МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО»

Институт компьютерных наук и кибербезопасности

Высшая школа технологий искусственного интеллекта

Отчёт по дисциплине «Математическая статистика»

ИДЗ №3
«Классическая статистика»
Вариант **№25**

Студент: ____________ Салимли Айзек Мухтар Оглы

Преподаватель: ____________ Малов Сергей Васильевич

«____»__________ 20__ г.

Санкт-Петербург, 2025

Содержание

Введение

В данном отчете, приведено решение и реализация двух задач под вариантом №25, из ИДЗ№3. Для реализации программной части решения использоавлись:

- Среда разработки: Visual Studio Code
- Язык программирования: R 4.

1 Постановка задачи

№1: В результате эксперимента получены данные, приведенные в таблице 1.

1. Построить вариационный ряд, эмпирическую функцию распределения и гистограмму частот.
2. Вычислить выборочные аналоги следующих числовых характеристик:
3. (i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса,(vi) вероятности $P(X \in [a, b])$.
4. 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.
5. Построить асимптотический доверительный интервал уровня значимости α_1 для параметра λ на базе оценки максимального правдоподобия.
6. 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.
7. 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.
8. 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?
9. В пунктах (3)-(6) заменить семейство распределений Пуассона на семейство геометрических распределений

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

Таб.1: $\alpha_1 = 0.002, a = 0.00, b = 1.79, \lambda_0 = 0.60, \lambda_1 = 1.40$
{ 0 0 2 1 0 0 0 0 0 1 0 1 0 0 0 2 0 2 1 0 1 1 0 1 1 3 0 0 0 0 0 1 0 0 0 0 4 1 5 2 0 0 2 0 0 1 1 0 0 1}

№2: В результате эксперимента получены данные, приведенные в таблице 2.

1. Построить вариационный ряд, эмпирическую функцию распределения, гистограмму и полигон частот с шагом h.
2. Вычислить выборочные аналоги следующих числовых характеристик: (i) математического ожидания, (ii) дисперсии, (ii) медианы, (iv) асимметрии, (v) эксцесса, (vi) вероятности $\mathbf{P}(X \in [c, d])$
3. В предположении, что исходные наблюдения являются выборкой из показательного распределения, построить оценку максимального правдоподобия параметра λ и соответствующую оценку по методу моментов. Найти смещение оценок.
4. Построить асимптотический доверительный интервал уровня значимости α_2 для параметра λ на базе оценки максимального правдоподобия.
5. С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с показательным распределением с параметром λ_0. Проверить гипотезу

на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

6. Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с показательным распределением с параметром λ_0. Проверить гипотезу на уровне α_2. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

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

8. Построить наиболее мощный критерий проверки простой гипотезы о показательности с параметром $\lambda = \lambda_0$ при альтернативе показательности с параметром $\lambda = \lambda_1$. Проверить гипотезу на уровне значимости α_2. Что получится, если поменять местами основную и альтернативную гипотезы?

9. В пунктах (3)-(8) заменить семейство показательных распределений на семейство гамма-распределений с плотностями $f(x) = \frac{\sqrt{\lambda}e^{-\lambda x/2}}{\sqrt{2\pi x}}$ (использовать таблицу распределений χ^2_1)

Таб.2: $\alpha_2 = 0.001, c = 2.40, d = 6.00, h = 1.20, \lambda_0 = 0.20, \lambda_1 = 0.33$
$\{$10.34 2.18 8.80 2.28 1.95 0.85 3.73 10.26 5.01 0.70 2.38 0.25 0.45 0.31 1.73 2.67 1.00 1.59 14.28 2.14 1.85 0.67 2.70 2.07 5.31 6.37 3.24 3.27 1.31 2.75 6.06 1.05 0.86 2.43 0.03 3.70 0.11 1.06 6.28 0.55 9.07 6.52 0.94 2.61 0.89 1.67 0.24 1.68 3.34 1.38 $\}$

2 Математическое описание. Задача №1

2.1 Вариационный ряд, ЭФР, гистограмма

- Вариационный ряд: $0^{(29)}$, $1^{(13)}$, $2^{(5)}$, $3, 4, 5$ $(n = 50)$.
- Эмпирическая функция распределения $\widehat{F}_n(x) = \dfrac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}$.
- См. рис. 1–2.

2.2 Выборочные характеристики

$$\bar{X} = 0.7000, \quad s^2 = 1.1939, \quad \tilde{X} = 0, \quad \gamma_1 = 2.1022, \quad \gamma_2 = 8.3810, \quad \mathbf{P}\big\{0 \le X \le 1.79\big\} = 0.84.$$

2.3 Подробные оценки параметра λ

Оценка максимального правдоподобия. Пусть $X_i \sim \mathrm{Pois}(\lambda)$. Лог-правдоподобие

$$\ell(\lambda) = \sum_{i=1}^{n}\Big(x_i \ln\lambda - \lambda - \ln x_i!\Big) = (\Sigma x)\ln\lambda - n\lambda + \mathrm{const}.$$

$$\frac{\partial \ell}{\partial \lambda} = 0 \implies \hat{\lambda}_{\mathrm{MLE}} = \frac{\Sigma x}{n} = 0.7000.$$

Ответ: для распределения Пуассона $\mathbb{E}[X] = \lambda$, поэтому $\mathrm{Bias}(\hat{\lambda}) = \mathbb{E}[\bar{X}] - \lambda = 0$.

