Студент: Киреев Константин
Группа: 8383
Вариант: 8
Дата: 4 декабря 2020 г.

Статистический анализ

Индивидуальное домашнее задание №2

В результате эксперимента получены данные, приведенные в таблицах 1 и 2.

Таблица 1.

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

Таблица 2.

$\alpha_1 = 0.10$	$a = 0.00$	$b = 1.37$	$\lambda_0 = 0.70$	$\lambda_1 = 1.40$

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

Решение.

Вариационный ряд:
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3 4
Построим таблицу частот для выборки.

Таблица 3.

x_j	0	1	2	3	4
m_j	25	18	5	1	1
p_j^*	$^1/_2$	$^9/_{25}$	$^1/_{10}$	$^1/_{50}$	$^1/_{50}$

Построим эмпирическую функцию распределения по полученным данным:

$$F(x) = \begin{cases} 0, x \leqslant 0 \\ 0.5, 0 < x \leqslant 1 \\ 0.86, 1 < x \leqslant 2 \\ 0.96, 2 < x \leqslant 3 \\ 0.98, 3 < x \leqslant 4 \\ 1, x > 4 \end{cases}$$

График представлен на рис. 1:

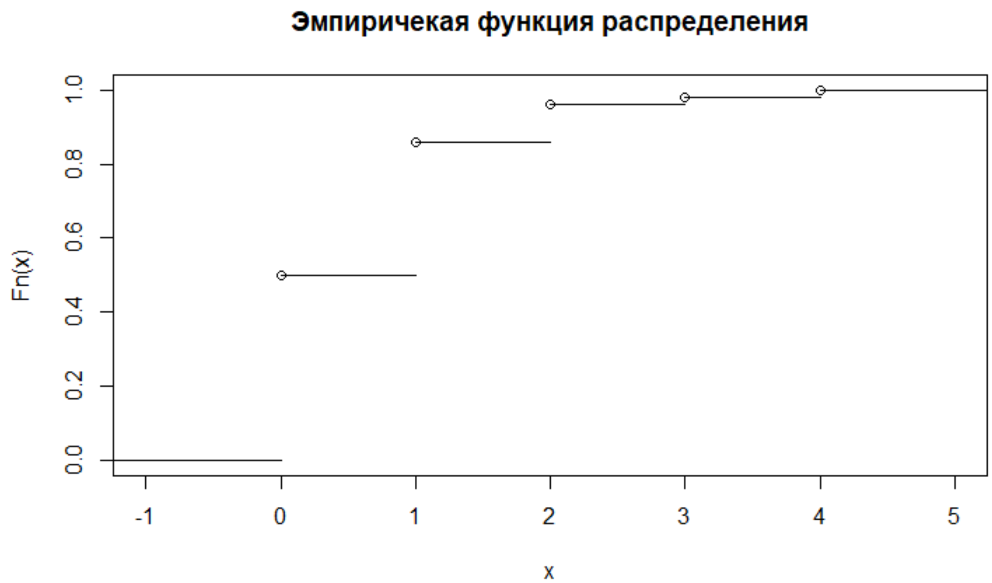

Рис. 1. Эмпирическая функция распределения

Гистограмма частот представлена на рис. 2:

