МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МОЭВМ

ИДЗ № 2
по дисциплине «Математическая статистика»

Студентка гр. 5381 ________________ Кочнева О.Р.

Преподаватель ________________ Чирина А.В.

Санкт-Петербург

2017

Постановка задачи:

Вар. 8 (538117)

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 $\alpha_1 = 0.20$; $a = 0.00$; $b = 0.78$; $\lambda_0 = 0.50$; $\lambda_1 = 1.50$.

1 0 0 0 1 1 0 0 0 3 1 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 2 2 1 1 0 1 0 0 2 0 1 0 0 1 2 0 0 0
0 0 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 $\alpha_2 = 0.20$; $c = -1.40$; $d = -0.80$; $h = 0.20$; $a_0 = -1.00$; $\sigma_0 = 0.50$; $a_1 = -1.50$; $\sigma_1 = 0.50$.

−1.251 −1.600 −1.615 −1.205 −1.118 −0.097 −1.224 −1.550 −0.539 −0.876 −0.975 −1.245 −0.853 −0.284 −1.366 −1.135
−1.106 −0.308 −2.044 −0.794 −1.945 −0.979 −1.043 −0.374 −1.062 −1.586 −1.314 −0.821 −0.970 0.062 −0.760 −0.669
−2.306 −1.275 −1.072 −1.025 −0.923 −1.543 −0.954 −0.552 −0.934 −0.918 −0.420 −0.017 −0.469 −1.199 −0.372 −1.429
−0.753 −0.793

Ход работы.

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

- Построение вариационного ряда:

Начальные данные

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

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

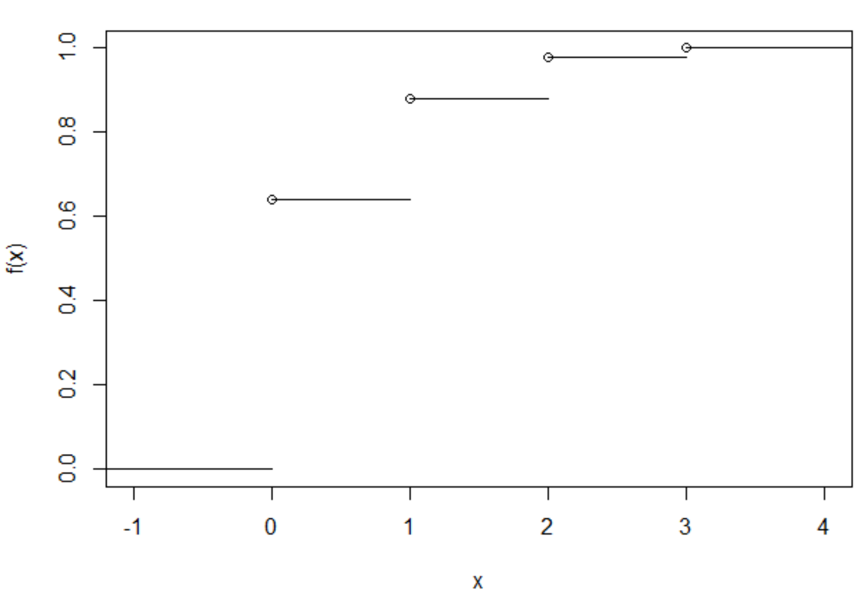
Таблица частот на основе вариационного ряда

Значение X_i	0	1	2	3
Частота	32	12	5	1

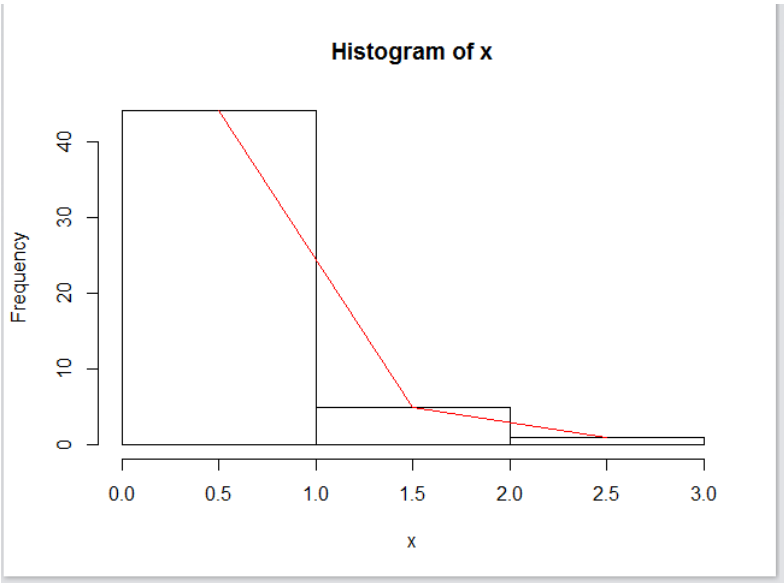
- Построение эмпирической функции распределения:

Эмпирическая функция распределения: $F(X) = \frac{1}{n}\sum_{i=1}^{n} ll\{x_i < X\}$

$$F = \begin{cases} 0, \text{если } x \le 0 \\ 0.64, \text{если } x \in (0,1] \\ 0.88, \text{если } x \in (1,2] \\ 0.98, \text{если } x \in (2,3] \\ 1, \text{если } x \ge 3 \end{cases}$$



- Гистограмма частот:



b) Вычислить выборочные аналоги следующих числовых характеристик:

1) Математическое ожидание: выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

   i=0.5

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

   ii=0.57

3) Медиана: - выборочная медиана, равная выборочной квантили порядка p:

$$\mu = \begin{cases} X_{[np]+1}, np \neq Z \\ [X_{np,n}, X_{np+1,n}], np \in Z \end{cases}$$

   iii=0

4) Асимметрия – выборочная асимметрия: $As = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$

   iv=1.3942446

5) Эксцесс – выборочный эксцесс: $Exc = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3$

   v=1.32041

6) Вероятность попадания в заданный промежуток: P(X ∈ [a, b]).
   vi=0.64

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.

$p(x)=\frac{\lambda^x}{x!}*e^{-\lambda}$ - плотность распределения Пуассона.

