**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра ВМ-2**

отчет

**по лабораторной работе №1**

**по дисциплине «Математическая статистика»**

Тема: **Классические методы мат. статистики**

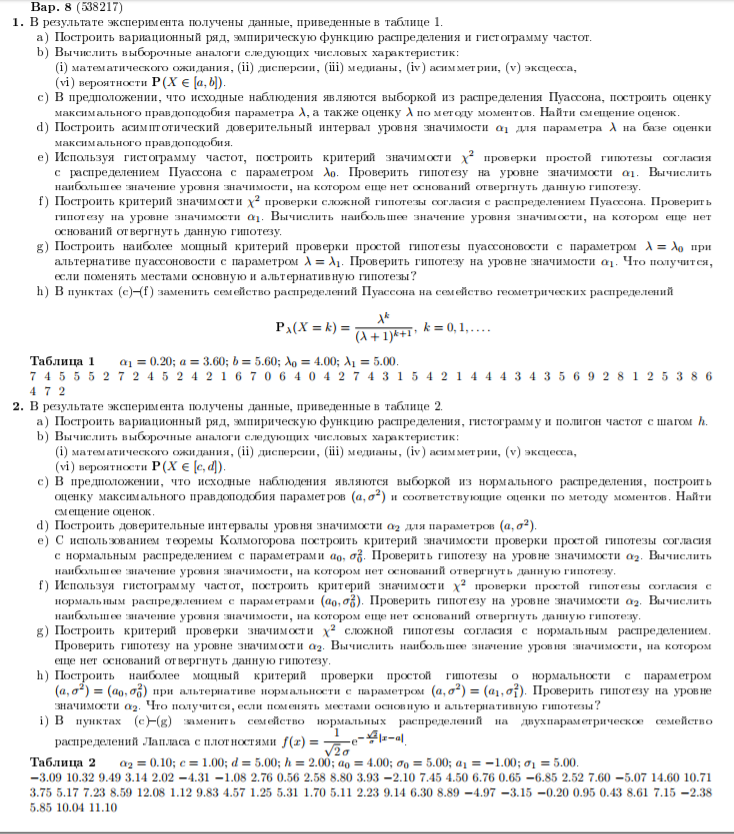
Вариант 8

|  |  |  |
| --- | --- | --- |
| Студентка гр. 5382 |  | Коппель Т.С. |
| Преподаватель |  | Чирина А.А. |

Санкт-Петербург

2017

**Постановка задачи**



**Ход работы:**

**Часть 1**

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

Чтобы получить вариационный ряд, считываем данные из текстового файла и сразу сортируем.



Рисунок 1. Входные данные

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| # | 2 | 4 | 9 | 4 | 12 | 7 | 4 | 5 | 2 | 1 |

*setwd ("C:\\MatStat")*

*dir()*

*#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*a):*

*x <- sort(scan ("input1.txt"))*

*print(x)#вариац ряд*

Чтобы сделать два графика – эмпирическую функцию распределения и гистограмму с полигоном, используем функцию “par(mfrow = c(1, 2))” .

*par(mfrow = c(1, 2))*

*f <- function(x, t){z <- x[x < t]; length(z)/length(x)}*

*xu <- unique (sort(x))*

*yu <- 0;*

*for(i in 1:length(xu)) yu[i] <- f(x, xu[i]); yu[length(xu)+1] <- 1*

*z <- stepfun(xu, yu)*

*plot(z, verticals = FALSE)*

*h1 <- hist(x, breaks = 0:9, right = F, col = "pink", main = "Гистограмма частот", xlab = " ", ylab = " ")*

*lines(h1$counts ~ h1$mids, col="black")*

*rug(x)*



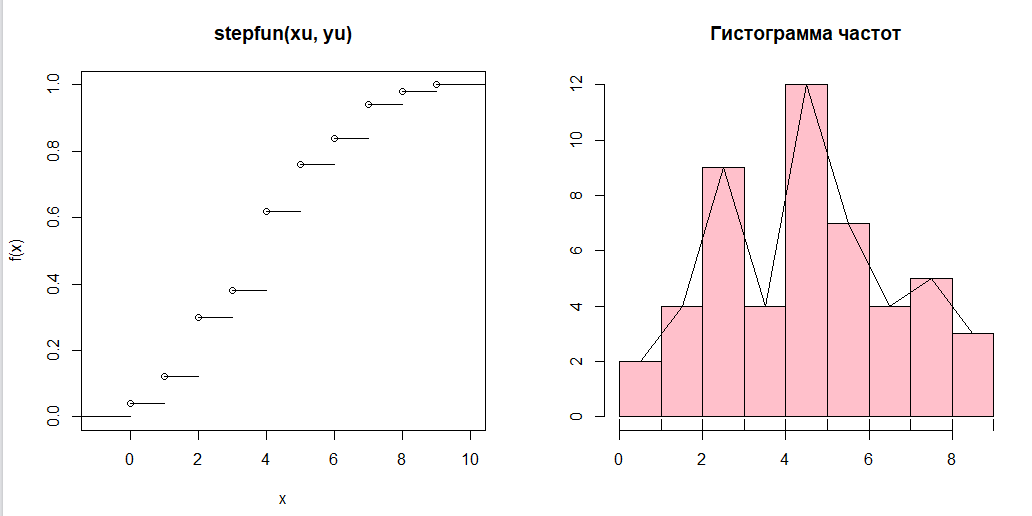


Рисунок 2. Получившиеся результаты

**b) Вычислить выборочные аналоги след. числовых характеристик:**

1. Математическое ожидание: 

