HW_3/9

Larin Anton

12/18/2020

ИДЗ 3 Статан

ИДЗ 9 Матпакеты

Вар. 12 (83832020)

Результаты статистического эксперимента приведенны в таблице 1. Требуется оценить характер (случайной) зависимости переменной Y от переменной X.

1. Построить графически результаты эксперимента. Сформулировать линейную регрессионную модель переменной Y по переменной X. Построить МНК оценки параметров сдвига β_0 и масштаба β_1. Построить полученную линию регрессии. Оценить визуально соответствие полученных данных и построенной оценки.
2. Построить и интерпретировать несмещенную оценку дисперсии. На базе ошибок построить гистограмму с шагом h. Проверить гипотезу нормальности ошибок на уровне α по χ^2. Оценить расстояние полученной оценки до класса нормальных распределений по Колмогорову. Визуально оценить данный факт.
3. В предположении нормальности ошибок построить доверительные интервалы для параметров β_0 и β_1 уровня доверия $1-\alpha$. Построить доверительный эллипс уровня доверия $1-\alpha$ для (β_0, β_1) (вычислить его полуоси).
4. Сформулировать гипотезу независимости переменной Y от переменной X. Провести проверку значимости.
5. Сформулировать модель, включающую дополнительный член с X^2. Построить МНК оценки параметров $\beta_1, \beta_2, \beta_3$ в данной модели. Изобразить графически полученную регрессионную зависимость.
6. Построить несмещенную оценку дисперсии. Провести исследование нормальности ошибок как в п.3.
7. В предположении нормальности ошибок построить доверительные интервалы для параметров $\beta_1, \beta_2, \beta_3$ уровня $1-\alpha$. Написать уравнение доверительного эллипсоида уровня доверия $1-\alpha$.
8. Сформулировать гипотезу линейной регрессионной зависимости переменной Y от переменной X и проверить ее значимость на уровне α.
9. Интерпретировать полученные результаты. Написать отчет.

Таблица 1 $\alpha_1 = 0.20$; $h = 1.40$.

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Y	9.60	6.02	3.23	8.41	3.63	8.25	6.25	7.48	7.61	5.64	4.40	4.09	7.04	8.93	6.02	8.49	7.08
X	1	3	0	3	1	2	1	2	1	2	2	3	1	2	1	3	2

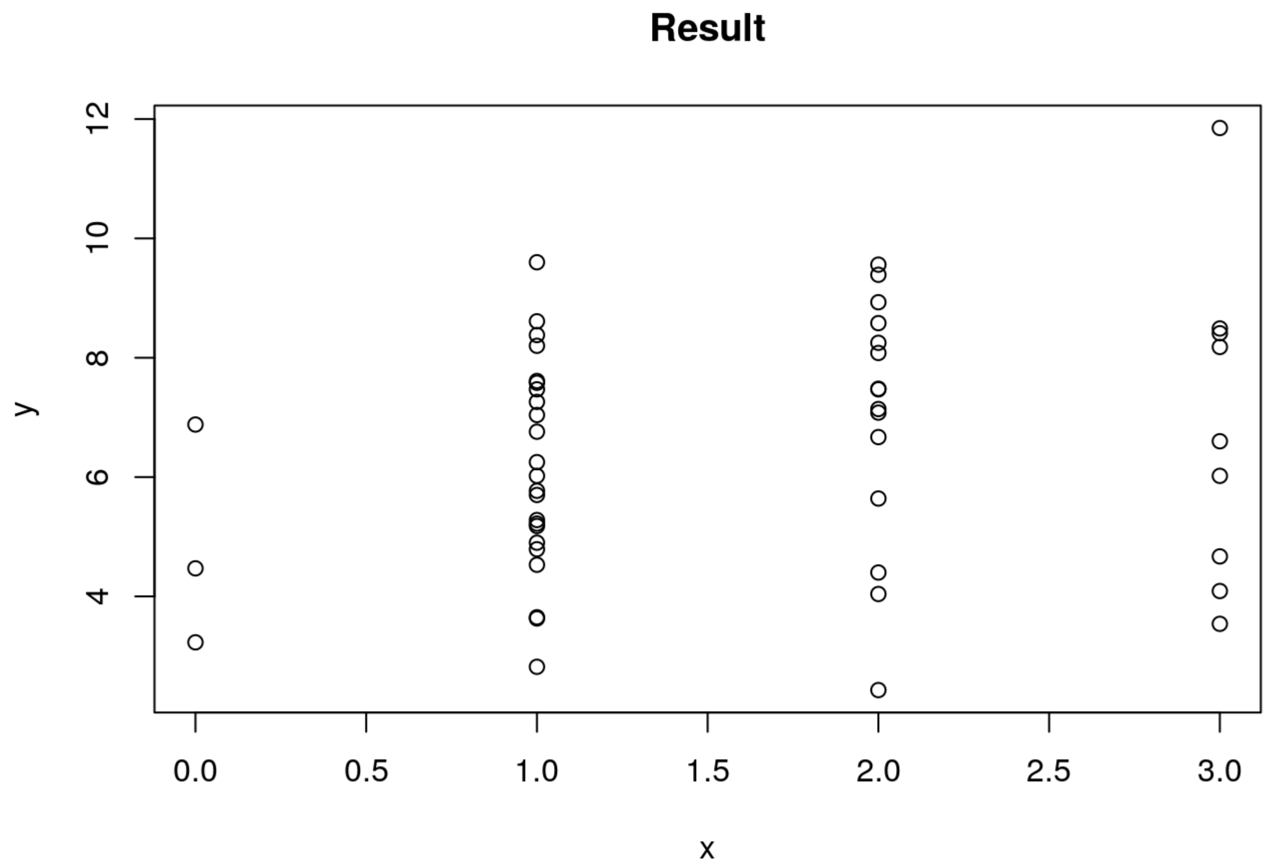
No	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Y	7.58	8.20	7.47	7.47	6.60	7.14	5.18	3.54	5.28	5.77	4.90	4.67	5.70	6.88	4.53	7.26	8.18
X	1	1	1	2	3	2	1	3	1	1	1	3	1	0	1	1	3

