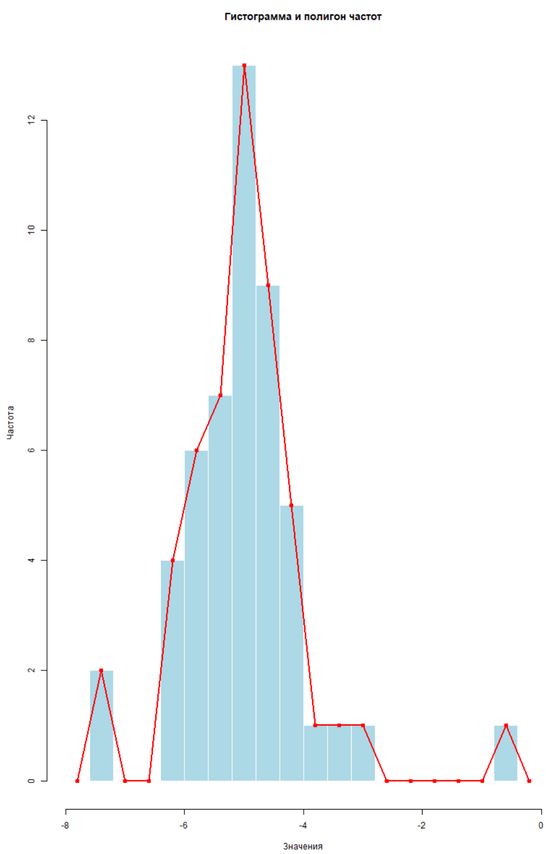
i) В пунктах (c)–(g) заменить семейство нормальных распределений на двухпараметрическое семейство распределений Лапласа с плотностями $f(x) = \dfrac{1}{\sqrt{2}\,\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x-a|}$.

Таблица 2 $\alpha_2 = 0.05$; $c = -6.00$; $d = -4.20$; $h = 0.40$; $a_0 = -3.00$; $\sigma_0 = 1.00$; $a_1 = -5.00$; $\sigma_1 = 1.00$.

−5.704	−4.882	−5.428	−4.305	−3.106	−7.321	−6.007	−0.676	−5.086	−5.795	−4.593	−4.123	−5.224	−4.911	−4.778	−7.262
−4.637	−4.581	−5.007	−4.794	−4.136	−6.111	−4.983	−5.230	−5.870	−5.005	−5.465	−4.399	−4.837	−4.200	−5.344	−3.305
−5.583	−4.943	−4.874	−4.795	−5.519	−4.423	−4.934	−4.785	−5.821	−5.073	−4.893	−5.733	−5.124	−4.742	−6.030	−5.900
−6.154	−3.731														

2.1 A





Вариационный ряд: -7.321, -7.262, -6.154, -6.111, -6.030, -6.007, -5.900, -5.870, -5.821, -5.795, -5.733, -5.704, -5.583, -5.519, -5.465, -5.428, -5.344, -5.230, -5.224, -5.124, -5.086, -5.073, -5.007, -5.005, -4.983, -4.943, -4.934, -4.911, -4.893, -4.882, -4.874, -4.837, -4.795, -4.794, -4.785, -4.778, -4.742, -4.637, -4.593, -4.581, -4.423, -4.399, -4.305, -4.200, -4.136, -4.123, -3.731, -3.305, -3.106, -0.676

2.2 B

Для выборки $X = (x_1, x_2, \dots, x_n)$ объёма n вычисляются следующие характеристики:

1. Выборочное среднее (эмпирическое математическое ожидание)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

2. Выборочная дисперсия

$$\text{Смещённая: } s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2}$$

$$\text{Несмещённая: } \tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3}$$

3. Выборочная медиана

$$\operatorname{med}(X) = \begin{cases} x_{(k+1)}, & \text{если } n = 2k+1 \\ \frac{x_{(k)} + x_{(k+1)}}{2}, & \text{если } n = 2k \end{cases} \tag{4}$$

где $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$ - вариационный ряд.