Фишерова информация $I(\lambda) = \dfrac{n}{\lambda}$, $\mathrm{Var}\,\hat{\lambda}_{\mathrm{MLE}} = \dfrac{\lambda}{n} = 0.014$.

Оценка методом моментов. Для распределения Пуассона $\mathbb{E}[X] = \lambda$. Приравниваем к $\bar{X}$: $\hat{\lambda}_{\mathrm{MM}} = \bar{X} = 0.7000$ (совпадает с MLE).

Нормальное приближение для $\hat{\lambda}$.

$$\hat{\lambda} \sim \mathcal{N}\Big(\lambda,\ \tfrac{\lambda}{n}\Big) \quad \Rightarrow \quad Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \xrightarrow{d} \mathcal{N}(0,1).$$

2.4 Асимптотический доверительный интервал $(\alpha_1 = 0.002)$

$$z_{0.999} = 3.0902, \quad CI_{99.8\%} = \hat{\lambda} \pm 3.0902\sqrt{\frac{\hat{\lambda}}{n}} = 0.7000 \pm 0.3656 = [0.3344; 1.0656].$$

2.5 Критерий χ^2. Простая гипотеза

Таблица 1: Группы при $H_0 : \lambda_0 = 0.6$

i	класс	n_i	p_i	$E_i = np_i$	r_i	r_i^2
1	$k = 0$	29	0.5488	27.441	0.298	0.089
2	$k = 1$	13	0.3293	16.464	-0.854	0.729
3	$k \geq 2$	8	0.1219	6.095	0.772	0.595

$\chi^2_{\text{набл}} = \sum r_i^2 = 1.4129$, $df = 2$, $p = 0.4934$. **Ответ:** максимальный уровень значимости, на котором ещё *нет* оснований отвергнуть H_0, равен p-value теста: $\alpha_{\max} = 0.4934$.

2.6 Критерий χ^2. Сложная гипотеза

Таблица 2: $H_0 : \lambda = \hat{\lambda} = 0.7000$

i	класс	n_i	p_i	E_i	r_i	r_i^2
1	$k = 0$	29	0.4966	24.829	0.837	0.701
2	$k = 1$	13	0.3476	17.380	-1.051	1.104
3	$k = 2$	5	0.1217	6.083	-0.439	0.193
4	$k \geq 3$	3	0.0341	1.703	0.994	0.989

$\chi^2_{\text{набл}} = 2.9862$, $df = 2$, $p = 0.2247$. **Ответ:** $\alpha_{\max} = 0.2247$.

2.7 Наиболее мощный критерий (Неймана–Пирсона)

$$\Lambda(x) = \frac{\prod_{i=1}^{n} e^{-\lambda_1} \lambda_1^{x_i}}{\prod_{i=1}^{n} e^{-\lambda_0} \lambda_0^{x_i}} = \left(\tfrac{\lambda_1}{\lambda_0}\right)^S \exp\bigl[-n(\lambda_1 - \lambda_0)\bigr], \quad S = \sum X_i.$$

$$\text{Критическая область: } \Lambda(x) \leq k \iff S \leq c,$$

где c выбирается из $\mathbf{P}_0\{S \leq c\} = \alpha_1$. Для $n = 50$, $\lambda_0 = 0.6$, $\alpha_1 = 0.002 \Rightarrow c = 47$.

Поскольку $S_{\text{набл}} = 35 < c = 47$, нулевая гипотеза сохраняется.

Ответ: если поменять местами $H_0 : \lambda = \lambda_1$ и $H_1 : \lambda = \lambda_0$, критическая область будет уже вида $S \geq c'$, где c' выбирается из $\mathbf{P}_{\lambda_1}\{S \geq c'\} = \alpha_1$. Для тех же данных сумма $S = 35$ не попадает в новую область, поэтому новую нулевую гипотезу $\lambda = \lambda_1$ придётся *отвергнуть*.