No	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Y	5.22	11.85	8.58	2.82	2.43	3.65	6.67	8.08	6.76	4.47	4.04	8.61	9.39	8.38	4.79	9.56
X	1	3	2	1	2	1	2	2	1	0	2	1	2	1	1	2

Image alt

#Ход работы ##Графические представление

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



Линейная регрессионная модель: $y = \beta_0 + \beta_1 * x$

1

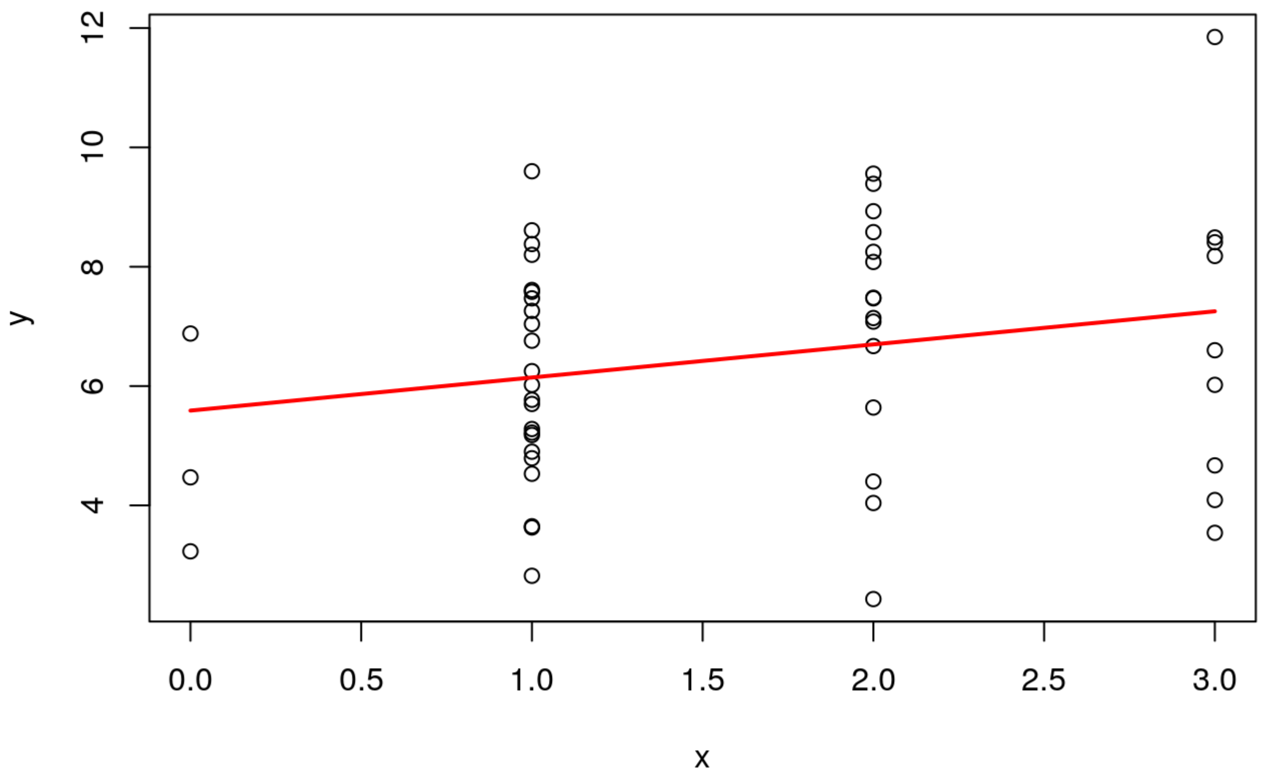
МНК оценка параметров сдвига β_0 и масштаба β_1

```
n<-length(y)
x0<-array(1,dim=n)
X<-t(matrix(c(x0,x),nrow=n,ncol=2))
Y<-as.matrix(y)
S<-X%*%t(X)
S1<-solve(S)
bhat<-S1%*%X%*%Y
```