*mean <- sum(x)/length(x)*

***mean = 4.02***

1. Дисперсия: 

*var <- sum(x^2)/length(x)-mean^2*

***var = 4.6596***

1. Медианы: - выборочная медиана, равная выборочной квантили порядка p:

при p=1/2

*med <- sort(x)[trunc(length(x)\*1/2+1)]*

***med = 4***

1. Асимметрии – выборочная асимметрия: 

*asm <- sum((x-mean)^3)/length(x)/var^(3/2)*

***asm = 0.2008713***

1. Эксцесса – выборочный эксцесс: 

*exc <- sum((x-mean)^4)/length(x)/var^2-3*

***esc = -0.61572***

1. Вероятность попадания в заданный промежуток: P(X ∈ [a, b]).

*a <- 3.60; b <- 5.60*

*p <- f(x, b) - f(x, a)*

***p = 0.38***

**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.**

*library(maxLik)*

*LL <- function (t)*

*{*

*sum(dcauchy(x, t[1], log = TRUE))*

*}*

*ml <- maxNR(LL, start = c(1)) #максимум функции правдоподобия*

*val <- ml$estimate; #оценка макс.правдоподобия*

*print(val)*

***val = 3.999189 === mean, т.е. выборочному среднему***

*Проверим теорией:*

 - плотность распределения Пуассона.

Метод максимального правдоподобия:

 =>  => 

 =>

Метод моментов:

математическое ожидание: ****, выборочный средний момент:

** =**> .

, значит  - несмещенная оценка.

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

Так как  имеет распределение Пуассона, то  => .

По методу максимального правдоподобия:

 , 

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



; = *0.20*



,

где  - квантиль порядка  стандартного нормального закона распределения.

.

.

*al <- 0.20*

*xal <- qnorm (1-al/2)*

*T<-array(dim = 2)*

*T[1]<-mean-xal\*sqrt(mean/length(x))*

*T[2]<-mean+xal\*sqrt(mean/length(x))*

*print(T) #доверительный интервал*

***[T[1], T[2]] = [3.656617 4.383383]***

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

Простая гипотеза Hо: , λo=4.00

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| # | 2 | 4 | 9 | 4 | 12 | 7 | 4 | 5 | 2 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| (-Inf; 1] | 6 | 0.09157819 | 4.578910 | 0.6641109 | 0.44104333 |
| (1; 2] | 9 | 0.14652511 | 7.326256 | 0.6183696 | 0.38238094 |
| (2; 3] | 4 | 0.19536681 | 9.768341 | -1.8456124 | 3.40628524 |
| (3; 4] | 12 | 0.19536681 | 9.768341 | 0.7140317 | 0.50984125 |
| (4; 5] | 7 | 0.15629345 | 7.814673 | -0.2914256 | 0.08492889 |
| (5; 6] | 4 | 0.10419563 | 5.209782 | -0.5300261 | 0.28092767 |
| (6; 7] | 5 | 0.05954036 | 2.977018 | 1.1724687 | 1.37468281 |
| (7; +Inf) | 3 | 0.05113362 | 2.556681 | 0.2772543 | 0.07686995 |
| xα = 9.80325 | |  |  | Χ2== 6.55696 | |

Делим последовательность на r = 8 интервала и вводим вектор border границ интервалов, имеющий размерность r-1, который потребуется для получения значений частот из гистограммы**: border <- c(1, 2, 3, 4, 5, 6, 7)**

Нижние границы элементов: **a1<-c(-Inf, 2, 3)** Верхние границы элементов: **b1<-c(1, 2, Inf)**



Критерий имеет вид:

*n <- length(x)*

*lambda0 <- 4*

*r <- 8*

*al <- 0.20*

*border <- c(1, 2, 3, 4, 5, 6, 7)*

*h <- hist(x, breaks = c(min(x), border, max(x)), plot = FALSE)*

*nu <- h$counts*

*print(nu)*

*p1<-array(dim = r)*

*p1[1]<- ppois(border[1], lambda0)*

*p1[r] <- 1-ppois(border[r-1], lambda0)*

*p1[2:(r-1)]<-ppois(border[2:(r-1)],lambda0)-ppois(border[1:(r-2)],lambda0)*

*print (p1)*

*print(n\*p1)*

*res <- array (dim = r)*

*res [1:r] <- (nu[1:r] - n\*p1[1:r])/sqrt(n\*p1[1:r])*

*print (res)*

*res2 <- array (dim = r)*

*res2 [1:r]<- (res[1:r])^2*

*print (res2)*

*Xi2<-sum(res2); print(Xi2)*

*xal<-qchisq(1-al, r-1);print(xal)*

*Xi2>xal*

**# FALSE**

Итак, получили, что χ2 < xα, следовательно, нужно **отвернуть** гипотезу Hо.

