**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра ВМ-2**

отчет

**по индивидуальному заданию №3**

**по дисциплине «Статический анализ»**

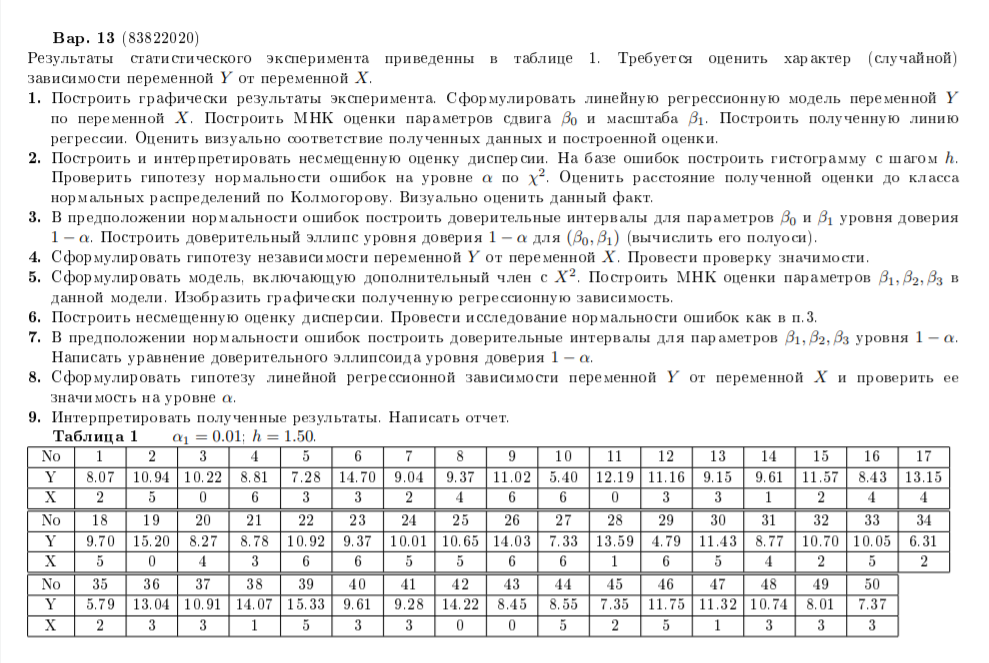
Вариант №13

|  |  |  |
| --- | --- | --- |
| Студент гр. 8382 |  | Мирончик П.Д. |
| Преподаватель |  | Малов С.В. |

Санкт-Петербург

2020

**Постановка задачи**

****

# Ход решения

# 1

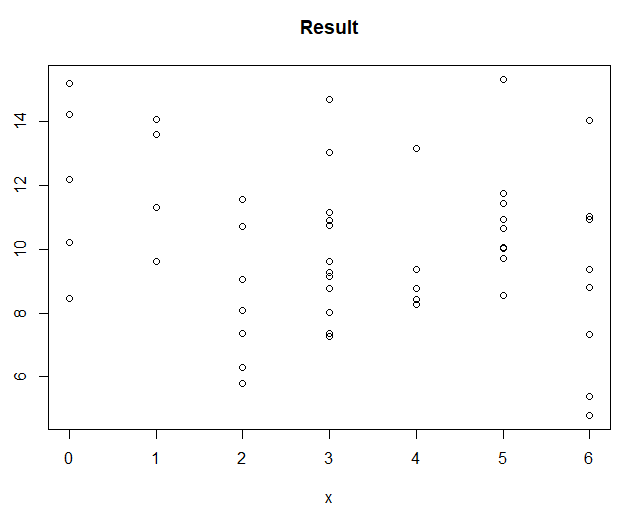
**Построить графически результаты эксперимента. Сформулировать линейную регрессионную модель переменной Y по переменной X. Построить МНК оценки параметров сдвига β0 и масштаба β1. Построить полученную линию регрессии. Оценить визуально соответствие полученных данных и построенной оценки.**

Графическое представление:

## x <- c(2, 5, 0, 6, 3, 3, 2, 4, 6, 6, 0, 3, 3, 1, 2, 4, 4, 5, 0, 4, 3, 6, 6, 5, 5, 6, 6, 1, 6, 5, 4, 2, 5, 2, 2, 3, 3, 1, 5, 3, 3, 0, 0, 5, 2, 5, 1, 3, 3, 3)