Результат:

$\beta_0 = 5.5889333$
$\beta_1 = 0.5546667$

Нарисуем полученную регрессионную зависимость



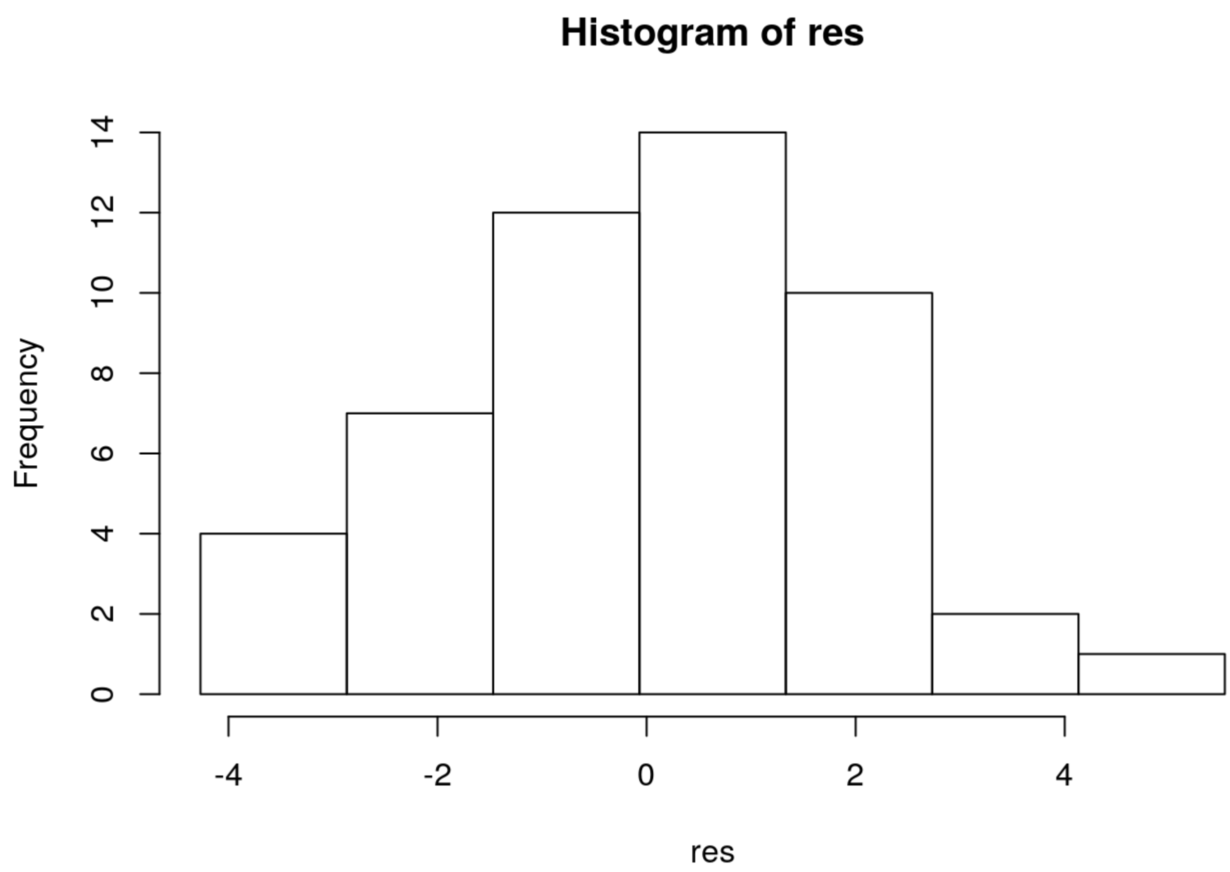
2

Построение несмещенной оценки дисперсии

```
  res<-Y-t(X)%*%as.matrix(bhat)
SS<-sum(res^2)
s2<-SS/(n-2)
```

Результат: 4.0704077

Построение гистограммы с шагом h = 1.4 на базе ошибок



Проверка гипотезы нормальности ошибок на уровне α по χ^2 $H_0 : Y - X^T\beta \sim (0, \sigma^2)$