Для нахождения наибольшего значения уровня значимости, на котором ещё нет оснований отвергнуть данную гипотезу, вычисляем функцию распределения  в точке , и вычитаем полученное значение из единицы:

*#находим наибольший уровень значимости, при котором нет оснований отвергнуть гипотезу:*

*al2<-1-pchisq(Xi2,r-1);*

*print(al2)*

***al2 = 0.4764169***

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

Но – основная гипотеза: Х ~ Pois ()

Поделим область на *r*=3 интервалов, аналогично предыдущему пункту.

Х2()=- зависит от , т.к. величины *pi* не фиксированы. Известно, что в случае регулярности эксперимента статистика  сходится по распределению к **

Критерий имеет вид:

*al <- 0.20*

*a1 <- c(-Inf, 4, 5, 6, 7)*

*b1 <- c(1, 2, 3, 4, Inf)*

*border <- c(1, 2, 3, 4, 5, 6, 7)*

*al <- 0.2; r <- 8; n <- 50*

*csq<-function (t){ #функция для X2*

*p<-pnorm(b1,0,t) - pnorm (a1,0,t);*

*f<-sum((nu-n\*p)^2/(n\*p));*

*print (f)*

*}*

*X2<-nlm(csq,p=mean(x))*

*xal1<-qchisq (1-al, r-2)*

*xal1*

***#xal1=8.55806***

*X2$minimum<=xal1#производим сравнение*

***#TRUE***

Итак, , и гипотеза **Принимается.**

*alpha2<-pchisq(X2$minimum,r-2) #наибольшее значение уровня значимости, на котором еще нет оснований отвергнуть данную гипотезу*

*alpha2*

#***alpha2=0***

**g) Построить наиболее мощный критерий проверки простой гипотезы пуассоновости с параметром** **при альтернативе пуассоновости с параметром** . **Проверить гипотезу на уровне значимости Что получится, если поменять местами основную и альтернативную гипотезы?**



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

 , где 





Наиболее мощный критерий:

Логарифмируем соотношение

Получим:

Обозначим через:

Тогда критерий:

Найдем и *p* из уравнения:



 , следовательно, 

Т.к. , то подбором (в цикле с помощью R) среди целых чисел можем найти такое наибольшее  (а после и α0), что:



Тогда 

*c<-0*

*alpha1<-0.02*

*alpha0<-1-ppois(c,lambda0\*n)-dpois(c, lambda0\*n)*

*while (alpha0 > alpha1)*

*{*

*c<-c+1;*

*alpha0<-1-ppois(c,lambda0\*n)-dpois(c, lambda0\*n)*

*}*

*print(c)*

***c = 228***

*p<-(alpha1-alpha0)/dpois(c,lambda0\*n)*

*print(p)*

***p = 0.07106808***

*print(alpha0)*

***alpha0 = 0.019712***

*lche<- sum(x)*

*lche>=c*

***FALSE***

Принимаем 

Критерий:



Теперь поменяем местами основную и альтернативную гипотезу.









Наиболее мощный критерий:

Логарифмируем соотношение

Получим:

Обозначим через:

Тогда критерий:

Отыщем и *p* из уравнения:



 , следовательно, 

Т.к. , то подбором (в цикле с помощью R) среди целых чисел можем найти наибольшее  (а после и α0)

Тогда с учётом уравнения выше 

*c<-0; lambda1 <-5*

*alpha0<-ppois(c,lambda1\*n)*

*while(alpha0<alpha1){*

*c<-c+1;*

*alpha0<-ppois(c,lambda1\*n)*

*}*

*c<-c-1*

*print(c)*

***c = 217***

*alpha0<-ppois(c,lambda1\*n)*

*p<-(alpha1-alpha0)/(dpois(c,lambda1\*n))*

*print(alpha0)*

***alpha0 = 0.01822694***

*print(p)*

***p = 0.6407782***

*lche<- sum(x)*

*lche<=c*

***TRUE***

**Принимаем альтернативу.**

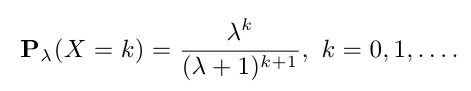
Критерий построен:

При замене основной и альтернативной гипотезы меняется также гипотеза, которую принимаем. Изменение происходит от 

на , но решение принимается в пользу разных гипотез.

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

**, *k = 0, 1, …***





Обозначим 

Найдём оценку максимального правдоподобия:







Для геометрического распределения математическое ожидание: ****, выборочный средний момент: ** =**> .

, значит  - несмещенная оценка.