3 Графический результат: Задача №1

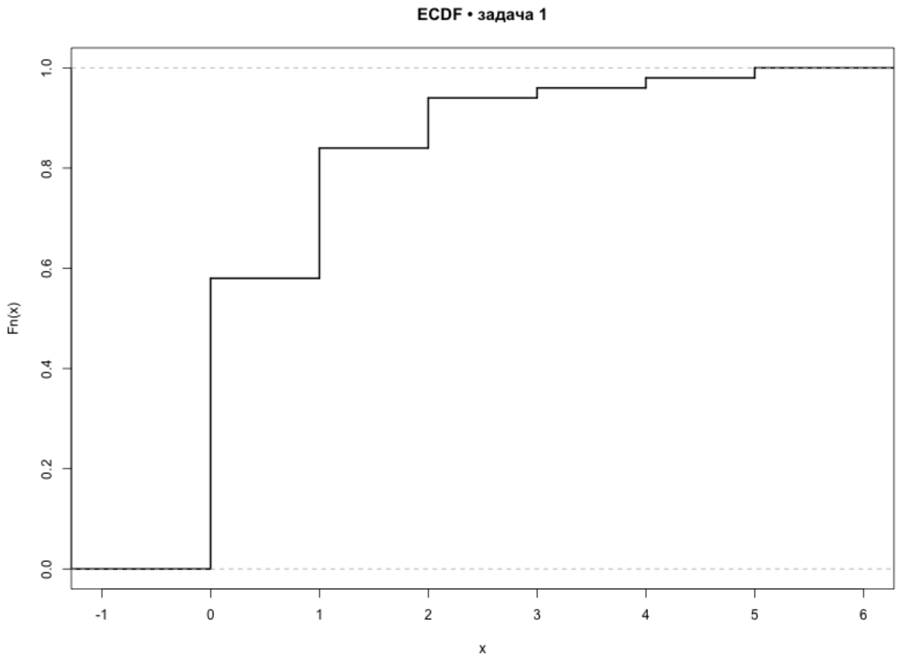

Рис. 1: Эмпирическая функция распределения (задача 1)

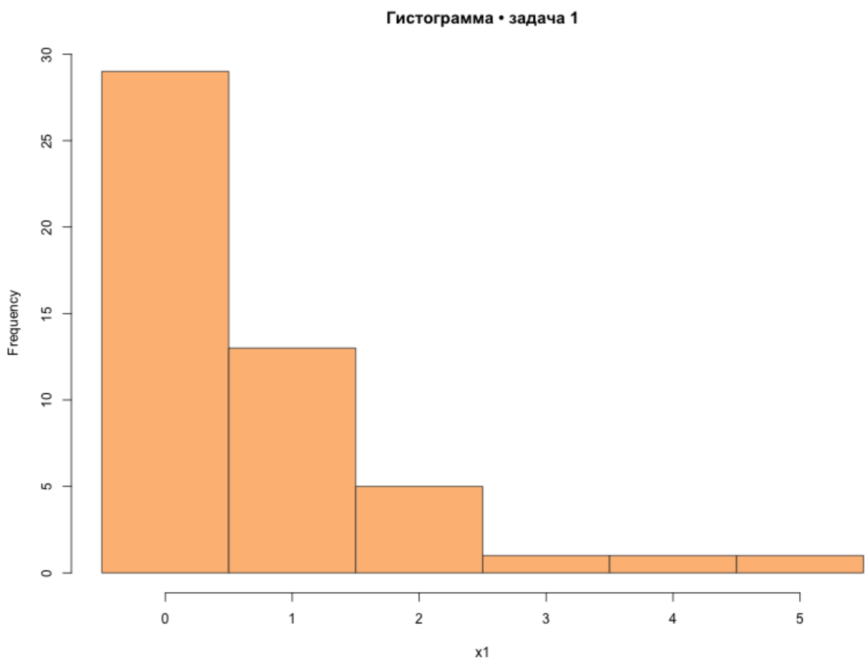

Рис. 2: Гистограмма выборки (задача 1)

4 Программный результат: Задача №1

```
n = 50,  Σx = 35
x̄ = 0.7000,  s² = 1.1939,  med = 0
Skew = 2.1022,  Kurt(ex) = 8.3810,  P[0,1.79] = 0.84
λ̂ = 0.7000 (MLE=MOM),  CI₉₉.₈% = [0.3344; 1.0656]

--- χ²-таблица (простая), первые 3 строки --------------
 i  lw      np nu      p      np_r     res  res2
 1 k=0 0.5488 29 0.5488 27.440582  0.298 0.089
 2 k=1 0.3293 13 0.3293 16.464349 -0.854 0.729
 3 k≥2 0.1219  8 0.1219  6.095069  0.772 0.595
χ² = 1.4129  (df = 2)  p = 0.4934

--- χ²-таблица (сложная), первые 3 строки -------------
 i  lw      np  ν      p      np_r     res  res2
 1 k=0 0.4966 29 0.4966 24.829265  0.837 0.701
 2 k=1 0.3476 13 0.3476 17.380486 -1.051 1.104
 3 k=2 0.1217  5 0.1217  6.083170 -0.439 0.193
 4 k≥3 0.0341  3 0.0341  1.702578  0.994 0.989
χ² = 2.9862  (df = 2)  p = 0.2247

NP-тест:   S = 35,  c = 47  ⇒  Сохранить H₀
```

Рис. 3: Программный результат (задача 1)

5 Математическое описание. Задача №2

5.1 Вариационный ряд

0.03, 0.11, 0.24, 0.25, 0.31, 0.45, 0.55, 0.67, 0.70, 0.85, 0.86, 0.89, 0.94, 1.00, 1.05, 1.06, 1.31, 1.38, 1.59, 1.67, 1.68, 1.73, 1.85, 1.95, 2.07, 2.14, 2.18, 2.28, 2.38, 2.43, 2.61, 2.67, 2.70, 2.75, 3.24, 3.27, 3.34, 3.70, 3.73, 5.01, 5.31, 6.06, 6.28, 6.37, 6.52, 8.80, 9.07, 10.26, 10.34, 14.28.