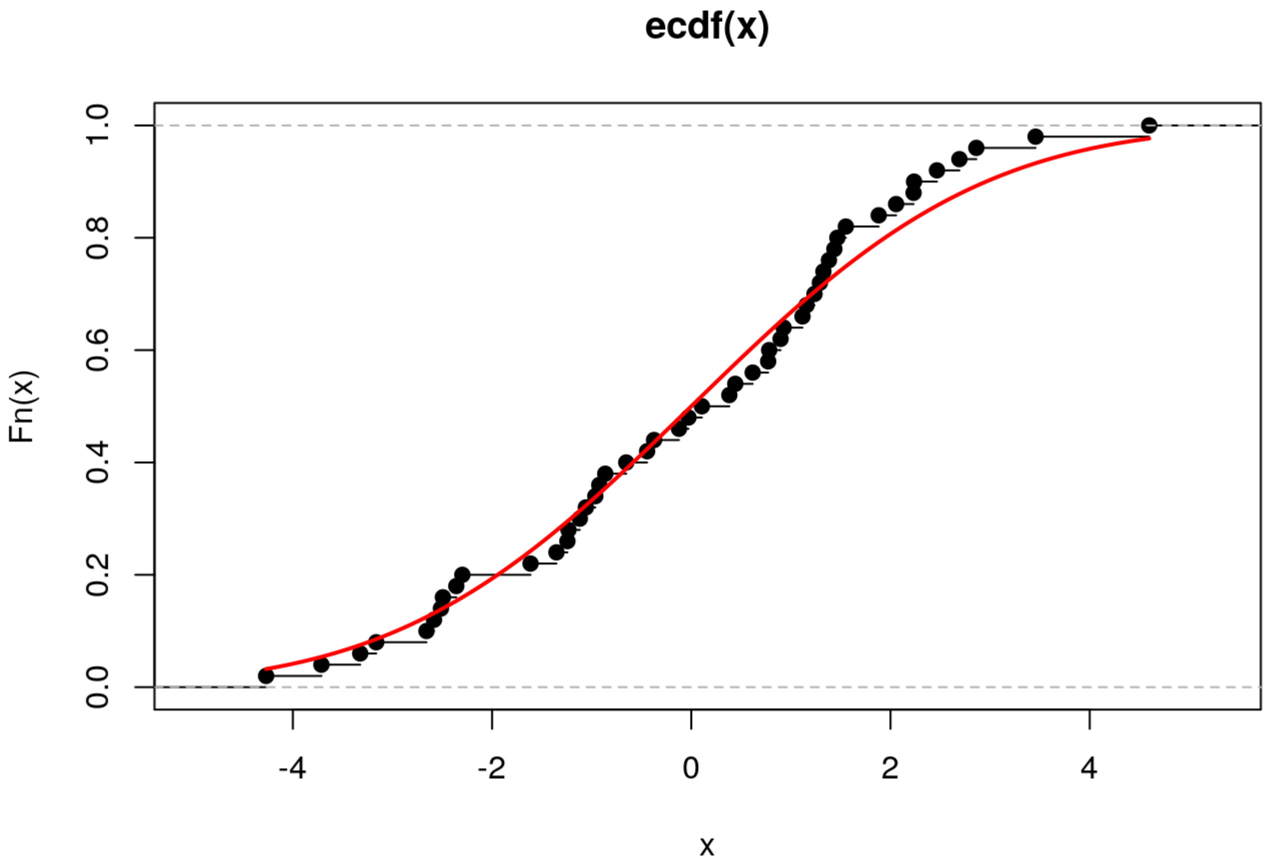
```
hh<-hist(res,breaks=brk,plot=FALSE)
nu<-hh$counts
breaks = hh$breaks;
r = length(breaks) - 1
l.b<-length(brk)
csq0<-function(s){
  if (s>0){
    p<-pnorm(brk[2:l.b],0,s)-pnorm(brk[1:(l.b-1)],0,s)
    return(sum((nu-n*p)^2/n/p))
  } else {
    return(Inf)
  }
}
csq.s<-nlm(csq0,p=sqrt(s2))$minimum
pv<-pchisq(csq.s,r - 2,lower.tail=FALSE)
```

