

СЕМ

Домашна работа №3

Изготвил:

Емил Кирилов

Група 1

Фак. Номер – 61763

## Задача 1

В задачата се иска да формулираме хипотеза за проверка на твърдението, че резултатите от измерванията на експерта се различават от тези на машината.

**Хипотезата** е следната:

**Но:**  $i$ -то измерване-експерт  $= i$ -то измерване-машина

**На:**  $i$ -то измерване-експерт  $\neq i$ -то измерване-машина

### Критична област

Имаме двустранна КО - при голяма разлика, независимо  $\pm$ , твърдението Но, че измерванията съвпадат, се отхвърля.

Пресмятаме квантила на дясната (положителна) страна на КО. Лявата е симетрична.

```
1 library(UsingR)
2
3 alpha = 0.1
4 df = length(blood) - 1
5 |
6 qt(alpha / 2, df, lower.tail = FALSE)
> qt(alpha / 2, df, lower.tail = FALSE)
[1] 6.313752
```

**Критична област за  $\alpha=0.1$  -  $(-\infty, -6.313752) \cup (6.313752, \infty)$**

### Тестване на хипотезата

```
> # тестова статистика и p-value
> (mean(blood$Machine - blood$Expert) * sqrt(length(blood$Machine)))/sd(blood$Machine - blood$Expert)
[1] 0.681617
> 2 * pt(q = ts, df = 14, lower.tail = FALSE)
[1] 0.5066011
>
> t.test(blood$Machine, blood$Expert, paired = TRUE)
```

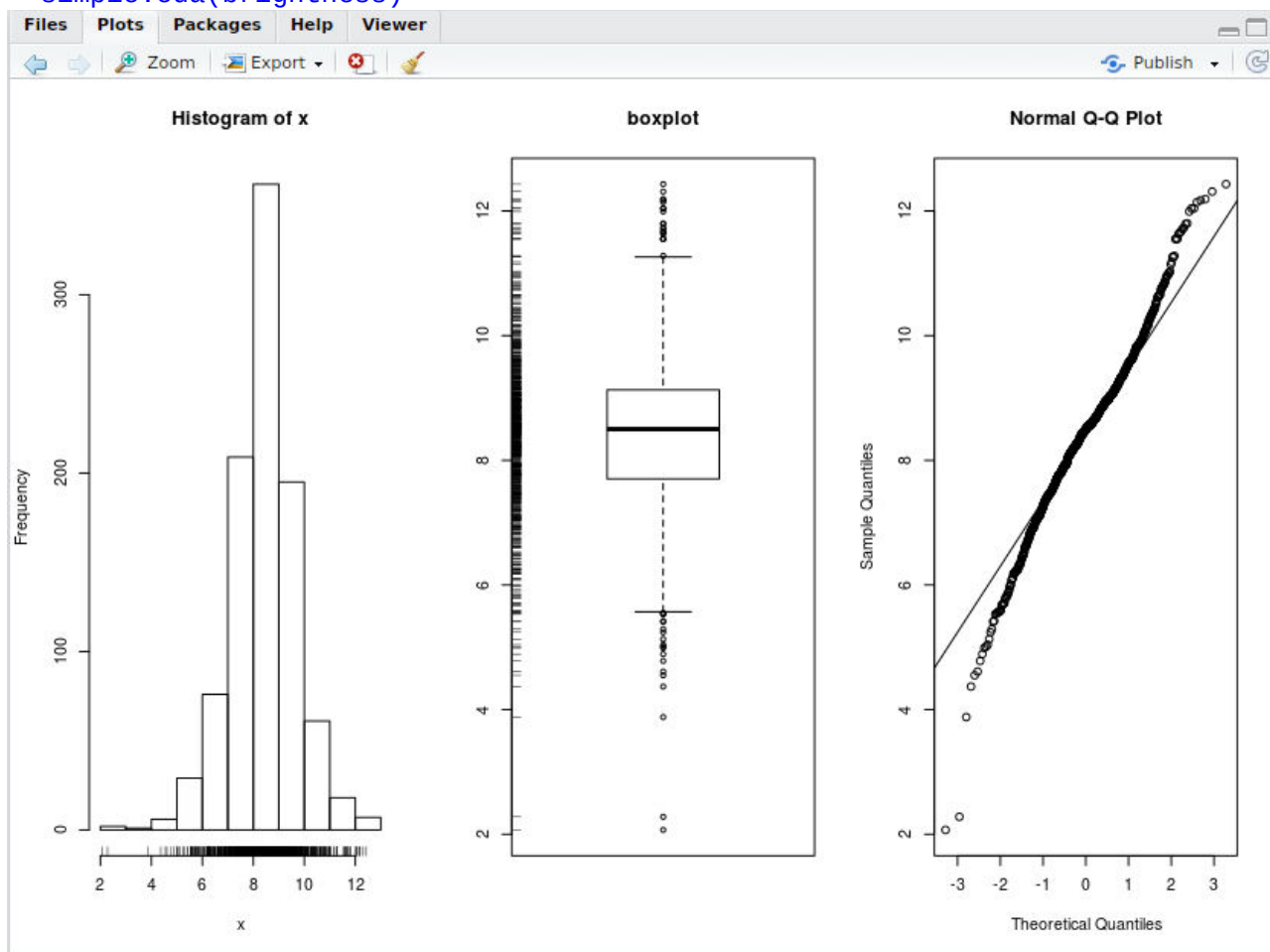
Paired t-test

```
data: blood$Machine and blood$Expert
t = 0.68162, df = 14, p-value = 0.5066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.146615  4.146615
sample estimates:
mean of the differences
```

R ни казва, че средното на разликата е **1**, а доверителният ни интервал е  $\sim(-2,4)$ , тоест  $H_0$  не с отхвърля. При толкова голямо **p-value**, а именно 0.5066, **няма как** да изберем  $\alpha > \text{p-value}$  при смислен тест, за да отхвърлим  $H_0$ .

## Задача 2

```
> simple.eda(brightness)
```



Данните изглеждат **симетрично** разположени и формата на хистограмата напомня на „камбанката“ в нормалното разпределение. На Q-Q графиката виждаме, че опашките леко **странят** от нормалното разпределение. Outlier-ите засилват това отдалечаване, а тях можем ясно да видим на boxplot-a.

Не знаем теоретичната дисперсия и използваме **t.test** за доверителния интервал.

```
> t.test(brightness, conf.level = 0.93)
```

One Sample t-test

```
data: brightness
t = 202.15, df = 965, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
93 percent confidence interval:
 8.342209 8.493277
sample estimates:
mean of x
 8.417743
```

## Задача 3

```
> prop.test(1000, 1337, alt = "greater", p = 0.7, conf.level = 0.95)
```

```
1-sample proportions test with continuity correction
```

```
data: 1000 out of 1337, null probability 0.7  
X-squared = 14.407, df = 1, p-value = 7.364e-05  
alternative hypothesis: true p is greater than 0.7  
95 percent confidence interval:  
 0.727541 1.000000  
sample estimates:  
      p  
0.7479432
```

R ни казва, че данните **не отхвърлят**  $H_0$ , която твърди, че повече от 70% от хората ползват интернет всеки ден.