Метод максимального правдоподобия:

$$L(\vec{x};\lambda)=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}*e^{-n*\lambda} \;=>\; LL(\vec{x};\lambda)=\ln(\lambda)*\sum_{i=1}^{n}x_i-\ln(\prod_{i=1}^{n}x_i!)-n*\lambda \;=>\; \frac{\partial LL}{\partial\lambda}=\frac{\sum_{i=1}^{n}x_i}{\lambda}-n$$

$$\frac{\partial LL}{\partial\lambda}=0 \;=>\; \hat{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}=\bar{x}$$

Метод моментов:

математическое ожидание: $E(x_1)=\lambda$, выборочный средний момент: $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$

$$=>\; \hat{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}.$$

$E(\hat{\lambda})=E(\bar{x})=\frac{1}{n}*\sum_{i=1}^{n}E(x_i)=\frac{1}{n}*n*\lambda=\lambda$, значит $\hat{\lambda}=\bar{x}$ - несмещенная оценка.

Оценка максимального правдоподобия: 0.5

d) Построить асимптотический доверительный интервал уровня значимости α_1= 0.20 для параметра λ на базе оценки максимального правдоподобия.

Так как x_i имеет распределение Пуассона, то $I_1(\lambda)=\frac{1}{\lambda}$ => $I(\lambda)=n*I_1(\lambda)=\frac{n}{\lambda}$

По методу максимального правдоподобия:

$$\hat{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}=\bar{x}\;,\quad \sqrt{I(\lambda)}*(\hat{\lambda}-\lambda)\Rightarrow N(0,1)$$

Эксперимент регулярен, значит, подстановка ОМП вместо параметра в информацию Фишера не нарушает асимптотической нормальности.

$$\sqrt{n*I_1(\hat{\lambda})}*(\hat{\lambda}-\lambda)\Rightarrow N(0,1)$$

$$\sqrt{\frac{n}{\bar{x}}}*(\bar{x}-\lambda)\Rightarrow N(0,1);\quad \alpha_1=0.10$$

$$p(T_1(\vec{x})\le\lambda\le T_2(\vec{x}))=1-\alpha_1$$

$$p(-x_\alpha\le\sqrt{\frac{n}{\bar{x}}}*(\bar{x}-\lambda)\le x_\alpha)=\Phi(x_\alpha)-\Phi(-x_\alpha)=2*\Phi(x_\alpha)-1=1-\alpha_1,$$

где $\Phi(x_\alpha)=\frac{1}{\sqrt{2*\pi}}\int_{-\infty}^{x_\alpha}e^{-\frac{t^2}{2}}dt$ - квантиль порядка x_α стандартного нормального закона распределения.

$$x_\alpha = \Phi^{-1}(1-\frac{\alpha_1}{2}).$$

$$p(\bar{x} - x_\alpha * \sqrt{\frac{\bar{x}}{n}} \le \lambda \le \bar{x} + x_\alpha * \sqrt{\frac{\bar{x}}{n}}) = 1-\alpha_1.$$

Полученный результат: [0.3718448 0.6281552]

e) Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением Пуассона с параметром λ_0=0.50. Проверить гипотезу на уровне значимости α_1=0.20. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Простая гипотеза H_o: $p(x) = \frac{\lambda_o^{\ x}}{x!} * e^{-\lambda_0}$, λ_o=0.5

$$\varphi(\underset{X}{\rightarrow}) = \begin{cases} 0, \ X^2 \le x_\alpha \\ 1, \ \ X^2 > x_\alpha \end{cases}$$

Таблица частот на основе вариационного ряда

Значение X_i	0	1	2	3
Частота	32	12	5	1

	n_k	p_k	np_k	$n_k - np_k$	$\frac{(n_k - np_k)^2}{np_k}$
(-Inf; 0]	32	0.6065307	30.32653	-29.72	0.09234461
(0; 1]	12	0.3032653	15.16327	-14.86	0.65990101
(1; 2]	5	0.07581633	3.790817	-3.715	0.38570171
(2; +Inf]	1	0.01263606	0.6318028	-0.6191667	0.21457519
x_α = 4.641628		$\sum p_k$ =0.9982		$X^2=\sum\frac{(n_k-np_k)^2}{np_k}$= 1.352523	

Итак, получили, что $\chi^2 < x_{\alpha,}$ следовательно, принимаем гипотезу $H_{o.}$

Для нахождения наибольшего значения уровня значимости, на котором ещё нет оснований отвергнуть данную гипотезу, вычисляем функцию распределения χ^2_{r-1} в точке X^2, и вычитаем полученное значение из единицы:

Наибольшее значение уровня значимости: 0.7167005

f) Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением Пуассона. Проверить гипотезу на уровне значимости α_1=0.20. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

$$H_0 : X_1, \dots, X_n \sim Pois(\lambda)$$
$$\sum_{k=1}^{r} \frac{(n_k - np_k(\lambda))^2}{np_k(\lambda)} \to \chi^2_{r-d-1}$$

Метод минимизации хи-квадрат

$$\underset{\lambda}{\operatorname{argmin}} \sum_{k=1}^{r} \frac{(n_k - np_k(\lambda))^2}{np_k(\lambda)}$$

```
minimum : num 0.51
estimate : num 1.23
gradient : num -2.34e-09
code : int 1
```

Получили оптимальную $\hat{\lambda} = 1.57$ и $\sum_{k=1}^{r} \frac{(n_k - np_k(\hat{\lambda}))^2}{np_k(\hat{\lambda})} = 0.51$

$t_{1-\alpha_1, r-d-1}$ – квантиль распределения хи-квадрат с $r - d - 1$ степенями свободы уровня $1 - \alpha_1$, где d – размерность оценки, $d = dim(\lambda) = 1$

$$t_{1-\alpha_1, r-d-1} = t_{0.2,2} = 3.218876$$
$$\sum_{k=1}^{r} \frac{\left(n_k - np_k(\hat{\lambda})\right)^2}{np_k(\hat{\lambda})} = 0.51 < 3.218876$$

=> Принимаем гипотезу H_0

`0.7749163` – наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

g) Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром $\lambda = \lambda_0$ =0.50 приальтернативе пуассоновости с параметром $\lambda = \lambda_1$=1.50. Проверить гипотезу на уровне значимости α_1=0.20. Что получится, если поменять местами основную и альтернативную гипотезы?

$$H_0 : X_1, \dots, X_n \sim Pois(0.50)$$
$$H_1 : X_1, \dots, X_n \sim Pois(1.50)$$
$$L_0(\vec{x}) = \frac{\left(\frac{1}{2}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-\left(\frac{1}{2}\right)n}$$
$$L_1(\vec{x}) = \frac{\left(\frac{3}{2}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-\left(\frac{3}{2}\right)n}$$

$$l\left(\vec{x}, \frac{1}{2}, \frac{3}{2}\right) = \frac{L_1(\vec{x})}{L_0(\vec{x})} = \frac{3^{\sum_{i=1}^{n} x_i} e^{-\left(\frac{3}{2}\right)n}}{1^{\sum_{i=1}^{n} x_i} e^{-\left(\frac{1}{2}\right)n}} = 3^{\sum_{i=1}^{n} x_i} e^{-n}$$