4. Выборочный коэффициент асимметрии

$$g_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \tag{5}$$

5. Выборочный коэффициент эксцесса

$$g_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3 \tag{6}$$

Результаты:
Среднее: -5.00324
Дисперсия: 1.045139
Медиана: -4.963
Асимметрия: 1.25438
Эксцесс: 5.384865
$P(X \in [-6, -4.2]) : 0.76$

2.3 C

Предполагаем, что выборка $X = (x_1, x_2, \dots, x_n)$ происходит из нормального распределения $N(a, \sigma^2)$ с плотностью:

$$f(x|a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

1. Оценки максимального правдоподобия (ОМП)
Функция правдоподобия:

$$L(a, \sigma^2) = \prod_{i=1}^{n} f(x_i | a, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right)$$

Логарифмическая функция правдоподобия:

$$\ell(a, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2$$

Уравнения правдоподобия:

$$\begin{cases} \frac{\partial \ell}{\partial a} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = 0 \\ \frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - a)^2 = 0 \end{cases}$$

Решения дают ОМП:

$$\hat{a}_{\text{МП}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad \hat{\sigma}^2_{\text{МП}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

2. Метод моментов
Приравниваем выборочные моменты к теоретическим:

$$\begin{cases} EX = a = \bar{x} \\ EX^2 = a^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{cases}$$

Оценки метода моментов:

$$\tilde{a} = \bar{x}, \quad \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

3. Смещение оценок

- Оценка $\hat{a}_{\text{МП}} = \tilde{a}$ является несмещённой:

$$E\hat{a}_{\text{МП}} = a$$

- Оценка $\hat{\sigma}^2_{\text{МП}}$ смещённая:

$$E\hat{\sigma}^2_{\text{МП}} = \frac{n-1}{n}\sigma^2$$

- Несмещённая оценка дисперсии:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Результаты
Оценки параметров нормального распределения: ОМП:$a = -5.00324\sigma^2 = 1.024236$
Метод моментов: $a = -5.00324$
$\sigma^2 = 1.045139$
Смещение оценки σ^2 : -0.02048471

2.4 D

Используем статистику:

$$T = \frac{\overline{X} - a}{S/\sqrt{n}} \sim t_{n-1}$$

где $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ — выборочное среднее, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ — исправленная выборочная дисперсия, t_{n-1} — распределение Стьюдента с $n-1$ степенями свободы.

Доверительный интервал уровня $1-\alpha$:

$$P\left(-t_{1-\alpha/2} \leq \frac{\overline{X} - a}{S/\sqrt{n}} \leq t_{1-\alpha/2}\right) = 1 - \alpha$$

где $t_{1-\alpha/2}$ — квантиль распределения Стьюдента.

Отсюда получаем интервал:

$$a \in \left(\overline{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$$

2. Доверительный интервал для дисперсии σ^2

Используем статистику:

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

Доверительный интервал уровня $1-\alpha$:

$$P\left(\chi^2_{\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}\right) = 1 - \alpha$$

Отсюда получаем интервал:

$$\sigma^2 \in \left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right)$$

Доверительные интервалы (95%):
Для a: (-5.29378 , -4.7127)
Для σ^2: (0.7292798 , 1.622942)

2.5 E

Статистика критерия Колмогорова вычисляется как максимальное отклонение между теоретической и эмпирической функциями распределения:

$$D_n = \sup_x |F_n(x) - F_0(x)|$$

На практике вычисляется как:

$$D_n = \max\left(\max_{1\leq i\leq n}\left|F_0(x_{(i)}) - \frac{i-1}{n}\right|, \max_{1\leq i\leq n}\left|F_0(x_{(i)}) - \frac{i}{n}\right|\right)$$

где $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ – вариационный ряд.