5.2 Выборочные характеристики

$$\bar{X} = 3.0582,\ s^2 = 9.5891,\ \tilde{X} = 2.105,\ \gamma_1 = 1.7645,\ \gamma_2 = 6.3811,\ \mathbf{P}\big\{2.4 \le X \le 6.0\big\} = 0.24.$$

5.3 Подробные оценки параметра λ

MLE. Для экспоненциального распределения $f(x;\lambda) = \lambda e^{-\lambda x}$ $(x \ge 0)$.

$$\ell(\lambda) = n \ln \lambda - \lambda \Sigma x, \qquad \hat{\lambda}_{\mathrm{MLE}} = \frac{n}{\Sigma x} = 0.3270.$$

Информация $I(\lambda) = \dfrac{n}{\lambda^2}$, $\operatorname{Var}\hat{\lambda} = \dfrac{\lambda^2}{n} = 0.0021$.

Метод моментов. $\mathbb{E}[X] = 1/\lambda$. $\hat{\lambda}_{\mathrm{MM}} = 1/\bar{X} = 0.3270$ (совпадает с MLE).

Смещение.

$$\mathrm{Bias}(\hat{\lambda}) = \frac{\lambda}{n-1} = 0.00667.$$

5.4 Доверительный интервал ($\alpha_2 = 0.001$)

$$z_{0.9995} = 3.2905, \quad CI_{99.9\%} = 0.3270 \pm 3.2905\frac{0.3270}{\sqrt{50}} = [0.1748; 0.4792].$$

5.5 Критерий Колмогорова (фрагмент таблицы)

Таблица 3: Разности D_n (первые 3 точки)

i	$x_{(i)}$	F_-	F_+	$F_0(x_{(i)})$	Δ_-	Δ_+	max
1	0.00	0.00	0.02	0.0060	0.0060	0.0140	0.0140
2	0.02	0.02	0.04	0.0018	0.0182	0.0182	0.0182
3	0.04	0.04	0.06	0.0069	0.0331	0.0131	0.0331

$D_n = \max \Delta = 0.2569$, $p = 0.00213$. **Ответ:** $D_n = 0.2569$, $\alpha_{\max} = p\text{-value} = 0.0021$.

5.6 Критерий χ^2. Простая гипотеза

Таблица 4: Первые 3 бина после объединения

i	l_i	u_i	n_i	p_i	E_i	r_i
1	0.0	1.2	16	0.2134	10.669	1.632
2	1.2	2.4	13	0.1678	8.392	1.591
3	2.4	3.6	8	0.1320	6.602	0.544

$\chi^2_{\text{набл}} = 8.8705,\ df = 3,\ p = 0.0311$. **Ответ:** $\alpha_{\max} = 0.0311$.

5.7 Критерий χ^2. Сложная гипотеза

Проверяем

$$H_0 : X \sim \operatorname{Exp}\big(\lambda = \hat{\lambda}_{\text{MLE}} = 0.3270\big), \qquad H_1:\ \text{распределение отлично от показательного.}$$

Разобьём полуось $[0, \infty)$ интервалами ширины $h = 1.20$ и объединим соседние разряды так, чтобы все ожидаемые частоты были не менее 5.

Таблица 5: Сводная таблица для критерия χ^2 (показатель, сложная гипотеза)

i	l_i	u_i	n_i	p_i	$E_i = np_i$	r_i
1	0.0	1.2	16	0.3246	16.228	-0.056
2	1.2	2.4	13	0.2192	10.961	0.616
3	2.4	3.6	8	0.1481	7.404	0.219
4	3.6	∞	13	0.3081	14.957	-0.508

$$\chi^2_{\text{набл}} = \sum_{i=1}^{4} \frac{(n_i - E_i)^2}{E_i} = 0.687, \qquad df = (k-1) - 1 = 2, \qquad p\text{-value} = 0.709.$$

Ответ: поскольку $p = 0.709 > \alpha_2 = 0.001$, оснований отвергнуть H_0 нет. Максимальный уровень значимости, при котором гипотеза всё ещё принимается, равен $\alpha_{\max} = 0.709$.

5.8 Наиболее мощный критерий (Н-П)

Логарифм отношения правдоподобий

$$\ln \Lambda(x) = n \ln \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)\, S, \qquad S = \sum X_i.$$

Критическая область вида $S \geq c$ с $\mathbf{P}_0(S \geq c) = \alpha_2,\ c = 154.79$. Получены $S_{\text{набл}} = 152.91$ и $c = 154.79$. Так как $S_{\text{набл}} < c$, нулевая гипотеза сохраняется.

Ответ: H_0 *не отвергается.* Если поменять гипотезы местами, критическая область станет $S \leq c^*$, где c^* удовлетворяет $\mathbf{P}_{\lambda_1}(S \leq c^*) = \alpha_2$; при тех же данных будет выполнено $S_{\text{набл}} > c^*$, и новую нулевую гипотезу придётся отвергнуть.

