

Задача 1.

Имаме две хипотези.

Но: **няма връзка** между ранга и честотата на ранявания и убийства

H1: **има връзка** между ранга и честотата на ранявания и убийства

Изпълняваме следния код:

```
x = c(43, 12, 8, 96, 32, 18, 123, 35, 44)
m = matrix(x, nrow=3, ncol=3)
chisq.test(m)
```

Резултат:

Pearson's Chi-squared test

data: m

X-squared = 6.749, df = 4, p-value = 0.1498

Извод: $p\text{-value} > 0.05$, следователно приемаме хипотезата Но, т.е. няма връзка между ранга и честотата на раняванията и убийствата.(случайните величини са независими)

Задача 2

A)

```
file = read.csv(file.choose(), sep=";", dec=",")
data = data.frame(file$y, file$x1, file$x2, file$x3, file$x4, file$x5)
plot(data)
y = file$y
x1 = file$x1
x2 = file$x2
x3 = file$x3
x4 = file$x4
x5 = file$x5
```

Правим първият модел.

```
l1=lm(y~x1+x2+x3+x4+x5)
```

```
summary(l1)
```

Получаваме като резултат:

```
Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.2495	-0.6291	-0.1188	0.5927	3.0593

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
----------	------------	---------	----------

```

(Intercept)  -6.7536      4.2197  -1.601   0.1153
x1             3.7814      1.5521   2.436   0.0182 *
x2            -0.0912      0.1206  -0.756   0.4530
x3            -1.0823      0.5571  -1.943   0.0573 .
x4            -0.3051      0.4982  -0.612   0.5428
x5             1.1672      0.5938   1.966   0.0545 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.059 on 54 degrees of freedom
Multiple R-squared:  0.8421,    Adjusted R-squared:  0.8275
F-statistic:  57.6 on 5 and 54 DF,  p-value: < 2.2e-16

```

Като забелязваме, че стойностите пред x2 и x4 са големи, което ни подсказва да направим нов модел от който да ги изключим. Добавяме и свободен член -1

```

l2=lm(y~x1+x3+x5-1)
summary(l2)

```

```

Call:
lm(formula = y ~ x1 + x3 + x5 - 1)

```

```

Residuals:
    Min       1Q   Median       3Q      Max
-2.1286 -0.6963 -0.2450  0.7906  3.0209

```

```

Coefficients:
    Estimate Std. Error t value Pr(>|t|)
x1    5.0254     0.5395   9.315 4.75e-13 ***
x3   -2.0224     0.1578 -12.814 < 2e-16 ***
x5    0.1244     0.0890   1.398  0.168
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 1.074 on 57 degrees of freedom
Multiple R-squared:  0.924,    Adjusted R-squared:  0.92
F-statistic: 231.1 on 3 and 57 DF,  p-value: < 2.2e-16

```

Б)

Полученото Adjusted R-squared: 0.92
Т.е. 92% се описват от модела.

В) Бихме могли да приемем, че коефициента пред x1 е 5, защото в summary(l2), Estimate Std. Е 5.0254, което е доста близо до 5.

Г)

```

z = data.frame(x1=2,x2=4.5,x3=3.9,x4=4,x5=6.5)
predict.lm(l2,z,interval='confidence',level=0.95)

```

Резултат:

```

    fit      lwr      upr
1 2.972222 2.705072 3.239373

```

Задача 3.

Имаме две хипотези.

Но : теглото на бебетата **зависи** от доходите на родителите

H1: теглото на бебетата **не зависи** от доходите на родителите

Първо ще намерим средното тегло на децата от всяка категория.

```
avgWt = seq(0,9)
for (i in 0:9) {
  avgWt[i+1] = mean(wt[inc == i])
}
```

След това пускаме shapiro.test за да проверим дали данните са нормално разпределени

```
shapiro.test(avgWt)
```

Получаваме $p\text{-value}=0.03 < 0.05$ и следователно отхвърляме, че са нормално разпределени. Тоест не са!

Затова използваме wilcox.test за да проверим хипотезата

```
wilcox.test(avgWt)
```

Полученото $p\text{-value}=0.001953 < 0.05$ и следователно отхвърляме хипотезата Но и приемаме алтернативата H1.
Т.е. теглото на бебетата **не зависи** от доходите на родителите.

Задача 4.

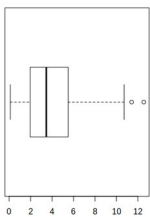
Първо зареждаме данните от файла и в променливи.