1. **Пункт c)**

*library(maxLik)*

*LL<-function(t)*

*{*

*sum(dgeom(x,t[1],log=TRUE))*

*}*

*ml<-maxNR(LL,start=c(1)) #максимум функции правдоподобия*

*val<-ml$estimate;*

*print (val) #оценка макс.правдоподобия*

***val = 3.999189***

1. **Пункт d)**



Найдём информацию Фишера.



ОМП параметра :





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









Где *,* т.е. *b* - квантиль стандартного нормального распределения.





*alpha<-0.20*

*T<-array(dim=2)*

*xal<-qnorm (1-alpha/2)*

*T[1]<-mean-xal\*sqrt((mean\*(mean+1))/length(x)) #левая граница Д.И.*

*T[2]<-mean+xal\*sqrt((mean\*(mean+1))/length(x)) #правая граница Д.И.*

*print(T)*

***T = [3.205828 4.834172]***

1. **Пункт e)**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| # | 2 | 4 | 9 | 4 | 12 | 7 | 4 | 5 | 2 | 1 |

Делим последовательность на r = 8 интервала. Границы интервалов берем такие же, как раньше. 

Критерий имеет вид

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| (-Inf; 1] | 6 | 0.38752363 | 19.376181 | -3.0387716 | 9.2341327 |
| (1; 2] | 9 | 0.13314704 | 6.657352 | 0.9079382 | 0.8243519 |
| (2; 3] | 4 | 0.10420203 | 5.210101 | -0.5301499 | 0.2810589 |
| (3; 4] | 12 | 0.08154941 | 4.077471 | 3.9234528 | 15.3934817 |
| (4; 5] | 7 | 0.06382128 | 3.191064 | 2.1322382 | 4.5464438 |
| (5; 6] | 4 | 0.04994709 | 2.497354 | 0.9508597 | 0.9041342 |
| (6; 7] | 5 | 0.0.908903 | 1.954451 | 2.1784777 | 4.7457651 |
| (7; +Inf) | 3 | 0.14072049 | 7.036025 | -1.5215637 | 2.3151561 |
| xα = 9.80325 | |  |  | Χ2== 38.24452 | |

*n <- length(x)*

*lambda0 <- 4*

*r <- 7*

*al <- 0.05*

*border <- c(1, 2, 3, 4, 5, 6)*

*h <- hist(x, breaks = c(min(x), border, max(x)), plot = FALSE)*

*nu <- h$counts*

*print(nu)*

*p1<-array(dim = r)*

*p1[1]<- pgeom(border[1], prob=0.02)*

*p1[r] <- 1-pgeom(border[r-1], prob=0.02)*

*p1[2:(r-1)]<-pgeom(border[2:(r-1)], prob=0.02)-pgeom(border[1:(r-2)], prob=0.02)*

*print (p1)*

*print(n\*p1)*

*res <- array (dim = r)*

*res [1:r] <- (nu[1:r] - n\*p1[1:r])/sqrt(n\*p1[1:r])*

*print (res)*

*res2 <- array (dim = r)*

*res2 [1:r]<- (res[1:r])^2*

*print (res2)*

*Xi2<-sum(res2); print(Xi2)*

*xal<-qchisq(1-al, r-1);print(xal)*

*Xi2>xal*

***#TRUE –принимаем гипотезу***

*#находим наибольший уровень значимости, при котором нет оснований отвергнуть гипотезу:*

*al2<-1-pchisq(Xi2,r-1); al2*

***#al2=0***

1. **Пункт f)**



Сложная гипотеза согласия: Но – основная гипотеза: Х ~ Geom (1/(+1))

Поделим область на *r*=3 интервалов, аналогично предыдущему пункту. *X*2- зависит от , т.к. величины *pi* не фиксированы. Известно, что в случае регулярности эксперимента статистика  сходится по распределению к **.

Критерий имеет вид:

*P<-function(a){*

*p[1]<-pgeom(b[1],1/(a+1))*

*i<-2*

*while(i<r){*

*p[i]<-pgeom(b[i],1/(a+1))-pgeom(b[i-1],1/(a+1));*

*i<-i+1;*

*}*

*p[r]<-1-pgeom(b[r-1],1/(a+1))*

*print(p)*

*}*

*X2<-function(a){g<-n\*P(a);f<-(nu-g)^2/g;sum(f)}*

*XM<-nlm(X2,p=mean) #проводим минимизацию,*

*xb<-qchisq(1-0.02,r-2) #вычисляем квантиль*

*XM$minimum<xb# гипотезу принимаем на заданном уровне знач.*

***FALSE***

*alpha2<-1-pchisq(XM$minimum,r-2)#наибольший уровень значимости, на котором*

*print(alpha2)*

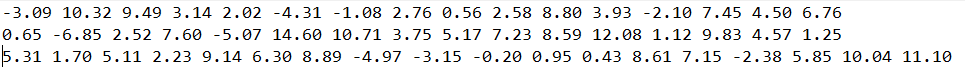
***alpha2 = 0.01040266***

Следовательно, нужно **отвергнуть гипотезу** H0.

**Часть 2**

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

Чтобы получить вариационный ряд, считываем данные из текстового файла и сразу сортируем.