Результат: FALSE

Оценим расстояние оценки до класса нормальных распределений оп Колмогорову

```
kolm.stat<-function(s){
  sres<-sort(res)
  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(s2))$minimum
```

Полученое расстояние: 0.0708533



3

Построим ДИ для параметров с уровнем доверия $1-\alpha(0.8)$

```
C<-diag(c(1,1))
ph<-bhat
V<-diag(S1)
xa<-qt(1-al/2,n-2)
s1<-sqrt(s2)
d<-xa*s1*sqrt(V)
CI<-data.frame(lw=ph-d,up=ph+d)
```

Для $\beta_0 : [4.7975981, 6.3802685]$
Для $\beta_1 : [0.1177249, 0.9916084]$

Доверительный эллипс можно вычислить как

$A\alpha = \{x, y : ((xy)^T - \beta)^T * (XX^T)^{-1} * ((xy)^T - \beta) \leq qs^2 x\alpha\}$
Где:

$\beta =$

```
##           [,1]
## [1,] 5.5889333
## [2,] 0.5546667
```

$(XX^T) =$

```
##             [,1]        [,2]
## [1,]  0.09111111 -0.04444444
## [2,] -0.04444444  0.02777778
```

4

Гипотеза независимости Y от X: $H0 : \beta 1 = 0$
Критерий:
$\Phi(x) = \begin{cases} 0, \alpha < PV \\ 1, \alpha > PV \end{cases}$

Найдем статистику F-критерия и P-значение:

```
FST<-bhat[2]^2/V[2]/s2
pv.f<-pf(FST,1,n-2,lower.tail=FALSE)
```