Наиболее мощный критерий:

$$\varphi(\underset{X}{\to}) = \begin{cases} 0, \ 3^{\sum X_i} * e^{-n} < C \\ p, 3^{\sum X_i} * e^{-n} = C \\ 1, 3^{\sum X_i} * e^{-n} > C \end{cases}$$

Логарифмируем соотношение $3^{\sum X_i} * e^{-n} < C$

Получим: $\sum X_i * \ln(3) - n \ < \ln(C)$

$$\sum X_i < \frac{\ln C + n}{\ln 3}$$

Обозначим через: $\tilde{C} = \frac{\ln C + n}{\ln 3}$

Отыщем $\tilde{C}$ и p из уравнения:

$$P_{\lambda_0}\left(l(\vec{x},\lambda_0,\lambda_1) > C\right) + pP_{\lambda_0}\left(l(\vec{x},\lambda_0,\lambda_1) = C\right) = P_{\lambda_0}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) + pP_{\lambda_0}\left(\sum_{i=1}^{n} x_i = \tilde{C}\right) = \alpha_1$$

$x_i \Rightarrow Pois(\lambda_0)$, следовательно, $\sum_{i=1}^{n} x_i \Rightarrow Pois(n\lambda_0)$

Т.к. $\sum_{i=1}^{n} x_i \in N_0$, то подбором (в цикле с помощью R) среди целых чисел можем найти такое наибольшее $\tilde{C}$ (а после и α_0), что:

$$\alpha_0 = P_{\lambda_0}\left(\sum_{i=1}^{n} x_i > \tilde{C}\right) = 1 - P_{n\lambda_0}\left(\tilde{C}\right) - p_{n\lambda_0}(\tilde{C}) < \alpha_1$$

Тогда $p = \dfrac{\alpha_1 - \alpha_0}{P_{\lambda_0}\left(\sum_{i=1}^{n} x_i = A\right)} = \dfrac{\alpha_1 - \alpha_0}{p_{n\lambda_0}(A)}$

Проведём вычисления в R

p = 0.4210315

c = 28

sum(x) = 25<28-гипотеза принимается

Тогда критерий:

$$\varphi(\underset{X}{\to}) = \begin{cases} 0, \ \sum X_i < 48 \\ 0.42, \sum X_i = 48 \\ 1, \sum X_i > 48 \end{cases}$$

<u>Теперь поменяем местами основную и альтернативную гипотезу.</u>

$$H_0: X_1, \dots, X_n \sim Pois(1.50)$$
$$H_1: X_1, \dots, X_n \sim Pois(0.50)$$
$$L_0(\vec{x}) = \frac{\left(\frac{3}{2}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-\left(\frac{3}{2}\right)n}$$
$$L_1(\vec{x}) = \frac{\left(\frac{1}{2}\right)^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-\left(\frac{1}{2}\right)n}$$

$$l\left(\underset{x}{\rightarrow},\frac{3}{2},\frac{1}{2}\right)=\frac{L_1(\vec{x})}{L_0(\vec{x})}=\frac{1^{\sum_{i=1}^n x_i}e^{-\left(\frac{1}{2}\right)n}}{3^{\sum_{i=1}^n x_i}e^{-\left(\frac{3}{2}\right)n}}=\left(\frac{1}{3}\right)^{\sum_{i=1}^n x_i}e^n$$

Наиболее мощный критерий:

$$\varphi(\underset{X}{\rightarrow})=\begin{cases}0, & \left(\frac{1}{3}\right)^{\sum_{i=1}^n x_i}e^n<C\\ p, & \left(\frac{1}{3}\right)^{\sum_{i=1}^n x_i}e^n=C\\ 1, & \left(\frac{1}{3}\right)^{\sum_{i=1}^n x_i}e^n>C\end{cases}$$

Логарифмируем соотношение $\left(\frac{1}{3}\right)^{\sum_{i=1}^n x_i}e^n<C$

Получим: $\sum X_i*\ln\left(\frac{1}{3}\right)+n\ <\ln(C)$

$$\sum X_i<-\frac{\ln C-n}{\ln 3}$$

Обозначим через: $\tilde{C}=-\frac{\ln C+n}{\ln 3}$

Отыщем $\tilde{C}$ и p из уравнения:

$$P_{\lambda_0}\left(l(\vec{x},\lambda_0,\lambda_1)>C\right)+pP_{\lambda_0}\left(l(\vec{x},\lambda_0,\lambda_1)=C\right)=P_{\lambda_0}\left(\sum_{i=1}^n x_i>\tilde{C}\right)+pP_{\lambda_0}\left(\sum_{i=1}^n x_i=\tilde{C}\right)=\alpha_1$$

$x_i\Rightarrow Pois(\lambda_0)$, следовательно, $\sum_{i=1}^n x_i\Rightarrow Pois(n\lambda_0)$

Т.к. $\sum_{i=1}^n x_i\in N_0$, то подбором (в цикле с помощью R) среди целых чисел можем найти такое наибольшее $\tilde{C}$ (а после и α_0), что:

$$\alpha_0=P_{\lambda_0}\left(\sum_{i=1}^n x_i>\tilde{C}\right)=1-P_{n\lambda_0}\left(\tilde{C}\right)-p_{n\lambda_0}(\tilde{C})<\alpha_1$$

Тогда $p=\dfrac{\alpha_1-\alpha_0}{P_{\lambda_0}\left(\sum_{i=1}^n x_i=A\right)}=\dfrac{\alpha_1-\alpha_0}{p_{n\lambda_0}(A)}$

Проведём вычисления в R

p = 0.4210315

c = 28

sum(x) = 25<28- Принимаем гипотезу H_0

Тогда критерий:

$$\varphi(\underset{X}{\rightarrow})=\begin{cases}0, & \sum X_i<48\\ 0.42, & \sum X_i=\ 48\\ 1, & \sum X_i>\ 48\end{cases}$$