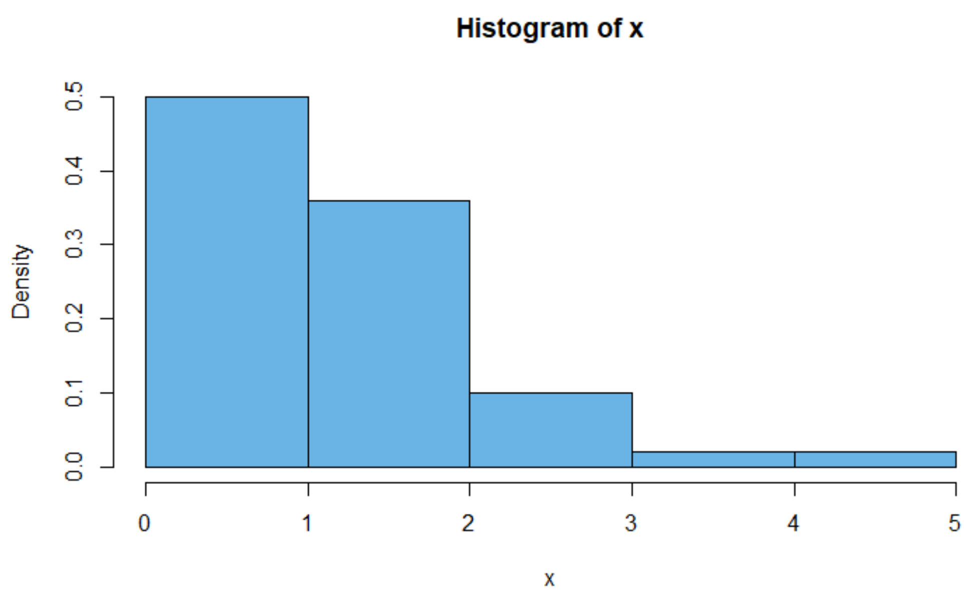

Рис. 2. Гистограмма частот

$\square$

Задача 2. *Вычислить выборочные аналоги следующих характеристик:*

- *математического ожидания*
- *дисперсии*
- *медианы*
- *ассиметрии*
- *эксцесса*
- *вероятности* $P(X \in [a,b])$

Решение.

- Математическое ожидание:

$$\bar{x}_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n} x_i m_i = \frac{1}{50}\cdot(18+10+7) = \frac{35}{50} = 0.7$$

- Дисперсия:

$$D_{\text{в}} = \bar{x^2} - \bar{x}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{50}\cdot(18+20+9+16) - 0.49 = \frac{63}{50} - 0.49 = 1.26 - 0.49 = 0.77$$

- Медиана:

$$Me = {}^{1}/_{2}$$

- Ассиметрия:

$$As = \frac{\mu_3^*}{\sigma^3} = 1.509608$$

- Эксцесс:

$$Ex = \frac{\mu_4^*}{\sigma^4} - 3 = 2.633496$$

- Вероятность:

$$P(x \in [a,b]) = P(x \in [0.00, 1.37]) = F(1.37) - F(0.00) = 0.86$$

$\square$

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

- Метод максимального правдоподобия

$$l(x, \lambda) = \frac{\lambda^{\sum\limits_1^n x_i} \cdot \exp(-\lambda n)}{\prod\limits_1^n x_i!} \Rightarrow ll(\bar{x}, \lambda) = \sum_1^n x_i ln\lambda - n\lambda - \sum_1^n ln x_i! \Rightarrow \frac{\partial ll(\bar{x}, \lambda)}{\partial \lambda} = \frac{1}{\lambda}\sum_1^n x_i - n = 0 \Rightarrow$$

$$\Rightarrow \hat{\lambda} = \frac{1}{n}\sum_1^n x_i = \bar{x}_{\text{в}} = 0.7$$

- Метод моментов

$$\begin{aligned}
&P(x, \theta) \\
&M_1^* = \bar{x}_{\text{в}}; \mathbb{E}X = \bar{x}_{\text{в}}; \\
&\mathbb{E}X = \int_{\mathbb{R}} xp(x, \theta)dx = \varphi(\theta) \\
&M_1 = \mathbb{E}X = \lambda; M_1^* = \bar{x}_{\text{в}} \Rightarrow \underline{\hat{\lambda} = \bar{x}_{\text{в}} = 0.7}
\end{aligned}$$

Чтобы найти смещение оценки, найдем:

$$\mathbb{E}\hat{\lambda} = \frac{1}{n}\sum_1^n \mathbb{E}x_i = \frac{n\lambda}{n} = \lambda \Rightarrow \text{оценки несмещенные}$$

$\square$

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

Решение.

$$\hat{\lambda} = \bar{x_{\text{в}}} = 0.7; \alpha_1 = 0.10; \gamma = 1 - \alpha_1 = 0.90;$$