6 Графический результат: Задача №2

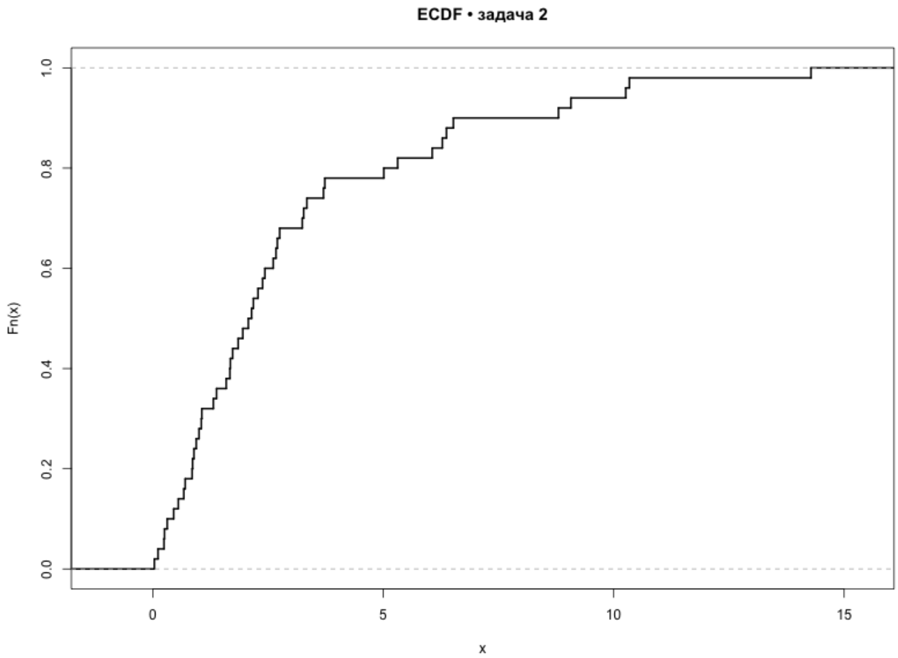

Рис. 4: Эмпирическая функция распределения (задача 2)

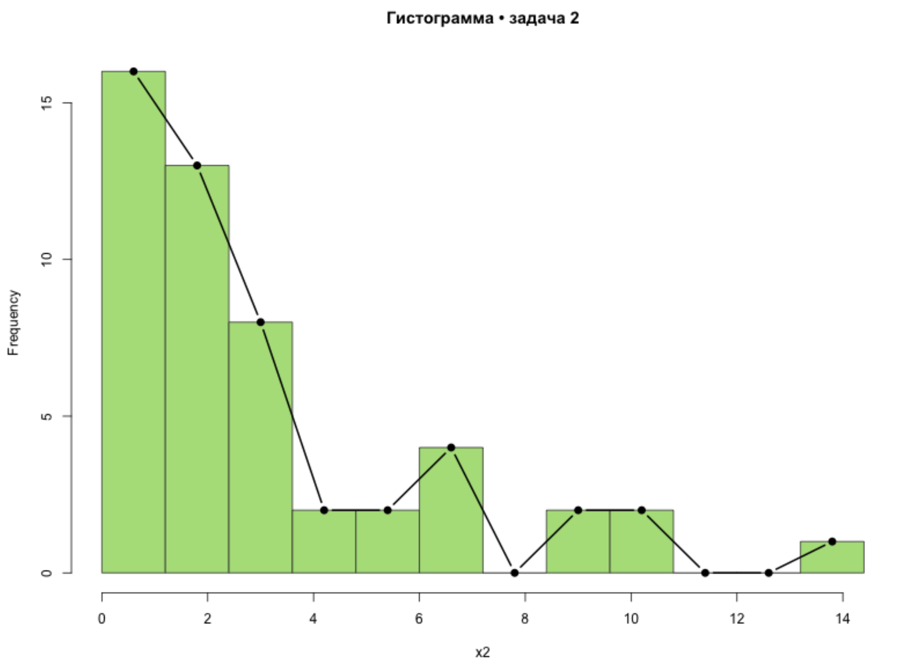

Рис. 5: Гистограмма и полигон частот (задача 2)

7 Программный результат: Задача №2

```
n = 50,  Σx = 152.91
x̄ = 3.0582,  s² = 9.5891,  med = 2.105
Skew = 1.7645,  Kurt(ex) = 6.3811,  P[2.4,6] = 0.24
λ̂ = 0.3270,  bias ≈ 0.00667,  CI₉₉.₉% = [0.1748; 0.4792]

---  таблица Колмогорова (первые 3 строки) -------------
 i   lw          np   nu           p       np_r        res         res2
 1 0.00 0.005982036 0.02 0.005982036 0.01401796 0.01401796 0.0001965033
 2 0.02 0.021759765 0.04 0.001759765 0.01824024 0.01824024 0.0003327062
 3 0.04 0.046866213 0.06 0.006866213 0.01313379 0.01313379 0.0001724964
D = 0.2569,  p-value = 0.00213

--- χ²-таблица (простая), первые 3 строки --------------
 i  lw  up  ν      p        np   res  res2
 1 0.0 1.2 16 0.2134 10.668607 1.632 2.664
 2 1.2 2.4 13 0.1678  8.392223 1.591 2.530
 3 2.4 3.6  8 0.1320  6.601557 0.544 0.296
χ² = 8.8705  (df = 3)  p = 0.0311

χ² (сложная) = 0.6866  (df = 2)  p = 0.7094

NP-тест:   S = 152.91,  c = 154.79  ⇒  Сохранить H₀
```

Рис. 6: Программный результат (задача 2)

Заключение

В заключение получены следующие результаты:

Задача №1

Выборочное среднее $= 0.7000$, Выборочная дисперсия $= 1.1939$, Медиана $= 0$,
Выборочная асимметрия $= 2.1022$, Выборочный эксцесс $= 8.3810$, $P[0, 1.79] = 0.84$;

$\hat{\lambda}_{\text{MLE}} = \hat{\lambda}_{\text{MOM}} = 0.7000$, $CI_{99.8\%} = [0.3344;\ 1.0656]$;

$\chi^2_{\text{простая}} = 1.4129$, $p = 0.4934$, $\chi^2_{\text{сложная}} = 2.9862$, $p = 0.2247$;

Тест Неймана–Пирсона: $S = 35$, $c = 47 \;\Rightarrow\; H_0$ сохраняется.