Проведём вычисления в R

p = 0.3181148

c = 67
sum(x) = 25<67- Принимаем альтернативу H_0

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

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

$$p_\lambda(x=k)=\frac{\lambda^k}{(\lambda+1)^{k+1}}=\left(\frac{\lambda}{\lambda+1}\right)^k\left(\frac{1}{\lambda+1}\right);k=0,1,...$$

$$q=\left(\frac{\lambda}{\lambda+1}\right);$$

$$p=\left(\frac{1}{\lambda+1}\right);k=0,1,...$$

Обозначим

Найдём оценку максимального правдоподобия:

$$L(\vec{x},\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{(\lambda+1)^{x_i+1}}=\left(\frac{\lambda}{\lambda+1}\right)^{\sum_{i=1}^{n}x_i}\left(\frac{1}{\lambda+1}\right)^n$$

$$\ln L(\vec{x},\lambda)=\ln(\lambda)\sum_{i=1}^{n}x_i-\ln(\lambda+1)\sum_{i=1}^{n}x_i-n\ln(\lambda+1)$$

$$\frac{\partial\ln L(\vec{x},\lambda)}{\partial\lambda}=\left(\frac{1}{\lambda}-\frac{1}{\lambda+1}\right)\sum_{i=1}^{n}x_i-\frac{n}{\lambda+1}=0$$

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

Для геометрического распределения математическое ожидание:

$E(x_1)=\frac{1-p}{p}=\lambda$, выборочный средний момент: $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$ => $\hat{\lambda}=\frac{\sum_{i=1}^{n}x_i}{n}$.

$E(\hat{\lambda})=E(\bar{x})=\frac{1}{n}*\sum_{i=1}^{n}E(x_i)=\frac{1}{n}*n*\lambda=\lambda$, значит $\hat{\lambda}=\bar{x}$ - несмещенная оценка.

Оценка максимального правдоподобия: 0.5

Построить асимптотический доверительный интервал уровня значимости α_1 = 0.20 для параметра λ на базе оценки максимального правдоподобия.

$x_i->Geom(\lambda)$

Найдём информацию Фишера.

$$f(x,\lambda)=\frac{\lambda^x}{(\lambda+1)^{x+1}}$$

$$\ln f(x,\lambda)=x\ln\lambda-(x+1)\ln(\lambda+1)$$

$$\frac{\partial^2 \ln f(x,\lambda)}{\partial\lambda^2}=-\frac{x}{\lambda^2}+\frac{x+1}{(\lambda+1)^2}$$

$$E(x)=\lambda$$

$$I_1(\lambda)=-E_\lambda\left(\frac{\partial^2 \ln f(x,\lambda)}{\partial\lambda^2}\right)=-E_\lambda\left(-\frac{x}{\lambda^2}+\frac{x+1}{(\lambda+1)^2}\right)=\frac{1}{\lambda}-\frac{1}{\lambda+1}=\frac{1}{\lambda(\lambda+1)}$$

$$I(\lambda)=nI_1(\lambda)=\frac{n}{\lambda(\lambda+1)}$$

ОМП $\hat{\lambda}$ параметра λ:

$$\hat{\lambda}=\overline{x}$$

$$\sqrt{I(\lambda)}\left(\hat{\lambda}-\lambda\right)\Rightarrow N(0;1)$$

Эксперимент регулярен, значит, подстановка ОМП вместо параметра в информацию Фишера не нарушает асимптотической нормальности.

$$\sqrt{nI_i(\hat{\lambda})}\left(\hat{\lambda}-\lambda\right)\Rightarrow N(0;1)$$

$$\sqrt{\frac{n}{\overline{x}(\overline{x}+1)}}\left(\overline{x}-\lambda\right)\Rightarrow N(0;1)$$

$$\alpha_1=0.1$$

$$p(T_1(\vec{x})<\lambda<T_2(\vec{x}))=1-\alpha_1$$

$$p(-b<\sqrt{\frac{n}{\overline{x}(\overline{x}+1)}}\left(\overline{x}-\lambda\right)<b)=\Phi(b)-\Phi(-b)=2\Phi(b)-1=1-\alpha_1$$

Где $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{\frac{-t^2}{2}}dt$, т.е. b - квантиль стандартного нормального распределения.

$$b=\Phi^{-1}(1-\frac{\alpha_1}{2})$$

$$p(\overline{x}-\frac{b\sqrt{\overline{x}(\overline{x}+1)}}{\sqrt{n}}<\lambda<\overline{x}+\frac{b\sqrt{\overline{x}(\overline{x}+1)}}{\sqrt{n}})=1-\alpha_1$$

Полученный результат: [0.3430426 0.6569574]

Используя гистограмму частот, построить критерий значимости χ^2 проверки простой гипотезы согласия с распределением Пуассона с параметром λ_0=0.50. Проверить гипотезу на уровне значимости α_1=0.20.

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

Простая гипотеза H_o: $p(x) = \frac{\lambda_o^{\,x}}{x!} * e^{-\lambda_0}$, λ_o=0.5

$$\varphi(\underset{X}{\rightarrow}) = \begin{cases} 0, & X^2 \le x_\alpha \\ 1, & X^2 > x_\alpha \end{cases}$$

Таблица частот на основе вариационного ряда

Значение X_i	0	1	2	3
Частота	32	12	5	1

	n_k	p_k	np_k	$n_k - np_k$	$\frac{(n_k - np_k)^2}{np_k}$
(-Inf; 0]	32	0.6666667	33.33333	-32.6666	0.05333333
(0; 1]	12	0.2222222	11.11111	-10.88889	0.07111111
(1; 2]	5	0.07407407	3.703704	-3.62963	0.45370370
(2; +Inf]	1	0.02469136	1.234568	-1.209877	0.04456790
x_α = 4.641628		$\sum p_k =$ 0.9876543		$X^2=\sum\frac{(n_k-np_k)^2}{np_k}=$ 0.622716	

Итак, получили, что $\chi^2 < x_\alpha$, следовательно, принимаем гипотезу H_o.

Для нахождения наибольшего значения уровня значимости, на котором ещё нет оснований отвергнуть данную гипотезу, вычисляем функцию распределения χ^2_{r-1} в точке X^2, и вычитаем полученное значение из единицы:
Наибольшее значение уровня значимости: 0.8912129

Построить критерий значимости χ^2 проверки сложной гипотезы согласия с распределением Пуассона. Проверить гипотезу на уровне значимости α_1=0.20. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

$$\sum_{k=1}^{r}\frac{(n_k - np_k(\lambda))^2}{np_k(\lambda)} \rightarrow \chi^2_{r-d-1}$$

Метод минимизации хи-квадрат

$$\underset{\lambda}{\operatorname{argmin}} \sum_{k=1}^{r} \frac{(n_k - np_k(\lambda))^2}{np_k(\lambda)}$$

```
minimum : num 7.38
estimate : num 0.44
gradient : num 6.57e-10
```

Получили оптимальную $\hat{\lambda} = 0.44$ и $\sum_{k=1}^{r} \frac{(n_k - np_k(\hat{\lambda}))^2}{np_k(\hat{\lambda})} = 7.38$

$t_{1-\alpha_1, r-d-1}$ – квантиль распределения хи-квадрат с $r - d - 1$ степенями свободы уровня $1 - \alpha_1$, где d – размерность оценки, $d = dim(\lambda) = 1$

$$t_{1-\alpha_1, r-d-1} = t_{0.2,2} = 3.218876$$

$$\sum_{k=1}^{r} \frac{\left(n_k - np_k(\hat{\lambda})\right)^2}{np_k(\hat{\lambda})} = 7.38 > 3.218876$$

=>Отвергаем гипотезу H_0

0.02493594 – наибольшее значение уровня значимости, на котором нет основан ий отвергнуть данную гипотезу.