## y <- c(8.07, 10.94, 10.22, 8.81, 7.28, 14.70, 9.04, 9.37, 11.02, 5.40, 12.19, 11.16, 9.15, 9.61, 11.57, 8.43, 13.15, 9.70, 15.20, 8.27, 8.78, 10.92, 9.37, 10.01, 10.65, 14.03, 7.33, 13.59, 4.79, 11.43, 8.77, 10.70, 10.05, 6.31, 5.79, 13.04, 10.91, 14.07, 15.33, 9.61, 9.28, 14.22, 8.45, 8.55, 7.35, 11.75, 11.32, 10.74, 8.01, 7.37)

## plot(x,y,main="Result")



МНК оценка параметров сдвига β0 и масштаба β1

## XM = matrix(c(array(1,dim=50),x),nrow=2,ncol=50,byrow=TRUE)

## S = XM %\*% t(XM)

## YM = as.matrix(y)

## S1 = solve(S)

## # МНК оценка

## bht = S1 %\*% XM %\*% YM;

## print(bht)

Результат: β0 = 11.0500639, β1 = -0.2796599

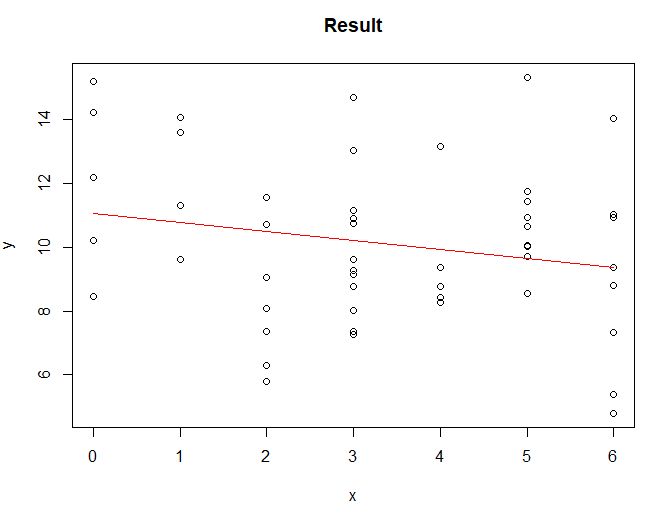
Линейная регрессионная модель: .

Постройка полученной регрессии:

## xx<-c(0,1,2,3,4,5,6)

## yy<-bht[1]+bht[2]\*xx

## points(xx,yy,"l",col="red")



Линия регрессии имеет отрицательный угол наклона, что совпадает с коэффициентом наклона, найденным по МНК.

# 2

**Построить и интерпретировать несмещенную оценку дисперсии. На базе ошибок построить гистограмму с шагом h. Проверить гипотезу нормальности ошибок на уровне α по χ2. Оценить расстояние полученной оценки до класса нормальных распределений по Колмогорову.**

Построение несмещенной оценки дисперсии:

## eps<-(YM-t(XM)%\*%bht)

## n = length(y)

## r = dim(XM)[1]

## SSE = sum(eps^2)

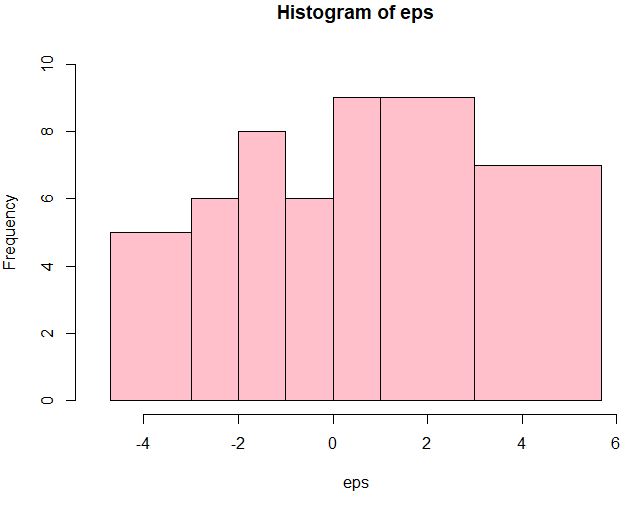
## sht = SSE/(n-r)

