Sheet

Иерусалимов Никита

ИДЗ№2

Вар. 6 (938221)

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 $\alpha_1 = 0.02$; $a = 4.55$; $b = 5.45$; $\lambda_0 = 7.00$; $\lambda_1 = 5.00$.
3 8 6 1 6 4 1 6 6 5 3 5 4 6 4 6 3 6 5 10 5 3 2 5 8 4 5 13 7 6 4 3 1 5 5 9 1 3 7 5 6 2 4 6 6 6 6 7 5 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 $f(x) = \frac{1}{\sqrt{2}\,\sigma} e^{-\frac{\sqrt{2}}{\sigma}|x-a|}$.

Таблица 2 $\alpha_2 = 0.10$; $c = 2.92$; $d = 3.06$; $h = 0.05$; $a_0 = 3.00$; $\sigma_0 = 0.10$; $a_1 = 2.88$; $\sigma_1 = 0.10$.
2.877 3.162 3.071 2.939 3.055 2.928 3.053 2.979 2.845 3.042 3.043 3.109 3.102 3.001 3.021 2.894 3.103 2.971 3.069 2.977 3.038 2.998 2.896 3.065 2.829 3.025 2.879 3.036 3.001 2.979 3.007 2.754 3.061 3.015 2.892 3.068 2.998 3.128 3.035 3.159 3.091 3.156 3.004 2.813 2.983 2.951 2.944 3.127 3.041 3.278

```
#install lib and include
```

Ход решения

Задание 1

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

```
x<-c(3, 8, 6, 1, 6, 4, 1, 6, 6, 5, 3, 5, 4, 6, 4, 6, 3, 6, 5, 10, 5, 3, 2, 5, 8, 4, 5, 13, 7, 6, 4, 3
6, 7, 5, 2)

VariationalSeries<-sort(x)
cat("Вариационный ряд: ",VariationalSeries)
```

```
Вариационный ряд:  1 1 1 1 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6
```

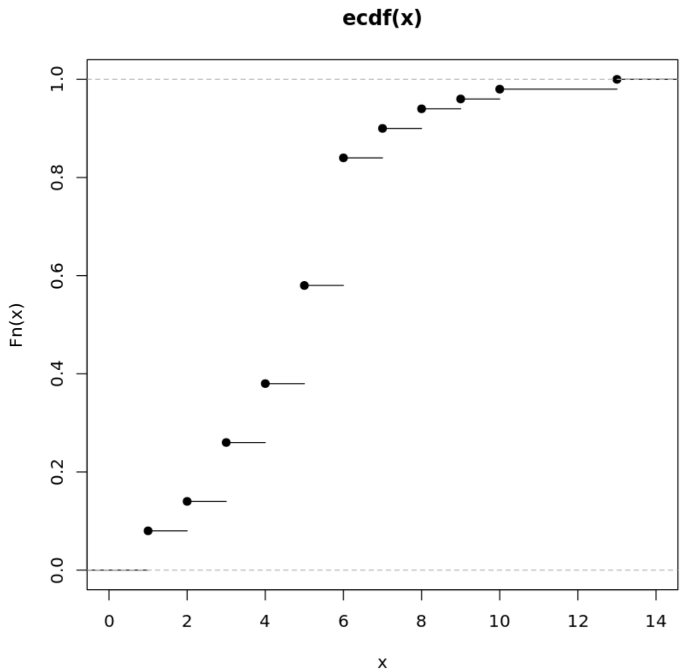
Формула эмпирической функции распределения:
$\frac{1}{n}\sum_{i=1}^{n}$

Построение эмпирической функции распределения:

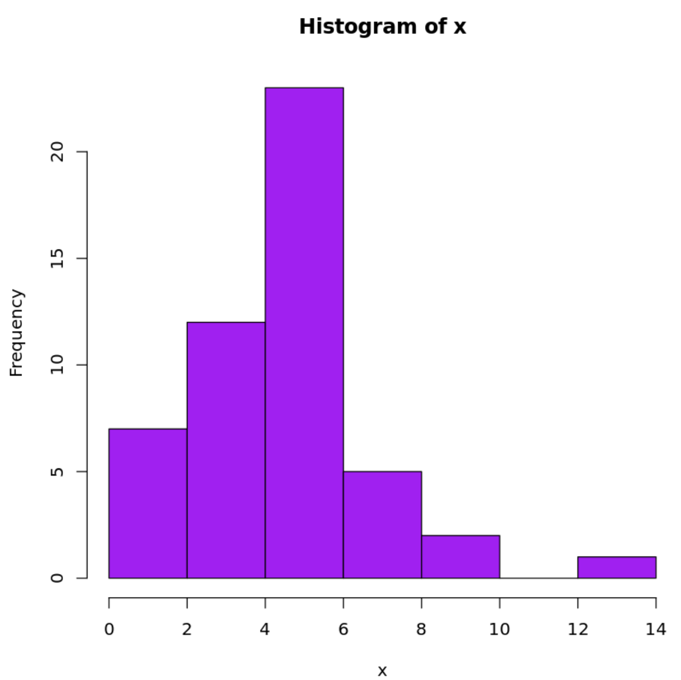
```
#создаем и вычисляем эмпирической функцию распределения
F<- function(x,t){z<-x[x<t]; length(z)/length(x)}
cat("F(X)=\n |",F(x,1),"    ,    X<=1   |\n",
"|",F(x,2)," ,   1<X<=2  |\n",
"|",F(x,3)," ,   2<X<=3  |\n",
"|",F(x,4)," ,   3<X<=4  |\n",
"|",F(x,5)," ,   4<X<=5  |\n",
"|",F(x,6)," ,   5<X<=6  |\n",
"|",F(x,7)," ,   6<X<=7  |\n",
"|",F(x,8),"  ,   7<X<=8  |\n",
"|",F(x,9)," ,   8<X<=9  |\n",
"|",F(x,10)," ,  9<X<=10  |\n",
"|",F(x,11)," ,  10<X<=13 |\n",
"|",F(x,13.000001),"    ,    X>13   |\n")



#Строим график
plot(ecdf(x))
```

```
F(X)=
 | 0     ,    X<=1   |
 | 0.08  ,   1<X<=2  |
 | 0.14  ,   2<X<=3  |
 | 0.26  ,   3<X<=4  |
 | 0.38  ,   4<X<=5  |
 | 0.58  ,   5<X<=6  |
 | 0.84  ,   6<X<=7  |
 | 0.9   ,   7<X<=8  |
```

```
| 0.94  ,   8<X<=9  |
| 0.96  ,  9<X<=10  |
| 0.98  ,  10<X<=13 |
| 1     ,    X>13   |
```




```
#Построение гистограммы частот
h <- hist(x,col="purple")
```



b) Вычислить выборочные аналоги следующих числовых характеристик:

(i) математического ожидания, (ii) дисперсии, (iii) медианы, (iv) асимметрии, (v) эксцесса, (vi) вероятности P(X ∈ [a, b]).

Формула мат. ожидания

$E(x) = \sum_{i=1}^{n} x_i p_i = \frac{1}{n} \sum_{i=1}^{n} x_i$

(i) Нахождение математического ожидания:

```
#E - мат. ожидание
E<-sum(x)/length(x)
cat("E(x) = ",E)
mean(x)
```

```
E(x) =  4.98
```

Формула дисперсии

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

(ii) Нахождение дисперсии

```
#D - дисперсия
D<-sum(x^2)/length(x)-E^2
cat("D(x) = ",D)
```

```
D(x) =  5.3796
```

Формула для нахождения медианы:

$M_e = \begin{cases} X_{np}, & np \in Z \\ \frac{X_{np} + X_{np+1}}{2}, & np \notin Z \end{cases} |p = \frac{1}{2}$