Задание 2.

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

- Построение вариационного ряда:

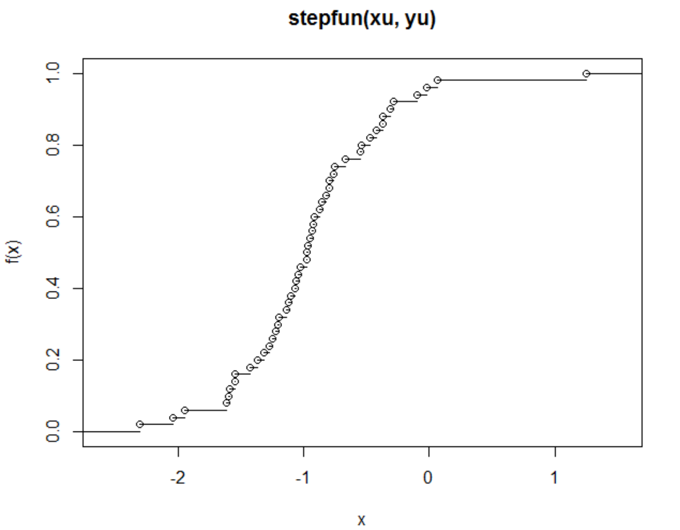
Начальные данные

```
[1] -1.251 -1.600 -1.615 -1.205 -1.118 -0.097 -1.224 -1.550 -0.539 -0.876 -0.975
[12]  1.245 -0.853 -0.284 -1.366 -1.135 -1.106 -0.308 -2.044 -0.794 -1.945 -0.979
[23] -1.043 -0.374 -1.062 -1.586 -1.314 -0.821 -0.970  0.062 -0.760 -0.669 -2.306
[34] -1.275 -1.072 -1.025 -0.923 -1.543 -0.954 -0.552 -0.934 -0.918 -0.420 -0.017
[45] -0.469 -1.199 -0.372 -1.429 -0.753 -0.793
```

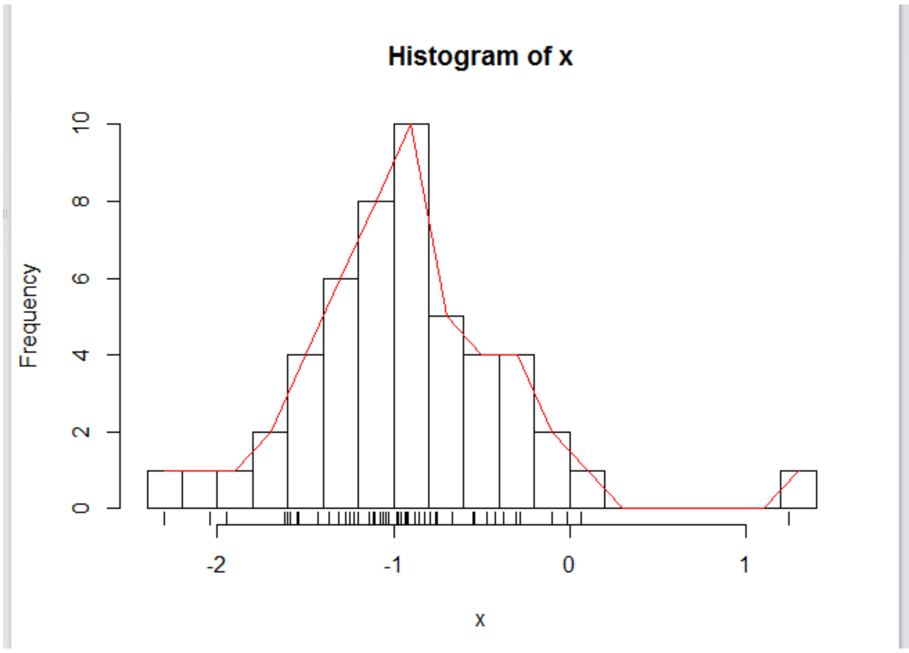
Вариационный ряд

```
[1] -2.306 -2.044 -1.945 -1.615 -1.600 -1.586 -1.550 -1.543 -1.429 -1.366 -1.314
[12] -1.275 -1.251 -1.224 -1.205 -1.199 -1.135 -1.118 -1.106 -1.072 -1.062 -1.043
[23] -1.025 -0.979 -0.975 -0.970 -0.954 -0.934 -0.923 -0.918 -0.876 -0.853 -0.821
[34] -0.794 -0.793 -0.760 -0.753 -0.669 -0.552 -0.539 -0.469 -0.420 -0.374 -0.372
[45] -0.308 -0.284 -0.097 -0.017  0.062  1.245
```

- Построение эмпирической функции распределения:



- Гистограмма частот:



b) Вычислить выборочные аналоги следующих числовых характеристик:

1) Математическое ожидание: выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

i=-0.9422

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

ii=0.34209876

3) Медиана: - выборочная медиана, равная выборочной квантили порядка р:

$$\mu = \begin{cases} X_{[np]+1}, np \neq Z \\ [X_{np,n}, X_{np+1,n}], np \in Z \end{cases}$$

iii=-0.97

4) Асимметрия – выборочная асимметрия: $As = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$

iv=0.76238987

5) Эксцесс – выборочный эксцесс: $Exc = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3$

v=2.607191535

6) Вероятность попадания в заданный промежуток:P(X ∈ [c, d]).

vi=0.48

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.

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

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

$$\frac{dLL(\vec{x}, a, \sigma^2)}{da} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} a\right) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - na\right)$$

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

$$\sum_{i=1}^{n} x_i - n\tilde{a} = 0$$

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

$$\frac{dLL(\vec{x}, a, \sigma^2)}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - a)^2$$

$$-\frac{n}{\tilde{\sigma}} + \frac{1}{\tilde{\sigma}^3}\sum_{i=1}^{n}(x_i - \tilde{a})^2 = 0$$

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{a})^2 = S_n^2$$

Оценка максимального правдоподобия:
$(\tilde{a}, \tilde{\sigma}^2) = (-0.9422, 0.3420988)$.