## print(sht)

Результат: 6.248361

Построение гистограммы с шагом h = 1.5 на базе ошибок

hh = hist(eps, breaks=c(min(eps), -3, -2, -1, 0, 1, 3, max(eps)), col="pink", freq=TRUE, xlim = c(-5, 6), ylim = c(0, 10))



Проверка гипотезы нормальности ошибок на уровне α по χ2:

𝐻0: 𝑌 − 𝑋𝑇 𝛽 ~(0, 𝜎 2)

## nu = hh$counts;

## breaks = hh$breaks;

## r = length(breaks) - 1

## csq0<-function(s){

## if (s>0){

## p<-pnorm(breaks[2:length(breaks)],0,s)-pnorm(breaks[1:(length(breaks)-1)],0,s)

## return(sum((nu-n\*p)^2/n/p))

## } else {

## return(Inf)

## }

## }

## csq.s<-nlm(csq0,p=sqrt(sht))$minimum

## xa1 = qchisq(0.99, r - 2)

## print(csq.s < xa1)

TRUE - значит, принимаем гипотезу.

Оценивание расстояния полученной оценки до класса нормальных распределений по Колмогорову:

## kolm.stat<-function(s){

## sres<-sort(eps)

## fdistr<-pnorm(sres,0,s)

## max(

## abs(c(0:(n-1))/n-fdistr),

## abs(c(1:n)/n-fdistr)

## )

## }

## ks.dist<-nlm(kolm.stat,p=sqrt(sht))$minimum

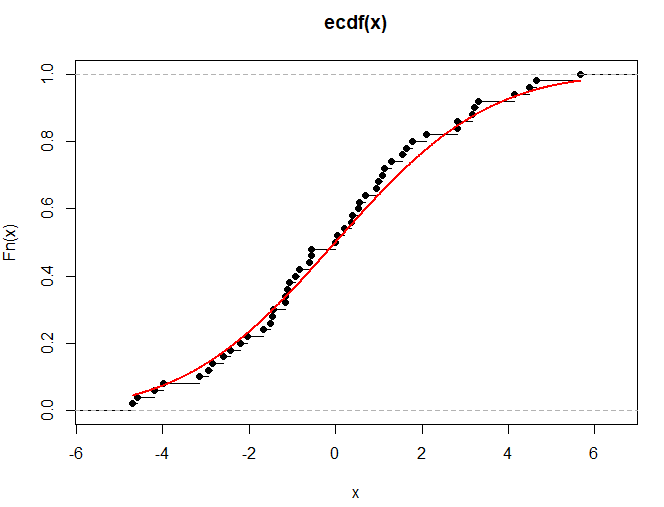
## plot.ecdf(eps)

## x2<-c(0:1000)\*(max(eps)-min(eps))/1000+min(eps)

## y2<-pnorm(x2,0,nlm(kolm.stat,p=sqrt(s2))$estimate)

## points(x2,y2,"l",col="red",lwd=2)

Полученное расстояние 0.06051206



# 3

**В предположении нормальности ошибок построить доверительные интервалы для параметров β0 и β1 уровня доверия 1-α. Построить доверительный эллипс уровня доверия 1-α для (β0, β1).**

Построим ДИ для параметров с уровнем доверия

## al = 0.1

## C<-matrix(c(c(1,0), c(0,1)),nrow=2, ncol=2)

## BVar<-matrix(nrow=2,ncol=2)

## XXt<-XM%\*%t(XM)

## bh<-matrix(nrow=2, ncol=2)

## xal<-qt(1-al/2,df=n-2)

## BVar<-t(C)%\*%solve(XXt)%\*%C

## b0low<-bhat[1]-xal\*sqrt(sht)\*sqrt(BVar[1,1])

## print(b0low)

## b0up<-bhat[1]+xal\*sqrt(sht)\*sqrt(BVar[1,1])

## print(b0up)

## b1low<-bhat[2]-xal\*sqrt(sht)\*sqrt(BVar[2,2])

## print(b1low)

## b1up<-bhat[2]+xal\*sqrt(sht)\*sqrt(BVar[2,2])

## print(b1up)

Получили

Для β0: [9.110742; 12.98939]

Для β1: [-0.786164; 0.2268443]