Получим 2.7210012 и 0.1055659
Результат: FALSE

5

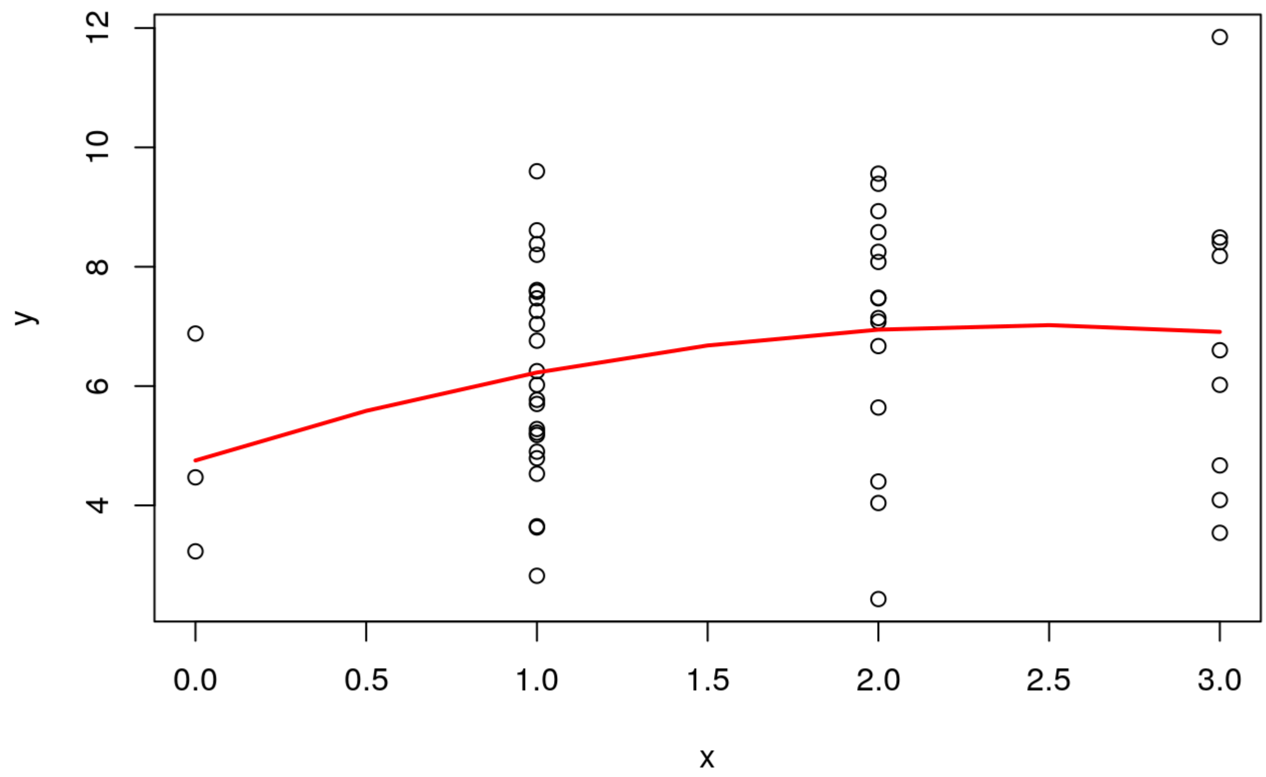
Добавим в модель член с X^2:
$Y = \beta_1 + \beta_2 x + \beta_3 x^2 + \varepsilon$

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