*setwd ("C:\\MatStat")*

*dir()*

*#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*a):*

*x <- sort(scan ("input2.txt"))*

*print(x)#вариац ряд*

Чтобы сделать два графика – эмпирическую функцию распределения и гистограмму с полигоном, используем функцию “par(mfrow = c(1, 2))” .

*par(mfrow = c(1, 2))*

*f<-function(x,t){z<-x[x<t]; length(z)/length(x)}*

*xu<-unique (sort(x))*

*yu<-0; for(i in 1:length(xu)) yu[i]<-f(x,xu[i]); yu[length(xu)+1]<-1*

*z<-stepfun(xu,yu)*

*plot(z,verticals=FALSE)*

*h1 <- hist (x, breaks=c((max(x)-min(x))/0.4), right = TRUE, freq = TRUE, col = "lightblue")*

*lines(h1$counts ~ h1$mids, col="red")*

*rug(x)*

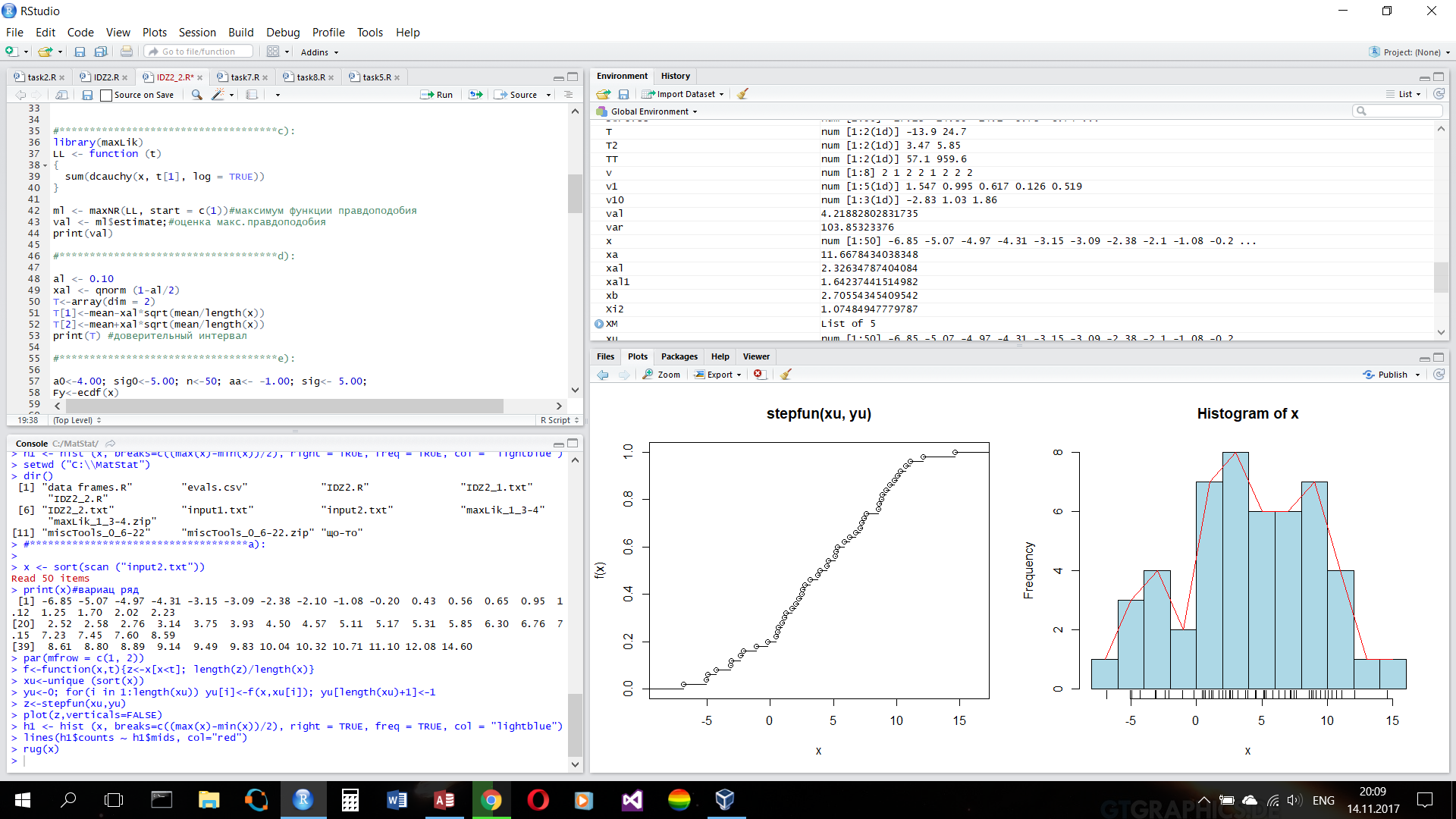


Рисунок 3. Получившиеся результаты

**b) Вычислить выборочные аналоги след. числовых характеристик:**

1. Математическое ожидание: 