```
data = read.csv(file.choose(), sep=',', dec='.')
x1=data$x1
x2=data$x2
x3=data$x3
x4=data$x4
x5=data$x5
x6=data$x6
```

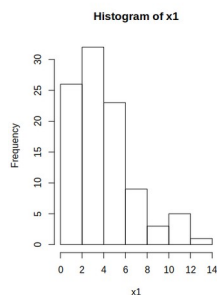
Започваме да изследваме графично променливите една по една.

X1

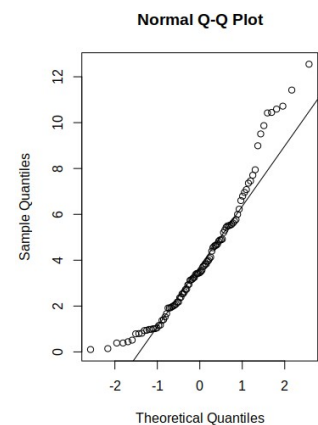
boxplot(x1, horizontal=T)



hist(x1)



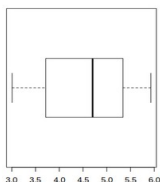
qqnorm(x1)
qqline(x1)



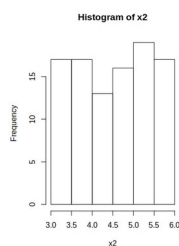
Според графиките заключаваме, че x1 е експоненциално разпределена.

X2

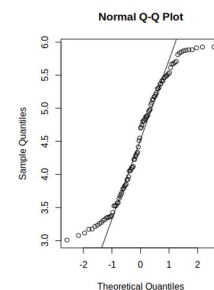
boxplot(x2, horizontal=T)



hist(x2)



qqnorm(x2)
qqline(x2)



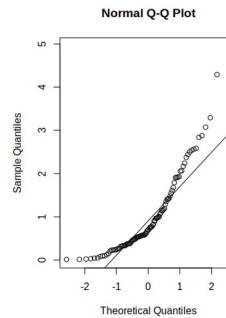
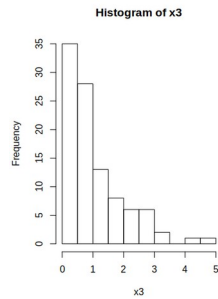
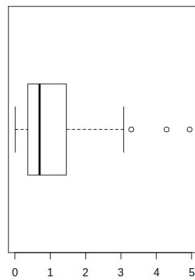
Според графиките заключаваме, че x_2 е равномерно разпределено.

X3

`boxplot(x3,horizontal=T)`

`hist(x3)`

`qqnorm(x3)`
`qqline(x3)`



Според графиките X3 е експоненциално разпределена.

X4

Аналогично можем да видим, че x_4 също е експоненциално разпределена

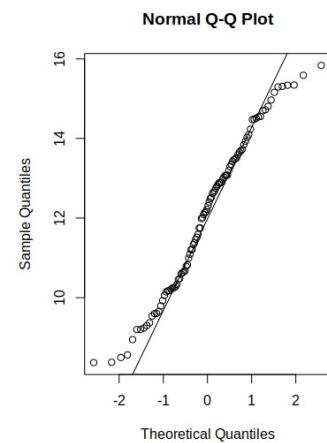
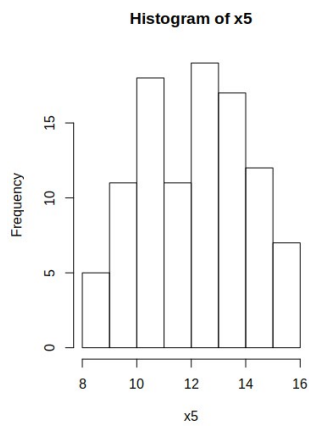
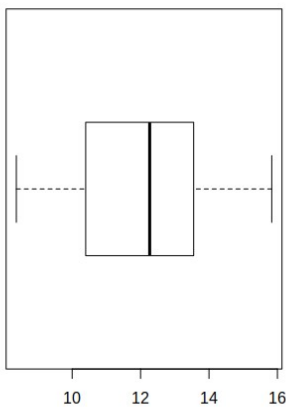
X5

`boxplot(x5,horizontal=T)`

`hist(x5)`

`qqnorm(x5)`

`qqline(x5)`



Според графиките x_5 е нормално разпределена

X6

Аналогично според графиките и X6 е нормално разпределена случайна величина.

Сега ще направим проверка на хипотезата.

Но: X_1 принадлежи на $X^2(4)$

H1: X_1 не принадлежи на $X^2(4)$

```
data = table(cut(x1, breaks = c(0,2,4,6,12)))
probTable = c(0,0,0,0)
probTable[1] = pexp(2, 4) - pexp(0, 4)
probTable[2] = pexp(4,4) - pexp(2,4)
probTable[3] = pexp(6,4) - pexp(4,4)
probTable[4] = 1 - pexp(6,4)
chisq.test(data,p=probTable)
```

Получаваме:

Chi-squared test for given probabilities

data: t

X-squared = 7.8164e+10, df = 3, p-value < 2.2e-16

Полученото p-value е прекалено ниско затова отхвърляме хипотезата и приемаме алтернативата H1. т.е. не принадлежи