(iii) Нахождение медианы:

```
if(length(VariationalSeries)%%2!=0){ #если остаток от деления по модулю не делится без остатка

M_e<-VariationalSeries[(length(VariationalSeries)-1)/2]

} else {

M_e<-(VariationalSeries[(length(VariationalSeries))/2]+VariationalSeries[(length(VariationalSeries)+2

}
cat("M_e = ", M_e)
```

```
M_e =  5
```

Формула для нахождения асимметрии:

$$A_s = \frac{E(x-E(x))^3}{(D(X))^{\frac{3}{2}}}$$

(iv) Нахождение асимметрии:

```
A_s<-sum((x-E)^3)/(length(x)*D^(3/2));
cat("A_s = ", A_s,"\n")
cat("Проверка с помощью внутр. функции = ", skewness(x))
```

```
A_s =  0.6686303
Проверка с помощью внутр. функции =  0.6686303
```

Формула для нахождения эксцесса:

$$E_k = \frac{E(x-E(x))^4}{(D(x))^2} - 3$$

(v) Нахождение эксцесса:

```
E_k<-(sum((x-E)^4)/(length(x)*D^2))-3;
cat("E_k = ", E_k)
```

```
E_k =  1.562338
```

Формула для нахождения вероятности $P(X \in [a,b])$:

$$P(X \in [a,b]) = F(b) - F(a)$$

(vi) Нахождение вероятности $P(X \in [a,b])$:

```
a <- 4.55
b <- 5.45
p <- F(x,b) - F(x,a)
cat("P(X∈[a,b]) = ", p)
```

```
P(X∈[a,b]) =  0.2
```

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.*

Нахождение оценки максимального правдоподобия параметра λ:

```
#install.packages("maxLik") #Если нет maxLik раскомментируйте и запустите
```

```
LL<-function(t){sum(dpois(x,t[1],log=TRUE))}

maxLL<-maxNR(LL,start=c(1))

OMP<-maxLL$estimate
cat("ОМП = ", OMP)
```

```
ОМП =  4.98
```

Сделаем теоретическую проверку.

Плотность распределения Пуассона

$p(k) = \frac{e^{-\lambda}\lambda^k}{k!}$

Оценка максимального правдоподобия

$L(\vec{x},\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i * e^{-n\lambda}}}{\prod_{i=1}^{n} x_i!}$

$LL(\vec{x},\lambda) = log(\lambda)\sum_{i=1}^{n} xi - n\lambda - log(\prod_{xi}^{n} x_i!)$

$\partial\partial\lambda LL(\vec{x},\lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0$

$\lambda\bar{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} = 4.98(\bar{x} = E)$

Нахождение оценки λ по методу моментов:

```
MM<-E
cat("ОМП = ", MM)
```

```
ОМП =  4.98
```

Сделаем теоретическую проверку.

Плотность распределения Пуассона

$p(k) = \frac{e^{-\lambda}\lambda^k}{k!}$

Метод моментов

$E(x1) = \lambda$
$\bar{x} = E(x1)$

$$\bar{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} = 4.98(\bar{x} = E)$$

Нахождение смещения оценок

$$E(\bar{\lambda}) = E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{n\lambda}{n} = \lambda = E(\lambda)$$

Следовательно, оценка $\bar{\lambda} = \bar{x}$ несмещённая.

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

Так как x_i имеет распределение Пуассона, то $I_1(\lambda) = \frac{1}{\lambda}$. Из этого следует, что $I(\lambda) = n * I_1(\lambda) = \frac{n}{\lambda}$.

По методу максимального правдоподобия:

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

$$\sqrt{I(\lambda)}(\bar{\lambda} - \lambda) => N(0,1)$$

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

$$\sqrt{n * I_1(\lambda)}(\bar{\lambda} - \lambda) => N(0,1)$$

$$\sqrt{\frac{n}{\bar{x}}}(\bar{x} - \lambda) => N(0,1), \alpha_1 = 0.02$$

$$p(T_1(\bar{x}) \leqslant \lambda \leqslant T_2(\bar{x})) = 1 - \alpha_1$$

$$p(-x_\alpha \leq \sqrt{\frac{n}{\bar{\bar{x}}}}(\overline{x} - \lambda) \leq_\alpha) = \Phi(x_\alpha) - \Phi(-x_\alpha) = 2 * \Phi(x_\alpha) - 1 = 1 - \alpha 1$$

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.