Доверительный эллипс можно вычислить как

𝐴𝛼 = {𝑥, 𝑦: (( 𝑥 𝑦 ) T − 𝛽) 𝑇 \* (𝑋 𝑋𝑇 ) −1 \*(( 𝑥 𝑦 ) T − 𝛽) ≤ 𝑞𝑠2 𝑥𝛼}

Значения параметров нам известны:

Значение (𝑋 𝑋𝑇 ) −1 также было вычислено ранее как S1:

# 4

**Сформулировать гипотезу независимости переменной Y от переменной X. Провести проверку значимости.**

Гипотеза независимости переменной Y от Х:

𝐻0: 𝛽1 = 0

Критерий:

Найдем статистику F-критерия и P-значение:

## FST <- bht[2]^2/S1[2, 2]/sht

## pv.f <- pf(FST, 1, n-2, lower.tail=FALSE)

Получили 2.193198 и 0.1451561: 𝛼<PV, значит гипотеза принимается.

И еще:

FST: 2.193198

Граница критической области: 7.194218

## qf(0.01, 1, n-2, lower.tail=FALSE)

# 5

**Сформулировать модель, включающую дополнительный член с X2. Построить МНК оценки параметров β1, β2, β3 в данной модели. Изобразить графически полученную регрессионную зависимость.**

Добавим в модель член с X2:

Y=β1+ β1x+ β1x2+eps

Найдем МНК оценки:

## x0 <- array(1, dim=n)

## X <- t(matrix(c(x0, x, x^2), nrow=n, ncol=3))

## Y <- as.matrix(y)

## S <- X%\*%t(X)

## S1 <- solve(S)

## bht <- S1%\*%X%\*%Y

Получили

𝛽1 = 11.7792131

𝛽2 = -0.9922159

𝛽3 = 0.1126049

И изобразим полученную регрессионную зависимость:

## plot(x, y)

## x1 <- seq(from = min(x), to = max(x), by = (0.5))

## y1 <- bht[1] + bht[2] \* x1 + bht[3] \* x1^2

## points(x1, y1, "l", col="red", lwd=2)

# 

# 6

**Построить несмещенную оценку дисперсии. Провести исследование нормальности ошибок как в п.2.**

Несмещенная оценка дисперсии:

## eps <- (YM - t(X)%\*%bht)

## n = length(y)

## r = dim(X)[1]

## SSE = sum(eps^2)

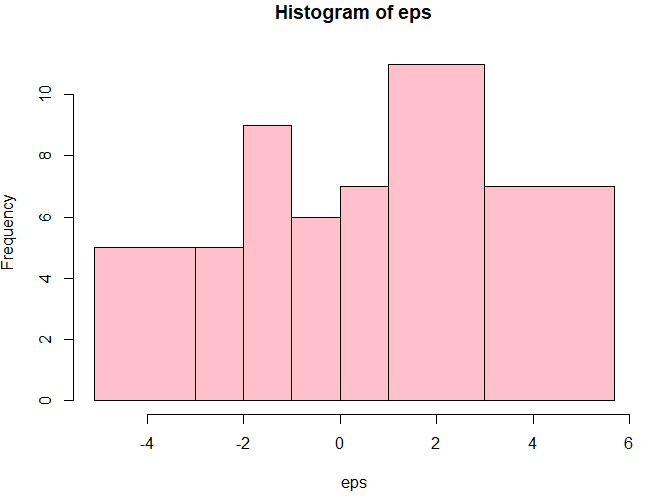
## sht = SSE/(n-r)

## print(sht)

Получили 6.219721.

Построим гистограмму на базе ошибок:

## hh = hist(eps, breaks=c(min(eps), -3, -2, -1, 0, 1, 3, max(eps)), col="pink", freq=TRUE)



Проверка гипотезы нормальности ошибок на уровне α по χ2:

## nu = hh$counts;

## breaks = hh$breaks;

## r = length(breaks) - 1

## csq0<-function(s){

## if (s>0){

## p<-pnorm(breaks[2:length(breaks)],0,s)-pnorm(breaks[1:(length(breaks)-1)],0,s)

## return(sum((nu-n\*p)^2/n/p))

## } else {

## return(Inf)

## }

## }

## csq.s<-nlm(csq0,p=sqrt(sht))$minimum

## xa1 = qchisq(0.99, r - 3)

## print(csq.s < xa1)

TRUE, значит гипотеза принимается.