Вычисление p-value

P-value вычисляется через функцию распределения статистики Колмогорова:

$$p = 1 - K\left(D_n\sqrt{n}\right)$$

где $K(t)$ – функция распределения Колмогорова:

$$K(t) = \frac{\sqrt{2\pi}}{t} \sum_{k=1}^{\infty} e^{-(2k-1)^2\pi^2/(8t^2)}$$

Для больших выборок ($n > 35$) используется аппроксимация:

$$p \approx 2\exp\left(-2nD_n^2\right)$$

Критическая область
Гипотеза H_0 отвергается на уровне значимости α, если:

$$D_n > D_\alpha$$

где D_α – критическое значение, которое можно найти в таблицах или вычислить как:

$$D_\alpha \approx \frac{C(\alpha)}{\sqrt{n}}$$

Для $\alpha = 0.05$, $C(\alpha) \approx 1.36$.

```
Точка максимального расхождения:

i     xi  lecdf  recdf   hcdf      ldif      rdif    maxdif
46 -4.123   0.9   0.92 0.1307187 0.7692813 0.7892813 0.7892813


Статистика Колмогорова D = 5.581061
P-value = 8.881784e-16
```

2.6 F

Разбиение на интервалы Для k интервалов с границами $-\infty = b_0 < b_1 < \ldots < b_k = +\infty$:

$$(b_j)_{j=0}^{k} = \text{brk} \cup \{-\infty, +\infty\} \tag{7}$$

Наблюдаемые частоты Число наблюдений в j-м интервале $(b_{j-1}, b_j]$:

$$n_j = \sum_{i=1}^{n} I_{\{b_{j-1} < x_i \le b_j\}}, \quad j = 1, \ldots, k \tag{8}$$

Теоретические вероятности Вероятность попадания в интервал при H_0:

$$p_j = \Phi\left(\frac{b_j - a_0}{\sigma_0}\right) - \Phi\left(\frac{b_{j-1} - a_0}{\sigma_0}\right) \tag{9}$$

где Φ – функция стандартного нормального распределения.

Ожидаемые частоты

$$e_j = n \cdot p_j \tag{10}$$

Вычисление статистики

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - e_j)^2}{e_j} \sim \chi^2_{k-1} \tag{11}$$

Нормированные остатки

$$r_j = \frac{n_j - e_j}{\sqrt{e_j}} \tag{12}$$

Таблица вычислений
Критическая область При уровне значимости α:

$$\chi^2 > \chi^2_{1-\alpha}(k-1) \tag{13}$$

p-значение

$$p = P(\chi^2_{k-1} > \chi^2_{\text{набл}}) \tag{14}$$

Результаты Критическое значение хи-квадрат: 15.50731
Статистика X2 = 1179.432
Отвергаем гипотезу (X2 > xal)? TRUE
P-значение: 2.664047e-249

Group	lower	upper	count	prob	expr	resid1	resid2
1	-Inf	-6.00	6	0.001349898	0.0674949	22.8350852	521.4411151
2	-6.00	-5.70	6	0.002117076	0.1058538	18.1162248	328.1976019
3	-5.70	-5.30	5	0.007257136	0.3628568	7.6980879	59.2605576
4	-5.30	-5.01	5	0.011491484	0.5745742	5.8382407	34.0850540
5	-5.01	-4.90	6	0.006500965	0.3250483	9.9537847	99.0778299
6	-4.90	-4.79	6	0.008010396	0.4005198	8.8478086	78.2837179
7	-4.79	-4.58	6	0.020326478	1.0163239	4.9434910	24.4381031
8	-4.58	-4.13	5	0.072184679	3.6092340	0.7320598	0.5359116
9	-4.13	Inf	5	0.870761888	43.5380944	-5.8405740	34.1123042

2.7 G

Оценка параметров методом минимизации χ^2 Параметры оцениваются путем минимизации статистики хи-квадрат:

$$(\hat{a}, \hat{\sigma}) =_{a,\sigma} \sum_{j=1}^{k} \frac{(n_j - e_j(a,\sigma))^2}{e_j(a,\sigma)} \tag{15}$$

где:

- $e_j(a,\sigma) = n \cdot p_j(a,\sigma)$
- $p_j(a,\sigma) = \Phi\left(\frac{b_j - a}{\sigma}\right) - \Phi\left(\frac{b_{j-1} - a}{\sigma}\right)$

Модифицированная статистика χ^2 После оценки параметров статистика имеет вид:

$$\chi^2_C = \sum_{j=1}^{k} \frac{(n_j - e_j(\hat{a},\hat{\sigma}))^2}{e_j(\hat{a},\hat{\sigma})} \tag{16}$$

Распределение статистики При справедливости H_0:

$$\chi^2_C \sim \chi^2_{k-3} \tag{17}$$

Число степеней свободы уменьшается на 2 (по числу оцененных параметров).