Оценка метода моментов:

$$EX = a, \quad DX = \sigma^2$$
$$\hat{a} = \bar{x}, \quad \hat{\sigma}^2 = S_n^2$$

$(\hat{a}, \hat{\sigma}^2) = (\tilde{a}, \tilde{\sigma}^2) = (-0.9422, 0.3420988)$.

Смещение оценок:

$$E\tilde{a} = E\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}E\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}\sum_{i=1}^{n} a = \frac{1}{n}na = a$$

$$y_i = x_i - a;$$

$$S_n^2(Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - a - \bar{x} + a)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S_n^2(X)$$

$$E\tilde{\sigma}^2 = E\frac{1}{n}\sum_{i=1}^{n}(x_i - a - \bar{x} + a)^2 = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n}(x_i - a)\right)^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}E(x_i - a)^2 - \frac{1}{n^2}\sum_{i=1}^{n}E(x_i - a)^2 - \frac{2}{n^2}\sum_{i<j\le n}Ey_iy_j$$

$$= \frac{1}{n}\sum_{i=1}^{n}E(x_i - a)^2 - \frac{1}{n^2}\sum_{i=1}^{n}E(x_i - a)^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \frac{1}{n^2} = \sigma^2 - \frac{\sigma^2}{n}$$

$$= \frac{n-1}{n}\sigma^2$$

$$\breve{\sigma}^2 = \tilde{S}_n^2 = \frac{n}{n-1}S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \tilde{a})^2$$

$$E\breve{\sigma}^2 = E\frac{n}{n-1}S_n^2 = \frac{n}{n-1}ES_n^2 = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

$(\tilde{a}, \breve{\sigma}^2) = (-0.9422, 0.3490804)$ – Несмещённая оценка

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

Построим доверительный интервал для a.

Согласно лемме Фишера $\sqrt{n-1}\frac{\bar{x}-a}{s} \Rightarrow S_{n-1}$

$$p\left(T_1 < \sqrt{n-1}\frac{\bar{x}-a}{s} < T_2\right) = S_{n-1}(T_2) - S_{n-1}(T_1) = 1 - \alpha_2$$

$t_\alpha < \sqrt{n-1}\frac{\bar{x}-a}{s} < t_\alpha$, где t_α - квантиль распределения Стьюдента уровня $1-\frac{\alpha_2}{2}$

$$\bar{x} - \frac{st_\alpha}{\sqrt{n-1}} < a < \bar{x} + \frac{st_\alpha}{\sqrt{n-1}},$$

Получили доверительный интервал для параметра a уровня доверия α_2.

[-2.0038627; 0.1194627]

Построим доверительный интервал для σ^2.

Согласно лемме Фишера $\frac{ns^2}{\sigma^2} \Rightarrow \chi^2_{n-1}$

Введём $x_{1\alpha}$ и $x_{2\alpha}$ - квантили распределения χ^2_{n-1} уровня $\frac{\alpha_2}{2}$ и $1-\frac{\alpha_2}{2}$ соответственно.

Тогда $p\left(\frac{ns^2}{x_{2\alpha}} < \sigma^2 < \frac{ns^2}{x_{1\alpha}}\right) = 1 - \alpha_2$

e) И использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с нормальным распределением с параметрами α_0, σ_0^2. Проверить гипотезу на уровне значимости α_2. Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.

$H_0 : a = a_0; \sigma^2 = \sigma_0^2$

Согласно теореме Колмогорова, при справедливости гипотезы $\sqrt{n}\sup|F_n(y) - F(y)| \Rightarrow K$, где K- распределение Колмогорова.

Обозначим $D_n = \sqrt{n}\sup|F_n(y) - F(y)|$

$$p(D_n < \lambda_{\alpha 2}) = 1 - \alpha_2$$

Согласно таблице распределения Колмогорова, **λ= 1,51**

Вычислим величину D_n = 0.3788687

Dn > λ, значит, не принимаем 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 ещё нет оснований отвергнуть данную гипотезу.

$$H_0: X_1, \dots, X_n \sim N(0.50, 0.50)$$
$$\sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k} \to \chi^2_{r-1}$$

	n_k	p_k	np_k	$n_k - np_k$	$\frac{(n_k - np_k)^2}{np_k}$
(-Inf; -2]	2	0.02275013	1.137507	-1.114756	0.65396972
(-2; -1]	21	0.47724987	23.862493	-23.385244	0.34337855
(-1; 0]	5	0.47724987	23.862493	-23.385244	0.05422406
(0; +Inf]	1	0.02275013	1.137507	-1.114756	0.65396972
x_α = 4.641628		$\sum p_k = 1$		$X^2 = \sum \frac{(n_k - np_k)^2}{np_k} =$ 1.705542	

Итак, получили, что $\chi^2 < x_\alpha$, следовательно, принимаем гипотезу H_0.

Для нахождения наибольшего значения уровня значимости, на котором ещё нет оснований отвергнуть данную гипотезу, вычисляем функцию распределения χ^2_{r-1} в точке X^2, и вычитаем полученное значение из единицы:

Наибольшее значение уровня значимости: 0.6357024

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 ещё нет оснований отвергнуть данную гипотезу.

$$H_0 : \theta \in \Theta \subset R^d ; d = 2; \theta = (a, \sigma^2)$$

Сложная гипотеза согласия: $H_0 : F(X) \equiv F(X, a, \sigma^2)$, где $F(X, a, \sigma^2)$ – функция нормального распределения с параметрами a, σ^2;

Поделим область на r=3 интервала, задав внутренние границы b=(-2;-1;0).

X^2- зависит от a, σ^2, т.к. величины p_i не фиксированы. Известно, что в случае регулярности эксперимента статистика $\tilde{X}^2(\theta) = \inf\limits_{\lambda \in \Theta} X^2(\theta)$ сходится по распределению к χ^2_{r-d-1}

$$t_{1-\alpha_2, r-d-1} = t_{0.8,1} = 1.642374$$

```
minimum : num 0.23
estimate : num -0.958
gradient : num 1.6e-07
```

Так как 0.23<1.642374, принимаем гипотезу на заданном уровне значимости. Наибольшее значение уровня значимости: 0.6318867

h) Построить наиболее мощный критерий проверки простой гипотезы о нормальности с параметрами $(a, \sigma^2) = (a_0, \sigma_0^2)$ 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?