Оценивание расстояния полученной оценки до класса нормальных распределений по Колмогорову:

## kolm.stat<-function(s){

## sres<-sort(eps)

## fdistr<-pnorm(sres,0,s)

## max(

## abs(c(0:(n-1))/n-fdistr),

## abs(c(1:n)/n-fdistr)

## )

## }

## ks.dist<-nlm(kolm.stat,p=sqrt(sht))$minimum

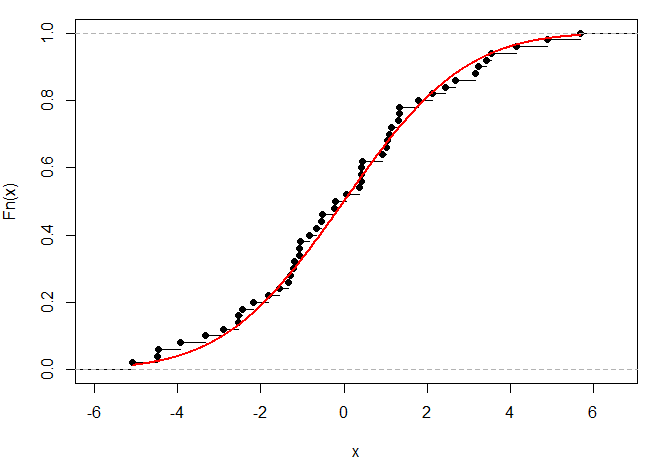
## plot.ecdf(eps)

## x2<-c(0:1000)\*(max(eps)-min(eps))/1000+min(eps)

## y2<-pnorm(x2,0,nlm(kolm.stat,p=sqrt(s2))$estimate)

## points(x2,y2,"l",col="red",lwd=2)

Получим расстояние 0.05779356



# 7

**В предположении нормальности ошибок построить доверительные**

**интервалы для параметров β1, β2, β3 уровня 1-α. Написать уравнение доверительного эллипсоида уровня доверия 1-** **α**

Построим ДИ для параметров:

## C<-diag(c(1,1,1))

## ph<-bht

## V<-diag(S1)

## xa<-qt(1-al/2,n-3)

## s1<-sqrt(s2)

## d<-xa\*s1\*sqrt(V)

## CI<-data.frame(lw=ph-d,up=ph+d)

Получили

β1: [9.1546564, 14.4037698]

β2: [-2.7957178, 0.8112861]

β3: [-0.1609643, 0.3861741]

Доверительный эллипсоид имеет формулу

𝐴𝛼 = {𝑥, 𝑦: (( 𝑥 𝑦 z ) T − 𝛽) 𝑇 \* (𝑋 𝑋𝑇 ) −1 \*(( 𝑥 𝑦 z ) T − 𝛽) ≤ 1}

Значения параметров, а также (𝑋 𝑋𝑇 ) −1, равны

𝛽 = [11.7792131, 0.9922159, 0.1126049]

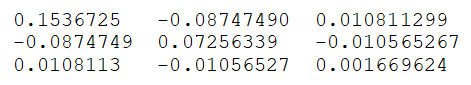
(𝑋 𝑋𝑇 ) −1:

## 0.1536725 -0.08747490 0.010811299

## -0.0874749 0.07256339 -0.010565267

## 0.0108113 -0.01056527 0.001669624

Получим неравенство:

𝐴𝛼 = {𝑥, 𝑦: (( 𝑥 𝑦 z ) T − (11.7792131, 0.9922159, 0.1126049)T) 𝑇 \* () \*(( 𝑥 𝑦 z ) T − (11.7792131, 0.9922159, 0.1126049)T) ≤ 1}

# 8

**Сформулировать гипотезу линейной регрессионной зависимости переменной Y от переменной X и проверить ее значимость на уровне α.**

Проведем проверку значимости регрессии. Если 𝑦 зависит от 𝑥, то параметры при 𝑥 𝛽3 = 0. Тогда нулевая гипотеза: 𝛽3 = 0.

Найдем статистику F-критерия и P-значение:

## FST <- bht[3]^2/S1[3, 3]/sht

## pv.f <- pf(FST, 1, n-2, lower.tail=FALSE)

Получили 1.221026 и 0.2746676: 𝛼=0.01<PV= 0.2746676, значит гипотеза принимается.