```

$\beta 1 = 4.7528712$
$\beta 2 = 1.8515702$
$\beta 3 = -0.3777389$

Нарисуем полученную регрессионную зависимость



6

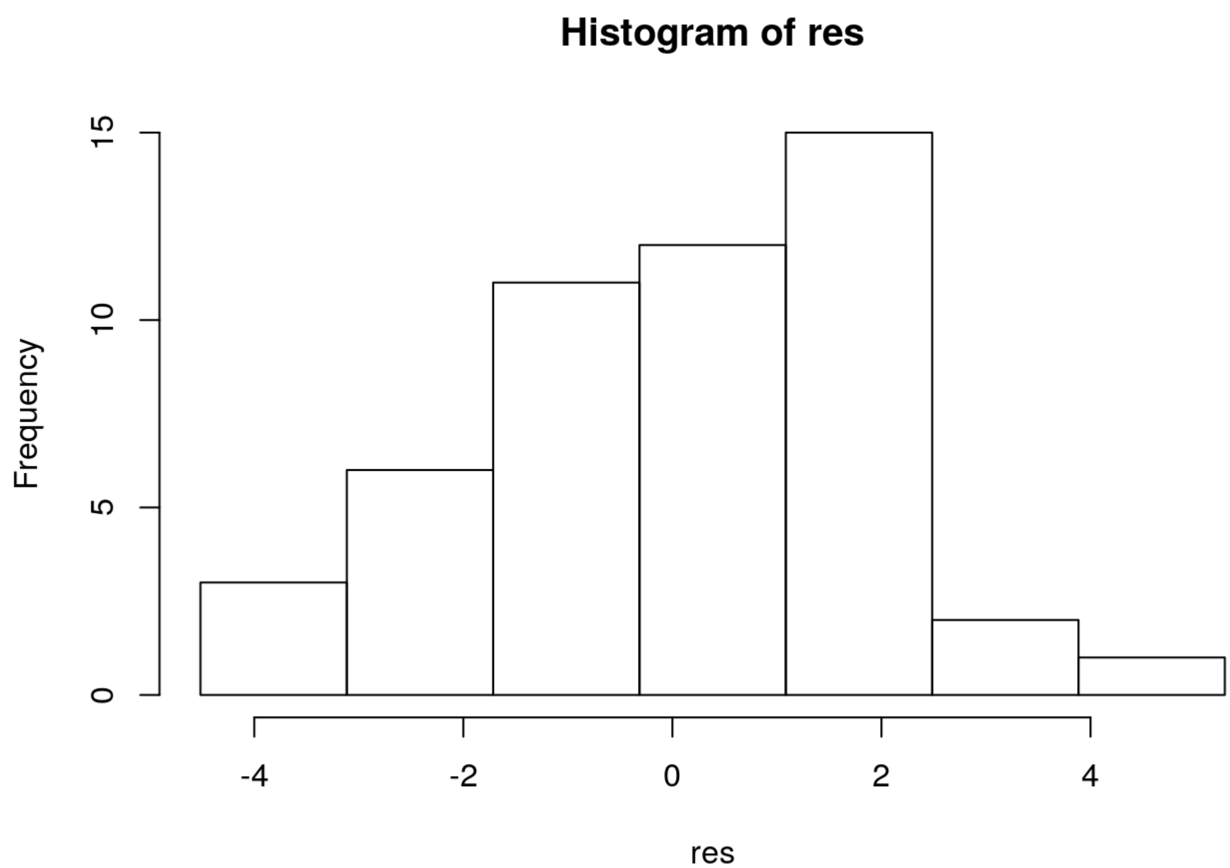
Несмещенная оценка дисперсии