$H_0 : (a, \sigma^2) = (a_0, \sigma_0{}^2) - \textit{основная}; H_1 : (a, \sigma^2) = (a_1, \sigma_1{}^2) - \textit{альтернативная}$

Согласно лемме Неймана-Пирсона, наиболее мощный критерий проверки гипотезы H_0 при альтернативе H_1 имеет вид:

$$\varphi(\vec{x}) = \begin{cases} 0, l(\vec{x}, a_0, \sigma_0{}^2, a_1, \sigma_1{}^2) < C \\ 1, l(\vec{x}, a_0, \sigma_0{}^2, a_1, \sigma_1{}^2) > C \end{cases}, \text{ где } C = const, C \geq 0$$

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

$$\left(\frac{\sigma_1}{\sigma_0}\right)^{-n} \exp\left(\sum_{i=1}^{n} \frac{(x_i - a_0)^2}{2\sigma_0{}^2} - \frac{(x_i - a_1)^2}{2\sigma_1{}^2}\right) = \left(\frac{\sigma_1}{\sigma_0}\right)^{-n} \exp\left(\left(\frac{1}{2\sigma_0{}^2} - \frac{1}{2\sigma_1{}^2}\right)\sum_{i=1}^{n} x_i^2 + \left(\frac{a_1}{\sigma_1{}^2} - \frac{a_0}{\sigma_0{}^2}\right)\sum_{i=1}^{n} x_i + \right.$$

$$\left. + \frac{n}{2}\left(a_0^2 - a_1^2\right)\right)$$

Т.к. по условию задачи $\sigma_1 = \sigma_0 = \sigma$, то $l(\vec{x}, a_0, \sigma_0{}^2, a_1, \sigma_1{}^2)$ примет вид:

$$l(\vec{x}, a_0, \sigma_0{}^2, a_1, \sigma_1{}^2) = \exp\left(\left(\frac{a_1 - a_0}{\sigma^2}\right)\sum_{i=1}^{n} x_i + \frac{n}{2}\left(a_0^2 - a_1^2\right)\right)$$

$$\exp\left(\left(\frac{a_1 - a_0}{\sigma^2}\right)\sum_{i=1}^{n} x_i + \frac{n}{2}\left(a_0^2 - a_1^2\right)\right) > C$$

$$\exp\left(\left(\frac{a_1 - a_0}{\sigma^2}\right)\sum_{i=1}^{n} x_i\right) > \exp\left(\frac{n}{2}\left(a_1^2 - a_0^2\right) + \ln(C)\right)$$

$$т.к. a_1 < a_0 \Rightarrow \sum_{i=1}^{n} x_i < \frac{\sigma^2 (n\left(a_1^2 - a_0^2\right) + \ln(C))}{2\left(a_1 - a_0\right)} = A$$

Значит, критерий можно переписать в виде:

$$\varphi(\vec{x}) = \begin{cases} 1, \sum_{i=1}^{n} x_i < A \\ 0, \sum_{i=1}^{n} x_i > A \end{cases}$$

i) Отыщем A из уравнения:

$$E\varphi(\vec{x}) = P\left(\sum_{i=1}^{n} x_i < A\right) = \alpha_2$$

$x_i \Rightarrow N(a_0, \sigma_0{}^2)$, следовательно, $\sum_{i=1}^{n} x_i \Rightarrow N(na_0, n\sigma_0{}^2)$

Тогда A - квантиль распределения $N(na_0, n\sigma_0{}^2)$ уровня α_2.

$\sum x$=-47.11>-52.97558=> принимаем гипотезу

Критерий построен.

$$\varphi(\underset{x}{\rightarrow}) = \begin{cases} 0, & \sum X_i < -52.97558 \\ 1, & \sum X_i > -52.97558 \end{cases}$$

Теперь поменяем местами основную и альтернативную гипотезу.

$H_0 : (a,\sigma^2) = (a_0,\sigma_0{}^2) - альтернативная; H_1 : (a,\sigma^2) = (a_1,\sigma_1{}^2) - основная$

Согласно лемме Неймана-Пирсона, наиболее мощный критерий проверки гипотезы H_1 при альтернативе H_0 имеет вид:

$$\varphi(\vec{x}) = \begin{cases} 0, l(\vec{x}, a_1, \sigma_1{}^2, a_0, \sigma_0{}^2) < C \\ 1, l(\vec{x}, a_1, \sigma_1{}^2, a_0, \sigma_0{}^2) > C \end{cases}, \text{ где } C = const, C \ge 0$$

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

$$\left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left(\sum_{i=1}^{n} \frac{(x_i - a_1)^2}{2\sigma_1{}^2} - \frac{(x_i - a_0)^2}{2\sigma_0{}^2}\right) = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left(\left(\frac{1}{2\sigma_1{}^2} - \frac{1}{2\sigma_0{}^2}\right)\sum_{i=1}^{n} x_i^2 + \left(\frac{a_0}{\sigma_0{}^2} - \frac{a_1}{\sigma_1{}^2}\right)\sum_{i=1}^{n} x_i + \right.$$

$$\left. + \frac{n}{2}\left(a_1^2 - a_0^2\right)\right)$$

Т.к. по условию задачи $\sigma_1 = \sigma_0 = \sigma$, то $l(\vec{x}, a_0, \sigma_0{}^2, a_1, \sigma_1{}^2)$ примет вид:

$$l(\vec{x}, a_0, \sigma_0{}^2, a_1, \sigma_1{}^2) = \exp\left(\left(\frac{a_0 - a_1}{\sigma^2}\right)\sum_{i=1}^{n} x_i + \frac{n}{2}\left(a_1^2 - a_0^2\right)\right)$$

$$\exp\left(\left(\frac{a_0 - a_1}{\sigma^2}\right)\sum_{i=1}^{n} x_i + \frac{n}{2}\left(a_1^2 - a_0^2\right)\right) > C$$

$$\exp\left(\left(\frac{a_0 - a_1}{\sigma^2}\right)\sum_{i=1}^{n} x_i\right) > \exp\left(\frac{n}{2}\left(a_0^2 - a_1^2\right) + \ln(C)\right)$$

$$т.к. a_1 > a_0 \Rightarrow \sum_{i=1}^{n} x_i > \frac{\sigma^2 (n\left(a_1^2 - a_0^2\right) + \ln(C))}{2\left(a_1 - a_0\right)} = A$$

Значит, критерий можно переписать в виде:

$$\varphi(\vec{x}) = \begin{cases} 1, \sum_{i=1}^{n} x_i > A \\ 0, \sum_{i=1}^{n} x_i < A \end{cases}$$

Отыщем A из уравнения:

$$E\varphi(\vec{x}) = P\left(\sum_{i=1}^{n} x_i > A\right) = 1 - P\left(\sum_{i=1}^{n} x_i < A\right) = \alpha_2$$

$x_i \Rightarrow N(a_1, \sigma_1{}^2)$, следовательно, $\sum_{i=1}^{n} x_i \Rightarrow N(na_1, n\sigma_1{}^2)$

Тогда A - квантиль распределения $N(na_0, n\sigma_0{}^2)$ уровня $1-\alpha_2$.