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

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

$$[4.02 - x_\alpha * 0.4949182; 4.02 + x_\alpha * 0.4949182]$$

```
a_1 <- 0.02 #вероятность того что не попадет в интервал
x_a <- qnorm(1-a_1/2)

T[1]<-E-xa*sqrt((E*(E+1))/length(x))

T[2]<-E+xa*sqrt((E*(E+1))/length(x))
cat("Левая граница = ", T[1],"\n")
cat("Правая граница = ", T[2])
```

```
Левая граница =  3.184626
Правая граница =  6.775374
```

e) Используя гистограмму частот, построить критерий значимости χ2 проверки простой гипотезы согласия с распределением Пуассона с параметром λ0. Проверить гипотезу на уровне значимости α1. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

Таблица частот имеет вид

```
table(x)
```

```
x
 1  2  3  4  5  6  7  8  9 10 13
 4  3  6  6 10 13  3  2  1  1  1
```

```
lambda.0 <- 7
nu <- h$counts
brk <- 1:13
lw <- c(-Inf, brk)
up <- c(brk, Inf)
p.0 <-ppois(up, lambda.0) - ppois(lw, lambda.0)
cat("f",(nu - length(x)*p.0)^2/(length(x)*p.0))
X <- sum((nu - length(x)*p.0)^2/(length(x)*p.0))
x.alpha <- qchisq(0.98, 13)
x.alpha
cat("\nВыборка  = ",X)
cat("\nнаибольшмй уровень значимости = ",1 - pchisq(X, 13))
```

```
f 120.7023 106.0274 159.5635 0.04219167 3.01222 7.450139 5.584365 0.03550991 9.471292 106.5984 3.3275
Выборка  =  523.0796
наибольшмй уровень значимости =  0
```

f) Построить критерий значимости χ2 проверки сложной гипотезы согласия с распределением Пуассона. Проверить гипотезу на уровне значимости α1. Вычислить наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу.

```
f <-  function(lambda){
  p.0 <-ppois(up, lambda) - ppois(lw, lambda)
  X <- sum((nu - length(x)*p.0)^2/(length(x)*p.0))
  return(X)
}
nlm.res <- nlm(f, mean(x)) #нелинейная минимизация
x.alpha <- qchisq(0.98, 12)

cat("x.alpha - ", x.alpha)
X.bar <- nlm.res$minimum
cat("\nX.bar - ", X.bar)
```

```
x.alpha -  24.05396
X.bar -  428.024

Warning message in nlm(f, mean(x)):
“NA/Inf replaced by maximum positive value”
```

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

```
lambda.1 <- 5
nu1 <- h$counts
brk1 <- 1:13
lw1 <- c(-Inf, brk1)
up1 <- c(brk1, Inf)
p.1 <-pgeom(up1, 1/(lambda.1 + 1)) - pgeom(lw1, 1/(lambda.1 + 1))
X <- sum((nu1 - length(x)*p.1)^2/(length(x)*p.1))
x.alpha <- qchisq(0.98, 13)
cat("Выборка  = ",X)
cat("\nнаибольшмй уровень значимости = ",1 - pchisq(X, 13))
```

```
Выборка  =  529.9512
наибольшмй уровень значимости =  0
```

```
f1 <-  function(lambda){
  p.2 <-pgeom(up1, 1/(lambda.1 + 1)) - pgeom(lw1, 1/(lambda.1 + 1))
  X <- sum((nu - length(x)*p.2)^2/(length(x)*p.2))
  return(X)
}
nlm.res <- nlm(f1, mean(x))
x.alpha <- qchisq(0.98, 13)

cat("x.alpha - ", x.alpha)
X.bar <- nlm.res$minimum
cat("\nX.bar - ", X.bar)
```

```
x.alpha -  25.47151
X.bar -  529.9512
```