```
res<-Y-t(X)%*%as.matrix(bht)
SS<-sum(res^2)
s2<-SS/(n-2)
```

Результат: 3.9820608

Гистограмма на базе ошибок

```
brk<-seq(min(res), max(res) + h, by=h)
hist(res,breaks=brk)
```



Проверка гипотезы нормальности ошибок на уровне α по χ^2

```
l.b<-length(brk)
brk[1]<- -Inf
brk[l.b]<-Inf
#r = length(breaks) - 1
hh<-hist(res,breaks=brk,plot=FALSE)
nu<-hh$counts
breaks = hh$breaks;
r = length(breaks) - 1

l.b<-length(brk)
csq0<-function(s){
  if (s>0){
    p<-pnorm(brk[2:l.b],0,s)-pnorm(brk[1:(l.b-1)],0,s)
    return(sum((nu-n*p)^2/n/p))
  } else {
    return(Inf)
  }
}
csq.s<-nlm(csq0,p=sqrt(s2))$minimum
```

```
## Warning in nlm(csq0, p = sqrt(s2)): NA/Inf replaced by maximum positive value
```

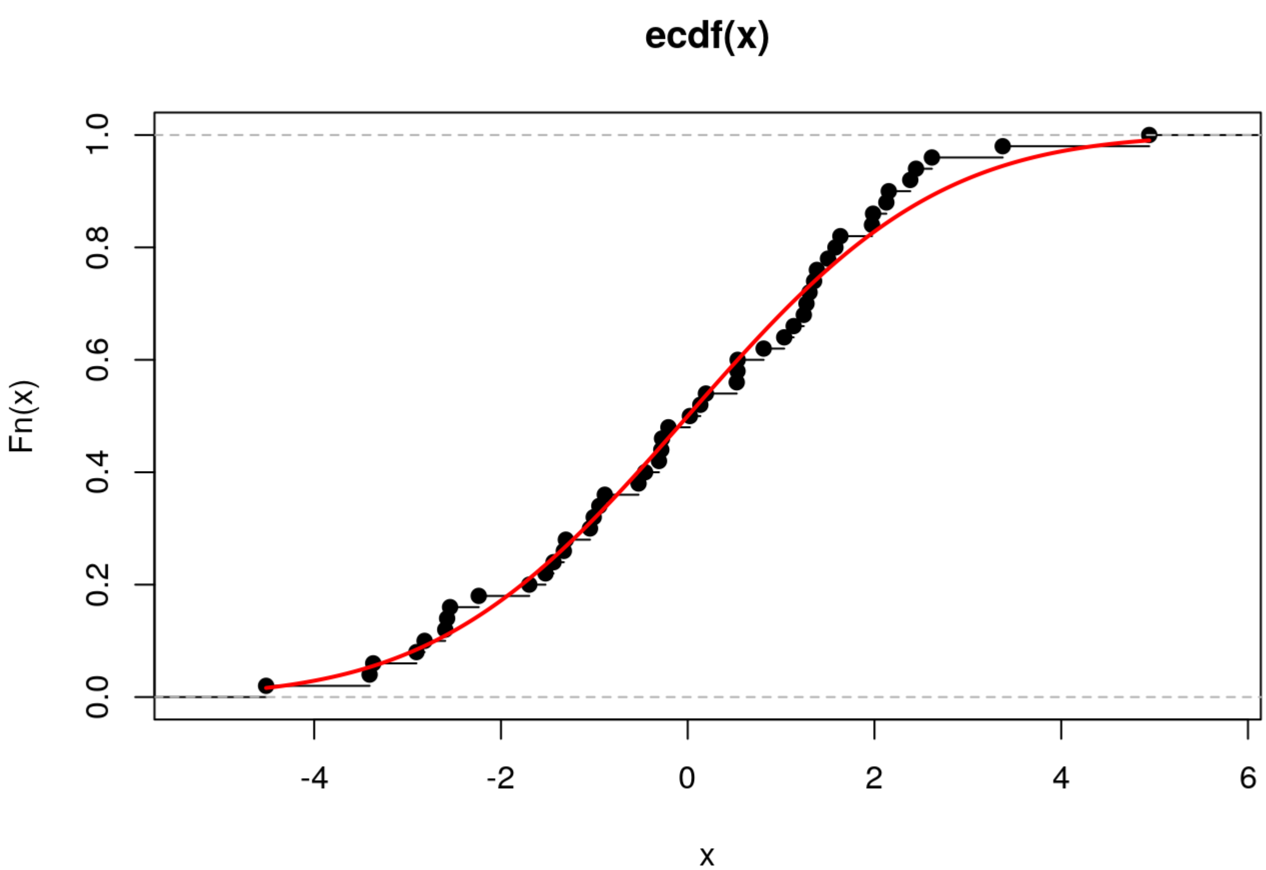
```
pv<-pchisq(csq.s,r-3,lower.tail=FALSE)
```

Результат: FALSE

Оценка расстояния до нормального рапределения по Колмагорову

```
kolm.stat<-function(s){
  sres<-sort(res)
  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(s2))$minimum


x2<-c(0:1000)*(max(res)-min(res))/1000+min(res)
y2<-pnorm(x2,0,nlm(kolm.stat,p=sqrt(s2))$estimate)
```



Полученое расстояние: 0.0677388

7

Построим ДИ для параметров

```
C<-diag(c(1,1,1))
ph<-bht #t(C)%*%bhat
V<-diag(S1) # diag(C%*%S1%*%t(C))
xa<-qt(1-al/2,n-2)
s1<-sqrt(s2)
d<-xa*s1*sqrt(V)
CI<-data.frame(lw=ph-d,up=ph+d)
```

Полученые интервалы:

```
##          lw        up
## 1  3.4410260 6.0647164
## 2  0.1622953 3.5408451
## 3 -0.8533866 0.0979088
```

Доверительный эллипсоид имеет форму

$A\alpha = \{x, y : ((xyz)^T - \beta)^T * (XX^T)^{-1} * ((xyz)^T - \beta) \leq 1\}$
Где:
$\beta =$

```
##            [,1]
## [1,]  4.7528712
## [2,]  1.8515702
## [3,] -0.3777389
```

$(XX^T) =$

```
##             [,1]       [,2]        [,3]
## [1,]  0.25594437 -0.3001346  0.07447286
## [2,] -0.30013459  0.4244056 -0.11552266
## [3,]  0.07447286 -0.1155227  0.03364738
```

8

Гипотеза линейной регрессионной зависимости Y от X: $H0 : \beta 3 = 0$
Критерий:

$$\Phi(x) = \begin{cases} 0, \alpha < PV \\ 1, \alpha > PV \end{cases}$$

Найдем статистику F-критерия и P-значение:

```
FST<-bht[3]^2/S1[2,2]/s2
pv.f<-pf(FST,1,n-2,lower.tail=FALSE)
```

Получим 0.0844295 и 0.7726338
Результат: TRUE