*mean <- sum(x)/length(x)*

***mean = 4.0318***

1. Дисперсия: 

*var <- sum(x^2)/length(x)-mean^2*

***var = 24.95651***

1. Медианы: - выборочная медиана, равная выборочной квантили порядка p:

при p=1/2

*med <- sort(x)[trunc(length(x)\*1/2+1)]*

***med = 4.5***

1. Асимметрии – выборочная асимметрия: 

*asm <- sum((x-mean)^3)/length(x)/var^(3/2)*

***asm = -0.1784126***

1. Эксцесса – выборочный эксцесс: 

*exc<-sum((x-mean)^4)/length(x)/var^2-3*

***esc =*** ***-0.6998372***

1. Вероятность попадания в заданный промежуток: P(X ∈ [a, b]).

*c <- 1.00; b <-5.00*

*p <- f(x, d) - f(x, c)*

***p = 0.26***

**c) В предположении, что исходные наблюдения являются выборкой из нормального распределения, построить оценку максимального правдоподобия параметров, а также оценку  по методу моментов. Найти смещение оценок.**











**d) Построить доверительные интервалы уровня значимости для параметров** .

Построим доверительный интервал для *а*.

Согласно лемме Фишера 



 где  - квантиль распределения Стьюдента уровня 



*al <- 0.20*

*xal <- qnorm (1-al/2)*

*T<-array(dim = 2)*

*T[1]<-mean-xal\*sqrt(mean/length(x))*

*T[2]<-mean+xal\*sqrt(mean/length(x))*

*print(T) #доверительный интервал*

***[T[1], T[2]] = [3.667884 4.395716]***

**e) С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с нормальным распределением с параметрами . Проверить гипотезу на уровне значимости . Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.**



Согласно теореме Колмогорова, при справедливости гипотезы

 , где *К-* распределение Колмогорова.

Обозначим 



Согласно таблице распределения Колмогорова, 

Вычислим величину  в *R*.

*a0<-4.00; sig0<-5.00; n<-50; aa<- -1.00; sig<- 5.00;*

*Fy<-ecdf(x)*

*Fnorm<-pnorm(x,a0,sig0)*

*Diff<-array(dim=50)*

*for(i in 1:n){Diff[i]<-abs(Fy(x[i])-Fnorm[i])}*

*D<-max(Diff)*

*Dn<-D\*sqrt(n)*

*print(Dn)*

*** = 0.4291471***

- значит **гипотезу принимаем**.

**f) Используя гистограмму частот, построить критерий значимости  проверки простой гипотезы согласия с нормальным распределением с параметрами. Проверить гипотезу на уровне значимости . Вычислить наибольшее значение уровня значимости, на котором ещё нет основания отвергнуть данную гипотезу.**

Простая гипотеза согласия: 

Поделим область на *r*=5 интервалов, введём вектор внутренних границ *b*, вычислим частоты попадания наблюдения в *i*-й интервал  (с помощью гистограммы), величину 

, где  - теоретическая вероятность попадания в интервал с номером *i*.

Выбрали следующий вектор границ *b* = (0.5, 0.9, 1.1, 1.4)

Известно, что .

Тогда запишем критерий проверки гипотезы:

Если , то принимаем гипотезу, иначе отвергаем.

 - квантиль порядка  распределения 

Для нахождения наибольшего значения уровня значимости, на котором ещё нет оснований отвергнуть данную гипотезу, вычисляем функцию распределения  в точке , и вычитаем полученное значение из единицы.

Вычисления в R*.*

*alpha2<-0.10*

*r<-5 #количество интервалов*

*a<-4*

*b<-array(dim=r-1) #вектор границ*

*b[1]<--2.715; b[2]<--1.923; b[3]<--0.132; b[4]<-0.724;*

*h<-hist(x,breaks=c(min(x),b,max(x)))#построение гистограммы*

*p[1]<-pnorm(b[1],a,sig0)*

*i<-2*

*while(i<=r-1){*

*p[i]<-pnorm(b[i],a,sig0)-pnorm(b[i-1],a,sig0);*

*i<-i+1;*

*}*

*p[r]<-1-pnorm(b[r-1],a,sig0)#конец заполнения вектора*

*yhu<-h$counts#получение вектора частот*

*v1<-(yhu-n\*p)^2/(n\*p)#вектор слагаемых величины X2*

*X2<-sum(v1)#вычисление величины X2*

*xa<-qchisq(1-alpha2,r-1)#вычисление квантиля*

*X2<xa*

***#TRUE***

*alpha3<-1-pchisq(X2,r-1)#находим наибольший уровень значимости, при котором нет оснований отвергнуть гипотезу*

*alpha3*

***#alpha3 = 0.7263292***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № интервала |  |  |  |  |  |  |
| 1 | -Inf | 0.5 | 6 | 0.04779035 | 0.71713839 | 0.5142874652 |
| 2 | 0.5 | 0.9 | 2 | 0.32165099 | 0.48410882 | 0.2343613504 |
| 3 | 0.9 | 1.1 | 2 | 0.26111732 | -1.11270183 | 1.2381053660 |
| 4 | 1.1 | 1.4 | 3 | 0.27823012 | 0.25205190 | 0.0635301619 |
| 5 | 1.4 | Inf | 37 | 0.09121122 | -0.03140536 | 0.0009862967 |
| xα = 7.77944 | | | | =1 | = 2.051271 | |

Итак, получили, что χ2 < xα, следовательно, нужно **принять** гипотезу Hо.