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

$$t_\gamma : \phi(t_\gamma) = 1 - \frac{\alpha_1}{2} \Rightarrow t_\gamma = 1.645$$

$$P(-t_\gamma \leqslant \frac{\sqrt{n}(\bar{x_{\text{в}}} - \lambda)}{\sqrt{\lambda}} \leqslant t_\gamma) \longrightarrow 1 - \alpha$$

$$n(\bar{x_{\text{в}}} - \lambda)^2 = t_\gamma^2 \lambda$$

$$\lambda^2 - 2\lambda(\bar{x_{\text{в}}} + \frac{t_\gamma^2}{2n}) + \bar{x_{\text{в}}} = 0$$

$$D = \frac{t_\gamma^2}{n}(\bar{x_{\text{в}}} + \frac{t_\gamma^2}{4n})$$

$$\lambda_{1,2} = \bar{x_{\text{в}}} + \frac{t_\gamma^2}{2n} \pm t_\gamma \sqrt{\frac{1}{n}(\bar{x_{\text{в}}} + \frac{t_\gamma^2}{4n})} = 0.7 + 0.027 \pm 0.1965 \Rightarrow [0.5305; 0.9235]$$

□

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.

$$\hat{\lambda_0} = 0.70; \alpha_1 = 0.10;$$

m_i – эмпирические частоты;

$m_i^{'}$ – выравнивающие частоты; $m_i^{'} = n \cdot p_i$

Простая гипотеза H_0 имеет вид:

$$H_0 : p(x) = \frac{\lambda_0^x}{x!}\exp(-\lambda_0)$$

Тогда искомая вероятность примет вид:

$$p_k = P(X = k) = \frac{0.7^k}{k!}\exp(-0.7)$$

Построим таблицу оценки методом χ^2.

Таблица 4.

x_i	0	1	2	3	4	$\sum$
m_i	25	18	5	1	1	50
p_i	0.497	0.348	0.122	0.029	0.005	1
$m_i^{'}$	25	17	6	1	1	50
$m_i - m_i^{'}$	0	1	-1	0	0	0
$\frac{(m_i - m_i^{'})^2}{m_i^{'}}$	0	$\frac{1}{17} = 0.059$	$\frac{1}{6} = 0.167$	0	0	$\chi^2_{\text{набл}}$

Итого

$$\chi^2_{\text{набл}} = \sum_1^k \frac{(m_i - m_i^{'})^2}{m_i^{'}} = 0.226$$

$$l = k - r - 1 = 5 - 1 - 1 = 3$$

$\chi^2_{\text{кр}} = \chi^2_{\alpha;l} = \chi^2_3$ на уровне значимости 0.1 = 6.251

$\chi^2_{\text{набл}} < \chi^2_{\text{кр}} \Rightarrow$ гипотеза H_0 принимается, выборка принадлежит распределению Пуассона. Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = 0.8914

□

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:

$$P < -function(a)\{$$
$$p < -0$$
$$p[1] < -ppois(0, a)$$
$$p[2] < -ppois(1, a) - sum(p)$$
$$p[3] < -ppois(2, a) - sum(p)$$
$$p[4] < -ppois(3, a) - sum(p)$$
$$p[5] < -1 - sum(p)$$
$$p\}$$
$$X2 < -function(a)\{g < -n \cdot P(a); f < -(nu - g)^2/g; sum(f)\}$$
$$nu < -c(25, 17, 6, 1, 1)$$
$$XM < -nlm(X2, 0.70)$$

В результате вычислений получим, что $\chi^2_{\text{набл}} = 1.5265 < \chi^2_{\text{крит}} = 4.6$ Таким образом, гипотеза принимается. Наибольшее значение уровня значимости, при котором еще нет оснований отвергнуть данную гипотезу = 0.74 □

Задача 7. *Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром* $\lambda = \lambda_0 = 0.70$ *при альтернативе пуассоновости с параметром* $\lambda = \lambda_1 = 1.40$. *Проверить гипотезу на уровне значимости* α_1. *Что получится, если поменять местами основную и альтернативную гипотезы?*

Решение.