Задача №2

Выборочное среднее $= 3.0582$, Выборочная дисперсия $= 9.5891$, Медиана $= 2.105$,
Выборочная асимметрия $= 1.7645$, Выборочный эксцесс $= 6.3811$, $P[c, d] = 0.24$;

$\hat{\lambda}_{\text{MLE}} = \hat{\lambda}_{\text{MOM}} = 0.3269$, Bias $= 0.0066$,
$CI_{99.9\%} = [0.1748;\ 0.4791]$;

Колмогорова: $D = 0.2569$, $p = 0.0021$;
$\chi^2_{\text{простая}} = 8.8705$, $p = 0.0311$;
$\chi^2_{\text{сложная}} = 0.6866$, $p = 0.7094$;

Тест Неймана–Пирсона: $S_{\text{набл}} = 152.91$, $c_{\text{кр}} = 154.79$, решение — «H_0 сохраняется».

Здесь $S = \sum_{i=1}^{n} X_i$ — статистика критерия, c — критическое значение, выбираемое так, чтобы при H_0 вероятность попасть в критическую область была равна α.

Приложение А

Скрипт задачи №1

```
  if (is.null(getOption("repos")[["CRAN"]]) ||
  getOption("repos")[["CRAN"]] == "@CRAN@") {
options(repos = c(CRAN = "https://cloud.r-project.org"))
}

pkgs <- c("moments", "ggplot2")
to.install <- setdiff(pkgs, rownames(installed.packages()))
if (length(to.install)) install.packages(to.install)

suppressPackageStartupMessages({
library(moments)
library(ggplot2)
})

if (!dir.exists("fig")) dir.create("fig")
theme_set(theme_bw())


## 1
x1 <- c(0,0,2,1,0,0,0,0,0,1,0,1,0,0,0,2,0,2,1,0,
        1,1,0,1,1,3,0,0,0,0,0,1,0,0,0,0,4,1,5,2,
        0,0,2,0,0,1,1,0,0,1)

n1          <- length(x1)
a1 <- 0;   b1 <- 1.79
lambda0.1 <- 0.60;   lambda1.1 <- 1.40
alpha1      <- 0.002

freq1 <- table(factor(x1, levels = 0:max(x1)))
df.freq1 <- data.frame(k = as.integer(names(freq1)),
                       n.k = as.integer(freq1),
                       rel = as.numeric(freq1) / n1)

##  ECDF
F.n1 <- ecdf(x1)
png("fig/emp_dist_1.png", 800, 600)
plot(F.n1, main = "ECDF * zadacha 1", verticals = TRUE,
   do.points = FALSE, lwd = 2)
dev.off()

##  Gistogramma
png("fig/hist_1.png", 800, 600)
hist(x1, breaks = seq(-0.5, max(x1) + 0.5, by = 1),
   col = "#FDBE85", border = "grey20",
   main = "Gistogramma * zadacha 1")
dev.off()

## 1.2  vyborochnye kharakteristiki
m1        <- mean(x1)
s2.1      <- var(x1)
skew1 <- (sqrt(n1*(n1-1))/(n1-2)) * skewness(x1)
kurt1 <- ((n1-1)/((n1-2)*(n1-3))) * ((n1+1)*kurtosis(x1) + 6)
p.ab.1   <- mean(x1 >= a1 & x1 <= b1)

## 1.3  otsenki lambda
lambda.hat.1 <- m1                 # MLE = MOM
bias.lambda.1 <- 0
```

```
## 1.4  doveritelnyi interval
z.a <- qnorm(1 - alpha1 / 2)
ci.low.1  <- lambda.hat.1 - z.a * sqrt(lambda.hat.1 / n1)
ci.high.1 <- lambda.hat.1 + z.a * sqrt(lambda.hat.1 / n1)

## 1.5  khi^2 (prostaya)
obs.simple <- c(sum(x1 == 0), sum(x1 == 1), sum(x1 >= 2))
exp.simple <- c(dpois(0, lambda0.1),
                dpois(1, lambda0.1),
                1 - dpois(0, lambda0.1) - dpois(1, lambda0.1)) * n1
chi2.simple <- sum((obs.simple - exp.simple)^2 / exp.simple)
p.simple.1  <- pchisq(chi2.simple, df = 2, lower.tail = FALSE)

chi2.table.1 <- data.frame(
i   = 1:3,
lw  = c("k=0", "k=1", "k>=2"),
np  = round(exp.simple / n1, 4),
nu  = obs.simple,
p   = round(exp.simple / n1, 4),
np_r= exp.simple,
res = round((obs.simple - exp.simple) / sqrt(exp.simple), 3),
res2= round((obs.simple - exp.simple)^2 / exp.simple, 3)
)

## 1.6  khi^2 (slozhnaya)
exp.comp <- dpois(0:5, lambda.hat.1) * n1
obs.comp <- c(freq1["0"], freq1["1"], freq1["2"],
              sum(freq1[c("3","4","5")]))
exp.comp <- c(exp.comp[1:3], sum(exp.comp[4:6]))

chi2.comp <- sum((obs.comp - exp.comp)^2 / exp.comp)
p.comp.1  <- pchisq(chi2.comp, df = length(obs.comp) - 2,
                    lower.tail = FALSE)

labels.comp <- c("k=0", "k=1", "k=2", "k>=3")[1:length(obs.comp)]

chi2.table.1c <- data.frame(
i     = 1:length(obs.comp),
lw    = labels.comp,
np    = round(exp.comp / n1, 4),
nu    = obs.comp,
p     = round(exp.comp / n1, 4),
np_r  = exp.comp,
res   = round((obs.comp - exp.comp) / sqrt(exp.comp), 3),
res2  = round((obs.comp - exp.comp)^2 / exp.comp, 3)
)

## 1.7  Neimana - Pirsona
S.1   <- sum(x1)
c.np.1 <- qpois(1 - alpha1, lambda0.1 * n1)
decision.np.1 <- ifelse(S.1 >= c.np.1,
                      "Otvergnyt H0", "Sokhranit H0")

cat(sprintf("n = %d,  Sum x = %d\nx = %.4f,  s^2 = %.4f,  med = 0\n",
          n1, sum(x1), m1, s2.1))
cat(sprintf("Skew = %.4f,  Kurt(ex) = %.4f,  P[%g,%g] = %.2f\n",
          skew1, kurt1, a1, b1, p.ab.1))
cat(sprintf("lambda^ = %.4f (MLE=MOM),  CI_99.8%% = [%.4f; %.4f]\n",
          lambda.hat.1, ci.low.1, ci.high.1))
```

```
cat("\n--- khi^2 tablica (prostaya), pervye 3 stroki --------\n")
print(chi2.table.1, row.names = FALSE)
cat(sprintf("khi^2 = %.4f  (df = 2)  p = %.4f\n",
            chi2.simple, p.simple.1))

cat("\n--- khi^2 tablica (slozhnaya), pervye 3 stroki -------\n")
print(chi2.table.1c, row.names = FALSE)
cat(sprintf("khi^2 = %.4f  (df = 2)  p = %.4f\n",
            chi2.comp, p.comp.1))

cat(sprintf("\nNP-test:  S = %d,  c = %d  =>  %s\n",
            S.1, c.np.1, decision.np.1))
```