**g) Построить критерий значимости проверки сложной гипотезы согласия с нормальным распределением. Проверить гипотезу на уровне значимости . Вычислить наибольшее значение уровня значимости, на котором ещё нет основания отвергнуть данную гипотезу.**



Сложная гипотеза согласия: , где  – функция нормального распределения с параметрами ;

Поделим область на *r*=4 интервалов, задав внутренние границы b=(0.6;0.9;51.3). *X*2- зависит от , т.к. величины *pi* не фиксированы. Известно, что в случае регулярности эксперимента статистика  сходится по распределению к ** Далее проведем вычисления в R:

*P<-function(aa,sig){*

*p[1]<-pnorm(b[1],aa,sig)*

*i<-2*

*while(i<=r-1){*

*p[i]<-pnorm(b[i],aa,sig)-pnorm(b[i-1],aa,sig);*

*i<-i+1;*

*}*

*p[r]<-1-pnorm(b[r-1],aa,sig);p;}*

*h<-hist(x,c(min(x),b,max(x)),plot=FALSE) #новая гистограмма*

*yhu<-h$counts*

*X2<-function(a,b){g<-n\*P(a,b);f<-(yhu-g)^2/g;sum(f)} #и величина X2 зависит от параметра*

*XM<-nlm(X2,mean(x),sqrt(var(x))) #проводим нелинейную минимизацию, отыскивая тем самым*

*yb<-qchisq(1-alpha2,r-3) #вычисляем квантиль*

*XM$minimum<yb# гипотезу принимаем на заданном уровне знач.*

***#TRUE***

*alpha3<-1-pchisq(XM$minimum,r-3) #наибольший уровень значимости, на котором*

*alpha3 #нет оснований отвергнуть гипотезу*

***alpha3 = 0.4085836***

## 

**h) Построить наиболее мощный критерий проверки простой гипотезы нормальности с параметрами  при альтернативе нормальности с параметрами  . Проверить гипотезу на уровне значимости . Что получится, если поменять местами основную и альтернативную гипотезы?**



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

 , где 



Т.к. по условию задачи , то  примет вид:



Значит, критерий можно переписать в виде:



Отыщем *А* из уравнения:



 , следовательно, 

Тогда *А* - квантиль распределения  уровня .

Проведём вычисления в R.

*A<- 0*

*A<-qnorm(alpha2,n\*a0,sqrt(n)\*sig0)*

*sum(y)<A*

***#TRUE***  #*гипотезу отвергаем на уровне* 

*A*

***A = 154.6903***

Критерий построен.



Теперь поменяем местами основную и альтернативную гипотезу.



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

 , где 



Т.к. по условию задачи , то  примет вид:





Значит, критерий можно переписать в виде:



Отыщем *А* из уравнения:



 , следовательно, 

Тогда *А* - квантиль распределения  уровня .

Проведём вычисления в R.

*a1<--1*

*A<-0*

*A<-qnorm(1-alpha2,n\*a1,sqrt(n)\*sig0)*

*A*

***#A = -4.69031***

*sum(y)<A*

***# FALSE #*** гипотезу отвергаем на уровне 

Критерий построен.



1. **В пунктах (с) – (g) заменить семейство нормальных распределений на двухпараметрическое семейство распределений Лапласа с плотностями**



**Пункт c)** В предположении, что исходные наблюдения являются выборкой из распределения Лапласа, построить оценку максимального правдоподобия параметров, а также оценку  по методу моментов.

Найдём оценку максимального правдоподобия:





Найдём оценку методом моментов



**Пункт d)** Построить асимптотические доверительные интервалы уровня значимости  для параметров  на базе оценки максимального правдоподобия.



ОМП параметра 



ОМП 



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











Где *,* т.е. *b* - квантиль стандартного нормального распределения.





Вычисления в R:

*T<-array(dim=2)*

*T[1]<-med-qnorm(1-alpha2/2)\*sum(abs(y-med))/n*

*T[2]<-med+qnorm(1-alpha2/2)\*sum(abs(y-med))/n*

*T*

***#4.088787 4.911213***

*TT<-array(dim=2)*

*TT[1]<-((sqrt(2)-qnorm(1-alpha2/2))\*sum(abs(y-med))/n)^2*

*TT[2]<-((sqrt(2)+qnorm(1-alpha2/2))\*sum(abs(y-med))/n)^2*

*TT*

***#0.003324677 0.584868254***

Получили асимптотические доверительные интервалы для параметров  уровня доверия 

Для *а* 

Для 

**Пункт e)** С использованием теоремы Колмогорова построить критерий значимости проверки простой гипотезы согласия с распределением Лапласа с параметрами . Проверить гипотезу на уровне значимости . Вычислить наибольшее значение уровня значимости, на котором нет оснований отвергнуть данную гипотезу.