Сформулируем гипотезы.
$H_0 : \lambda = \lambda_0 = 0.70$
$H_1 : \lambda = \lambda_1 = 1.40$
По лемме Неймана-Пирсона:

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

$$l(\bar{x}, 0.7, 1.4) = \frac{L(\bar{x}, 1.4)}{L(\bar{x}, 0.7)} = 2^{\sum x_i} \cdot \exp(n * (\lambda_0 - \lambda_1)) = 2^{\sum x_i} \cdot \exp(-0.7n)$$

$$ll(\bar{x}, \lambda_0, \lambda_1) = -\sum x_i \cdot ln2 - 0.7n < lnC$$

$$\sum x_i > \frac{-\,lnC - 0.7n}{ln2}$$

$$\hat{C} = \frac{-\,lnC - 0.7n}{ln2}$$

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 $\hat{C}$ и p из уравнения:

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

$$= P_{\lambda_0}(\sum_1^n x_i > \hat{C}) + p \cdot P_{\lambda_0}(\sum_1^n x_i = \hat{C}) = \alpha_1 = 0.1$$

$$x_i \to P_{ois}(\lambda_0)$$

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

Подбором среди целых чисел можно найти такое наибольшее $\hat{C}$ и α_0, что

$$\alpha_0 = P_{\lambda_0}(\sum_1^n x_i > \hat{C}) = 1 - P_{n\lambda_0}(\hat{C}) - p_{n\lambda_0}(\hat{C}) < \alpha_1$$

$$p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_0}(\sum_1^n x_i = A)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_0}(A)}$$

В результате расчета получим: $\alpha_0 = 0.09867$; $\hat{C} = 41$; $p = 0.03499$

$$\sum_1^n x_i = 35$$

$$35 < 41 \Rightarrow \text{ Таким образом, принимаем гипотезу} H_0$$

Теперь поменяем местами основную и альтернативную гипотезы.

$H_0 : \lambda = \lambda_1 = 1.40$
$H_1 : \lambda = \lambda_0 = 0.70$

$$l(\bar{x}, 1.4, 0.7) = \frac{L(\bar{x}, 0.7)}{L(\bar{x}, 1.4)} = (\frac{1}{2})^{\sum x_i} \cdot \exp(n * (\lambda_1 - \lambda_0)) = (\frac{1}{2})^{\sum x_i} \cdot \exp(0.7n)$$

$$ll(\bar{x}, \lambda_0, \lambda_1) = -\sum x_i \cdot ln(\frac{1}{2}) + 0.7n < lnC$$

$$\sum x_i < \frac{lnC - 0.7n}{ln(\frac{1}{2})}$$

$$\hat{C} = \frac{lnC - 0.7n}{ln(\frac{1}{2})}$$

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 из уравнения:

$$P_{\lambda_1}(\sum_1^n x_i > \hat{C}) + p \cdot P_{\lambda_1}(\sum_1^n x_i = \hat{C}) = \alpha_1 = 0.1$$

$$p = \frac{\alpha_1 - \alpha_0}{P_{\lambda_1}(\sum_1^n x_i = A)} = \frac{\alpha_1 - \alpha_0}{p_{n\lambda_1}(A)}$$

В результате расчета получим: $\alpha_0 = 0.081593$; $\hat{C} = 58$; $p = 1.049973$

$$\sum_1^n x_i = 35$$

$$35 < 58 \Rightarrow \text{ Таким образом, отвергаем гипотезу} H_0$$

При замене основной и альтернативной гипотезы меняется также 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:

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

- Метод максимального правдоподобия

$$l(\bar{x},\lambda)=\prod_1^n \frac{\lambda^{x_i}}{(\lambda+1)^{x_i+1}}=\frac{\lambda^{\sum\limits_1^n x_i}}{(\lambda+1)^{\sum\limits_1^n x_i+n}}$$

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

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

$$\frac{\partial ll}{\partial\lambda}=0\to\hat{\lambda}=\frac{1}{n}\sum_1^n x_i=\bar{x}=0.7$$

- Метод моментов

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

Чтобы найти смещение оценки, найдем:

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

$\square$

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

Решение.