Приложение В

Скрипт задачи №2

```
  if (is.null(getOption("repos")[["CRAN"]]) ||
  getOption("repos")[["CRAN"]] == "@CRAN@") {
options(repos = c(CRAN = "https://cloud.r-project.org"))
}

pkgs <- c("moments", "ggplot2")
to.install <- setdiff(pkgs, rownames(installed.packages()))
if (length(to.install)) install.packages(to.install)

suppressPackageStartupMessages({
library(moments)
library(ggplot2)
})

if (!dir.exists("fig")) dir.create("fig")
theme_set(theme_bw())


#2
x2 <- c(10.34, 2.18, 8.80, 2.28, 1.95, 0.85, 3.73, 10.26, 5.01, 0.70,
      2.38, 0.25, 0.45, 0.31, 1.73, 2.67, 1.00, 1.59, 14.28, 2.14,
      1.85, 0.67, 2.70, 2.07, 5.31, 6.37, 3.24, 3.27, 1.31, 2.75,
      6.06, 1.05, 0.86, 2.43, 0.03, 3.70, 0.11, 1.06, 6.28, 0.55,
      9.07, 6.52, 0.94, 2.61, 0.89, 1.67, 0.24, 1.68, 3.34, 1.38)

n2 <- length(x2)
h       <- 1.20
c2      <- 2.40
d2      <- 6.00
lambda0.2 <- 0.20
lambda1.2 <- 0.33
alpha2    <- 0.001

bins2 <- seq(0, ceiling(max(x2)/h)*h, by = h)

## ECDF
png("fig/emp_dist_2.png", 800, 600)
plot(ecdf(x2), main = "ECDF * zadacha 2", verticals = TRUE,
   do.points = FALSE, lwd = 2)
dev.off()

## Gistogramma s poligonom
hist2 <- hist(x2, breaks = bins2, plot = FALSE)
png("fig/hist_2.png", 800, 600)
h2 <- hist(x2, breaks = bins2, col = "#B2DF8A",
         border = "grey20", main = "Gistogramma * zadacha 2")
lines((bins2[-1] + bins2[-length(bins2)])/2, h2$counts,
    type = "b", pch = 19, lwd = 2)
dev.off()

## 2.1 vyborochnye kharakteristiki
## 2.1 vyborochnye kharakteristiki
m2      <- mean(x2)
s2.2    <- var(x2)

skew2   <- (sqrt(n2*(n2-1))/(n2-2)) * skewness(x2)          # b1
kurt2   <- ((n2-1)/((n2-2)*(n2-3))) * ((n2+1)*kurtosis(x2) + 6)   # b2
```

```
p.cd.2 <- mean(x2 >= c2 & x2 <= d2)


## 2.2  otsenki lambda
lambda.hat.2  <- 1 / m2
bias.lambda.2 <- lambda.hat.2 / (n2 - 1)

## 2.3  doveritelnyi interval
z.b <- qnorm(1 - alpha2 / 2)
ci.low.2  <- lambda.hat.2 - z.b * lambda.hat.2 / sqrt(n2)
ci.high.2 <- lambda.hat.2 + z.b * lambda.hat.2 / sqrt(n2)

## 2.4  Kolmogorov (prostaya)
ks2 <- ks.test(x2, "pexp", rate = lambda0.2)

##  tablica (pervye 3 stroki)
xs2      <- sort(x2)
F.emp.u  <- (1:n2) / n2
F.emp.l  <- (0:(n2-1)) / n2
F.theor  <- pexp(xs2, rate = lambda0.2)
delta.l  <- abs(F.theor - F.emp.l)
delta.r  <- abs(F.theor - F.emp.u)
delta.m  <- pmax(delta.l, delta.r)
KS.head <- head(data.frame(
i    = 1:n2,
lw   = F.emp.l,
np   = F.theor,
nu   = F.emp.u,
p    = delta.l,
np_r = delta.r,
res  = delta.m,
res2 = delta.m^2
), 3)