Согласно теореме Колмогорова, при справедливости гипотезы

 , где *К-* распределение Колмогорова.

Обозначим 



Согласно таблице распределения Колмогорова, 

Вычислим величину  в *R*.

*Flapl<-function(x,a,sig){*

*if (x>a) p<-1-exp(-sqrt(2)\*(x-a)/sig)/2*

*else p<-exp(sqrt(2)\*(x-a)/sig)/2*

*p;}*

*Fy<-ecdf(x);*

*Flaplvec<-array(dim=50);*

*for( i in 1:n){Flaplvec[i]<-Flapl(y[i],5.5,0.5)}*

*Diff<-array(dim=50);*

*for(i in 1:n){Diff[i]<-abs(Fy(y[i])-Flaplvec[i])}*

*D<-max(Diff);*

*Dn<-D\*sqrt(n);*

*Dn*

[1] 7.07031

- значит гипотезу отвергаем. Наибольший уровень значимости, на котором нет оснований отвергнуть данную гипотезу, согласно распределению Колмогорова, ничтожно мал (не удалось отыскать).

**Пункт f)** Используя гистограмму частот, построить критерий значимости  проверки простой гипотезы согласия с распределением Лапласа с параметрами. Проверить гипотезу на уровне значимости . Вычислить наибольшее значение уровня значимости, на котором ещё нет основания отвергнуть данную гипотезу.

Простая гипотеза согласия: 

Поделим область на *r*=5 интервалов, введём вектор внутренних границ *b*, вычислим частоты попадания наблюдения в *i*-й интервал  (с помощью гистограммы), величину .

, где  - теоретическая вероятность попадания в интервал с номером *i*.

Выбрали такой же вектор границ , как и ранее

Известно, что .

Тогда запишем критерий проверки гипотезы:

Если , то принимаем гипотезу, иначе отвергаем.

 - квантиль порядка  распределения 

Для нахождения наибольшего значения уровня значимости, на котором ещё нет оснований отвергнуть данную гипотезу, вычисляем функцию распределения  в точке , и вычитаем полученное значение из единицы.

Вычисления в R*.*

*h<-hist(y,breaks=c(min(y),b,max(y)),plot=FALSE) #построение гистограммы*

*p<-array(dim=5)#вектор теоретических вероятностей*

*p[1]<-Flapl(b[1],a,sig0)*

*i<-2*

*while(i<=r-1){*

*p[i]<-Flapl(b[i],a,sig0)-Flapl(b[i-1],a,sig0);*

*i<-i+1;*

*}*

*p[r]<-1-Flapl(b[r-1],a,sig0)*

*yhu<-h$counts#получение вектора частот*

*v1<-(yhu-n\*p)^2/(n\*p)#вектор слагаемых величины X2*

*X2<-sum(v1)#вычисление величины X2*

*xa<-qchisq(1-alpha2,r-1)#вычисление квантиля*

*X2<xa#гипотеза опроверглась*

***# FALSE***

*alpha3<-1-pchisq(X2,r-1)#находим наибольший уровень значимости, при*

*alpha3#котором нет оснований отвергнуть гипотезу*

*#0 #уровень значимости (очень мал)*

**Пункт g)** Построить критерий значимости проверки сложной гипотезы согласия с распределением Лапласа. Проверить гипотезу на уровне значимости . Вычислить наибольшее значение уровня значимости, на котором ещё нет основания отвергнуть данную гипотезу.



Сложная гипотеза согласия: , где  – функция распределения Лапласа с параметрами ;

Поделим область на *r*=4 интервалов, задав внутренние границы b=(0.0,0.05,0.08,0.14, 0.18). *X*2- зависит от , т.к. величины *pi* не фиксированы. Известно, что в случае регулярности эксперимента статистика  сходится по распределению к ** Далее проведем вычисления в R:

*P<-function(aa,sig){*

*p[1]<-Flapl(b[1],aa,sig)*

*i<-2*

*while(i<=r-1){*

*p[i]<-Flapl(b[i],aa,sig)-Flapl(b[i-1],aa,sig);*

*i<-i+1;*

*}*

*p[r]<-1-Flapl(b[r-1],aa,sig);p;}*

*h<-hist(x,c(min(y),b,max(x)),plot=FALSE) #новая гистограмма*

*yhu<-h$counts*

*X2<-function(a,b){g<-n\*P(a,b);f<-(yhu-g)^2/g;sum(f)} #и величина X2 зависит от параметра*

*XM<-nlm(X2,mean(y),sqrt(var(y))) #проводим нелинейную минимизацию, отыскивая тем самым*

*alpha3<-1-pchisq(XM$minimum,r-3)#наибольший уровень значимости, на котором*

*alpha3*

***# alpha3 = 1.481391e-09***