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

$$\hat{I} = -\frac{\partial^2 ll}{\partial\lambda^2}(\hat{\lambda}) = -\frac{\partial^2 ll}{\partial\lambda^2}(\hat{X}) = n(\frac{1}{\bar{X}} - \frac{1}{\bar{X}+1}) = 42.017$$

$$\sigma^2(\hat{\lambda}) = \hat{I}^{-1} = 0.024$$

$$\sigma = \sqrt{\hat{I}^{-1}} = 0.154$$

Доверительный интервал будет иметь вид

$$[\hat{\lambda} - x_\alpha\sigma, \hat{\lambda} + x_\alpha\sigma]$$

$$x_\alpha = \phi^{-1}(1 - \frac{\alpha_1}{2}) = 1.645$$

Получен доверительный интервал $[0.4467, 0.9534]$

$\square$

Задача 11. *Используя гистограмму частот, построить критерий значимости* χ^2 *проверки простой гипотезы согласия с геометрическим распределением с параметром* $\lambda_0 = 0.70$. *Проверить гипотезу на уровне значимости* $\alpha_1 = 0.10$. *Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.*

Решение.

$$H_0 : X_1, ..., X_n \sim Geom\left(\frac{1}{0.7+1}\right) = Geom\left(\frac{1}{1.7}\right)$$

Построим таблицу оценки методом χ^2.

Таблица 5.

x_i	0	1	2	3	4	$\sum$
m_i	25	18	5	1	1	50
p_i	0.59	0.242	0.099	0.040	0.017	1
$m_i^{'}$	29	12	5	2	2	50
$m_i - m_i^{'}$	-4	6	0	-1	-1	0
$\frac{(m_i - m_i^{'})^2}{m_i^{'}}$	$\frac{16}{29} = 0.55$	$\frac{36}{12} = 3$	0	$\frac{1}{2} = 0.5$	$\frac{1}{2} = 0.5$	$\chi^2_{\text{набл}}$

Итого

$$\chi^2_{\text{набл}} = \sum_1^k \frac{(m_i - m_i^{'})^2}{m_i^{'}} = 4.55$$

$$l = k - r - 1 = 5 - 1 - 1 = 3$$

$$\chi^2_{\text{кр}} = \chi^2_{\alpha;l} = \chi^2_3 \text{ на уровне значимости } 0.1 = 6.251$$

$\chi^2_{\text{набл}} < \chi^2_{\text{кр}} \Rightarrow$ гипотеза H_0 принимается, выборка принадлежит геометрическому распределению. Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = 0.79213 □

Задача 12. *Построить критерий значимости χ^2 проверки сложной гипотезы согласия с геометрическим распределением. Проверить гипотезу на уровне значимости $\alpha_1 = 0.10$. 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:

$$\underset{\lambda}{argmin} \sum_1^k \frac{(m_i - np_i(\lambda))^2}{np_i(\lambda)}$$

Задача реализована в R следующим скриптом:

$$P < -function(a)\{$$
$$p < -0$$
$$p[1] < -pgeom(0, a)$$
$$p[2] < -pgeom(1, a) - sum(p)$$
$$p[3] < -pgeom(2, a) - sum(p)$$
$$p[4] < -pgeom(3, a) - sum(p)$$
$$p[5] < -1 - sum(p)$$
$$p\}$$
$$X2 < -function(a)\{g < -n \cdot P(a); f < -(nu - g)^2/g; sum(f)\}$$
$$nu < -c(25, 17, 6, 1, 1)$$
$$XM < -nlm(X2, 1/(1 + 0.70))$$

Получили оптимальную $\hat{\lambda} = \dfrac{1}{0.56559} - 1 = 0.768$

Построим таблицу оценки методом χ^2.

Таблица 6.

x_i	0	1	2	3	4	$\sum$
m_i	25	18	5	1	1	50
p_i	0.566	0.246	0.107	0.046	0.020	1
$m_i^{'}$	28	13	6	2	1	50
$m_i - m_i^{'}$	-3	5	-1	-1	0	0
$\dfrac{(m_i - m_i^{'})^2}{m_i^{'}}$	$\dfrac{9}{28} = 0.32$	$\dfrac{25}{13} = 1.92$	$\dfrac{1}{6} = 0.17$	$\dfrac{1}{2} = 0.5$	0	$\chi^2_{\text{набл}}$

В результате вычислений получим, что $\chi^2_{\text{набл}} = 2.91 < \chi^2_{\text{крит}} = 4.6$

Таким образом, гипотеза принимается. Наибольшее значение уровня значимости, при котором еще нет оснований отвернуть данную гипотезу = 0.572998 $\square$