Шаг 1. Оценка параметров Используется метод Ньютона для минимизации:

$$nlm(f = \chi^2\text{-статистика}, p = \text{начальные приближения}) \tag{18}$$

Шаг 2. Вычисление статистики По полученным оценкам $\hat{a}$ и $\hat{\sigma}$:

$$\chi^2_C = \sum_{j=1}^{k} \frac{(n_j - n\hat{p}_j)^2}{n\hat{p}_j} \tag{19}$$

Шаг 3. Проверка гипотезы Критическая область:

$$\chi^2_C > \chi^2_{1-\alpha}(k-3) \tag{20}$$

Результаты

Результаты для критерия хи-квадрат с оцененными параметрами:
Оценки параметров: mean = -5.07886 sd = 0.7077514
Минимизированная статистика X2C = 10.99842
Критическое значение: 12.59159
Отвергаем гипотезу (X2C > xal)? FALSE
P-значение: 0.08842517

```
Таблица для критерия хи-квадрат (оцененные параметры):
Group lower   upper   count      prob        expr        resid1       resid2
1      -Inf   -6.00     6     0.09654327   4.827164    0.5338158   0.28495934
2     -6.00   -5.70     6     0.09353043   4.676521    0.6120057   0.37455100
3     -5.70   -5.30     5     0.18727386   9.363693   -1.4260364   2.03357970
4     -5.30   -5.01     5     0.16140593   8.070297   -1.0807757   1.16807619
5     -5.01   -4.90     6     0.06100255   3.050127    1.6890569   2.85291333
6     -4.90   -4.79     6     0.05865760   2.932880    1.7909508   3.20750493
7     -4.79   -4.58     6     0.10113515   5.056758    0.4194568   0.17594404
8     -4.58   -4.13     5     0.15043714   7.521857   -0.9195131   0.84550433
9     -4.13     Inf     5     0.09001408   4.500704    0.2353520   0.05539056
```

2.8 H

Статистика отношения правдоподобия:

$$\Lambda(X) = \frac{\mathcal{L}(\mu_0, \sigma_0^2)}{\mathcal{L}(\mu_1, \sigma_1^2)} = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\left(\frac{(X_i-\mu_0)^2}{\sigma_0^2} - \frac{(X_i-\mu_1)^2}{\sigma_1^2}\right)\right]\right\} \tag{21}$$

Логарифмическая форма Удобнее работать с логарифмом:

$$\ln\Lambda(X) = n\ln\left(\frac{\sigma_1}{\sigma_0}\right) - \frac{1}{2}\left[\frac{S_{00}}{\sigma_0^2} - \frac{S_{11}}{\sigma_1^2}\right] \tag{22}$$

где $S_{ab} = \sum_{i=1}^{n}(X_i-\mu_a)(X_i-\mu_b)$

Критическая область Критерий имеет вид:

$$\ln\Lambda(X) \le c_\alpha \tag{23}$$

где c_α выбирается из условия:

$$P(\ln \Lambda(X) \leq c_\alpha | H_0) = \alpha \tag{24}$$

При замене H_0 и H_1 местами:

- Критическая область становится $\ln \Lambda(X) \geq c'_\alpha$
- Мощность критерия изменится
- Условие оптимальности перестанет выполняться

Результаты

Критическое значение: -3.4967
Логарифм отношения правдоподобия для тестовых данных: 75.08311
Отвергаем H0?: TRUE
При замене гипотез местами:
Новое критическое значение: -77.11835
Значит в этом случае мы не отвергаем H0.
Данные согласуются с $\mathrm{N}(a_1, \sigma_1^2)$, а не с $\mathrm{N}(a_0, \sigma_0^2)$
Если поменять гипотезы местами, то $\mathrm{N}(a_1, \sigma_1^2)$ не отвергается.