Проведём вычисления в R.

$\sum x$=-47.11>-72.02442=> принимаем альтернативу

Критерий построен.

$$\varphi(\underset{X}{\rightarrow}) = \begin{cases} 0, & \sum X_i < -72.02442 \\ 1, & \sum X_i > -72.02442 \end{cases}$$

j) В пунктах (c) – (g) заменить семейство нормальных распределений на двухпараметрическое семейство распределений Лапласа с плотностями

$$p(x) = \frac{1}{\sqrt{2}\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x-a|} = \frac{\alpha}{2} e^{-\alpha|x-\beta|}$$

$$\alpha = \frac{\sqrt{2}}{\sigma}; \quad \beta = a$$

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

Найдём оценку максимального правдоподобия:

$$L(\vec{x}, a, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2}} e^{\frac{-\sqrt{2}|x_i - a|}{\sigma}} = \sigma^{-n} 2^{-\frac{n}{2}} e^{-\sum_{i=1}^{n} \frac{-\sqrt{2}|x_i - a|}{\sigma}}$$

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

$$\begin{cases} \dfrac{\partial \ln L(\vec{x}, a, \sigma)}{\partial a} = -\dfrac{\sqrt{2}}{\sigma}\sum_{i=1}^{n}(-1)sign(x_i - a) = 0 \\ \dfrac{\partial \ln L(\vec{x}, a, \sigma)}{\partial \sigma} = -\dfrac{n}{\sigma} + \dfrac{\sqrt{2}}{\sigma^2}\sum_{i=1}^{n}|x_i - a| = 0 \end{cases}$$

$$\begin{cases} \hat{a} = z_{1/2} = -0.641 \\ \hat{\sigma}^2 = \left(\dfrac{\sqrt{2}}{n}\sum_{i=1}^{n}\left|x_i - z_{1/2}\right|\right)^2 = 2.377604 \end{cases}$$

d)

Найдём оценку методом моментов

$$\begin{cases}\mu_1(a,\sigma^2)=a\\ \mu_2(a,\sigma^2)=\sigma^2+a^2\end{cases}$$

$$\left.\begin{cases}\overline{x}=a\\ \dfrac{1}{n}\sum_{i=1}^{n}x_i^{\ 2}=\dfrac{4}{\sigma^2}+a^2\end{cases}\right| \Rightarrow \begin{cases}\tilde{a}=\overline{x}=-0.9422\\ \tilde{\sigma}^2=\dfrac{4}{\dfrac{1}{n}\sum_{i=1}^{n}x_i^{\ 2}-\left(\overline{x}\right)^2}=\dfrac{4}{s^2}=0.3420988\end{cases}$$

d) Построить асимптотические доверительные интервалы уровня значимости α_2 для параметров (a,σ^2) на базе оценки максимального правдоподобия.

$$x_i->DE(a,\sigma)$$

$$I_i(\sigma)=I_i(a)=\frac{2}{\sigma^2};$$

$$I(a)=I(\sigma)=nI_i(a)=\frac{2n}{\sigma^2};$$

ОМП $\hat{a}$ параметра a

$\hat{a}=z_{1/2}$

ОМП $\hat{\sigma}^2=\left(\dfrac{\sqrt{2}}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|\right)^2$

$$\sqrt{I(a)}\left(\hat{a}-a\right)\Rightarrow N(0;1)$$

$$\sqrt{I(\sigma)}\left(\hat{\sigma}-\sigma\right)\Rightarrow N(0;1)$$

Эксперимент регулярен, значит, подстановка ОМП вместо параметра в информацию Фишера не нарушает асимптотической нормальности.

$$\sqrt{I(\hat{a})}\left(\hat{a}-a\right)\Rightarrow N(0;1)$$

$$\sqrt{I(\hat{\sigma})}\left(\hat{\sigma}-a\right)\Rightarrow N(0;1)$$

$$\sqrt{\frac{2n}{\left(\dfrac{\sqrt{2}}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|\right)^2}}\left(z_{1/2}-a\right)\Rightarrow N(0;1)$$

$$\sqrt{\frac{2n}{\left(\dfrac{\sqrt{2}}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|\right)^2}}\left(\frac{\sqrt{2}}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|-\sigma\right)\Rightarrow N(0;1)$$

$$\frac{n\left(z_{1/2}-a\right)}{\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|}\Rightarrow N(0;1)$$

$$\frac{\left(\sqrt{2}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|-n\sigma\right)}{\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|}\Rightarrow N(0;1)$$

$\alpha_2 = 0.2$

$$p(T_1(\vec{x})<a<T_2(\vec{x}))=1-\alpha_2$$

$$p(T_3(\vec{x})<\sigma<T_4(\vec{x}))=1-\alpha_2$$

$$p(-b<\frac{n\left(z_{1/2}-a\right)}{\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|}<b)=\Phi(b)-\Phi(-b)=2\Phi(b)-1=1-\alpha_2$$

$$p(-b<\frac{\left(\sqrt{2}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|-n\sigma\right)}{\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|}<b)=\Phi(b)-\Phi(-b)=2\Phi(b)-1=1-\alpha_2$$

Где $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{\frac{-t^2}{2}}dt$, т.е. b - квантиль стандартного нормального распределения.

$$b=\Phi^{-1}(1-\frac{\alpha_2}{2})$$

$$p(z_{1/2}-\frac{b}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|<a<z_{1/2}+\frac{b}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|)=1-\alpha_2$$

$$p\left(\frac{(-b+\sqrt{2})}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|<\sigma<\frac{(b+\sqrt{2})}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|\right)=1-\alpha_2$$

$$p\left(\left(\frac{(-b+\sqrt{2})}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|\right)^2<\sigma^2<\left(\frac{(b+\sqrt{2})}{n}\sum_{i=1}^{n}\left|x_i-z_{1/2}\right|\right)^2\right)=1-\alpha_2$$

Получили асимптотические доверительные интервалы для параметров (a,σ^2) уровня доверия α_2

Для a $[-1.5037406; -0.4362594]$

Для σ^2 $[0.003052681; 1.260527729]$