## 2.5  khi^2 (prostaya)
prob.exp0 <- diff(pexp(bins2, rate = lambda0.2))
exp2      <- prob.exp0 * n2
obs2      <- hist2$counts

combine <- function(obs, exp) {
o <- obs; e <- exp; i <- 1
while (i <= length(e)) {
  if (e[i] < 5) {
    if (i == 1) {
      e[2] <- e[2] + e[1];  o[2] <- o[2] + o[1]
      e <- e[-1];  o <- o[-1]
    } else {
      e[i-1] <- e[i-1] + e[i];  o[i-1] <- o[i-1] + o[i]
      e <- e[-i];  o <- o[-i];  i <- i - 1
    }
  }
  i <- i + 1
}
list(obs = o, exp = e)
}
tmp <- combine(obs2, exp2)
obs2.c <- tmp$obs;  exp2.c <- tmp$exp

chi2.simple.2 <- sum((obs2.c - exp2.c)^2 / exp2.c)
p.simple.2    <- pchisq(chi2.simple.2, df = length(obs2.c)-1,
```

```r
                    lower.tail = FALSE)

chi2.table.2 <- data.frame(
i     = 1:length(obs2.c),
lw    = head(bins2, -1),
up    = tail(bins2, -1),
nu    = obs2.c,
p     = round(exp2.c / n2, 4),
np    = exp2.c,
res   = round((obs2.c - exp2.c) / sqrt(exp2.c), 3),
res2 = round((obs2.c - exp2.c)^2 / exp2.c, 3)
)
chi2.head.2 <- head(chi2.table.2, 3)

## 2.6  khi^2 (slozhnaya)
prob.exphat <- diff(pexp(bins2, rate = lambda.hat.2))
tmp  <- combine(obs2, prob.exphat * n2)
obs2.comp <- tmp$obs;  exp2.comp <- tmp$exp
chi2.comp.2 <- sum((obs2.comp - exp2.comp)^2 / exp2.comp)
p.comp.2     <- pchisq(chi2.comp.2, df = length(obs2.comp)-2,
                    lower.tail = FALSE)

## 2.7  Neimana - Pirsona
S.2   <- sum(x2)
c.np.2 <- qgamma(alpha2, shape = n2, scale = 1/lambda0.2)
decision.np.2 <- ifelse(S.2 >= c.np.2,
                       "Otvergnyt H0", "Sokhranit H0")

cat(sprintf("n = %d,  Sum x = %.2f\nx = %.4f,  s^2 = %.4f,  med = %.3f\n",
          n2, sum(x2), m2, s2.2, median(x2)))
cat(sprintf("Skew = %.4f,  Kurt(ex) = %.4f,  P[%g,%g] = %.2f\n",
          skew2, kurt2, c2, d2, p.cd.2))
cat(sprintf("lambda^ = %.4f,  bias ~= %.5f,  CI_99.9%% = [%.4f; %.4f]\n",
          lambda.hat.2, bias.lambda.2, ci.low.2, ci.high.2))

cat("\n--- tablica Kolmogorova (pervye 3 stroki) ------------\n")
print(KS.head, row.names = FALSE)
cat(sprintf("D = %.4f,  p-value = %.5f\n",
          ks2$statistic, ks2$p.value))

cat("\n--- khi^2 tablica (prostaya), pervye 3 stroki --------\n")
print(chi2.head.2, row.names = FALSE)
cat(sprintf("khi^2 = %.4f  (df = %d)  p = %.4f\n",
          chi2.simple.2, length(obs2.c)-1, p.simple.2))

cat(sprintf("\nkhi^2 (slozhnaya) = %.4f  (df = %d)  p = %.4f\n",
          chi2.comp.2, length(obs2.comp)-2, p.comp.2))

cat(sprintf("\nNP-test:  S = %.2f,  c = %.2f  =>  %s\n",
          S.2, c.np.2, decision.np.2))
```