

Вероятности и Статистика

Домашно 2

Владимир Ананиев, 81459

[Code ▾](#)

Зад. 1 Таблицата 'morley' съдържа резултатите от пет експеримента за измерване на скоростта на светлината. Как са разпределени данните в четвъртия експеримент? Постройте 97% доверителен интервал по данните от него.

[Hide](#)

```
library(UsingR)
attach(morley)
```

The following objects are masked from morley (pos = 3):

Expt, Run, Speed

The following objects are masked from morley (pos = 11):

Expt, Run, Speed

The following objects are masked from morley (pos = 12):

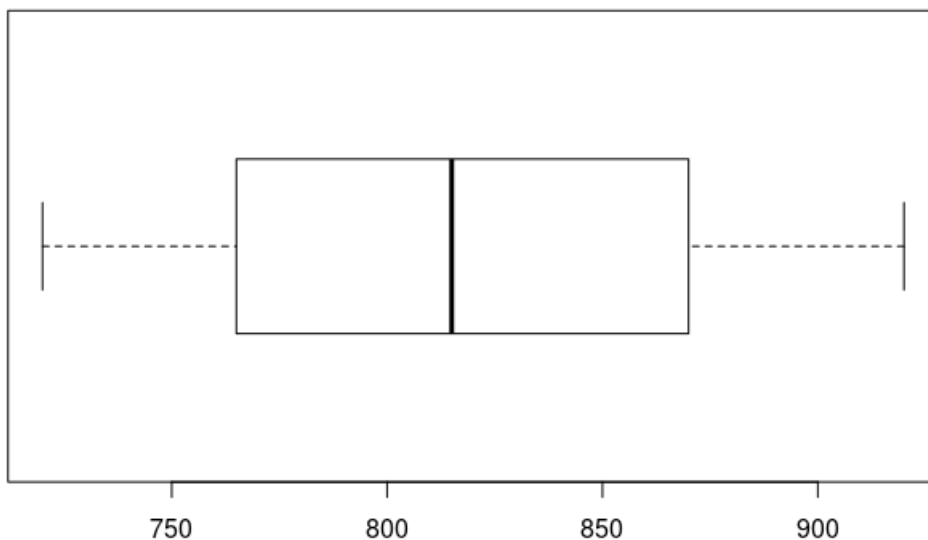
Expt, Run, Speed

[Hide](#)

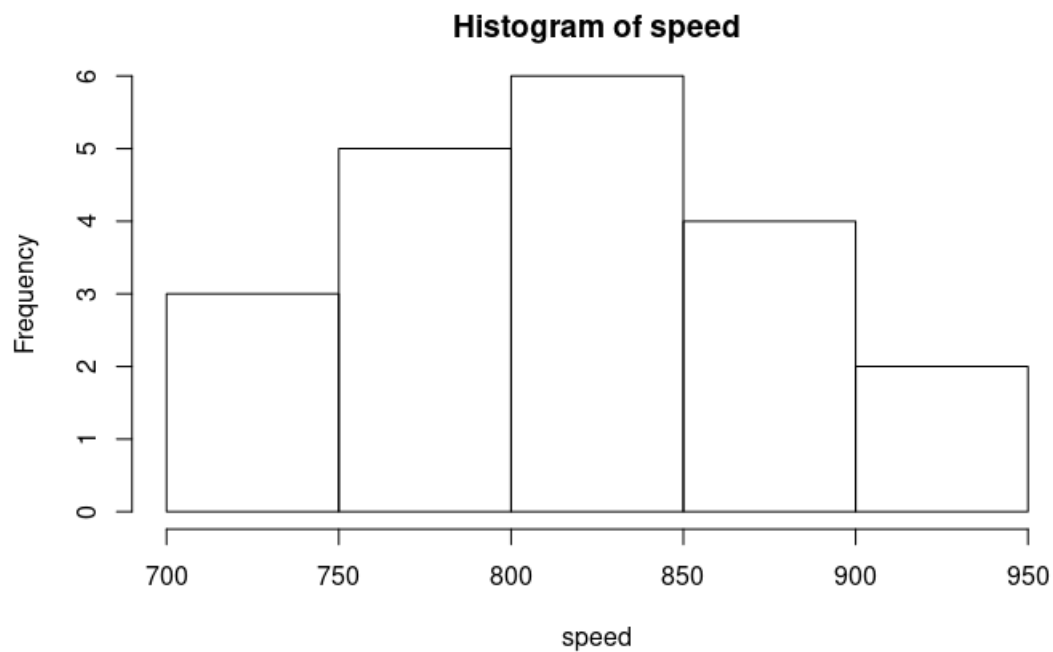
```
morley
```

[Hide](#)

```
speed <- morley[Expt == 4, 'Speed']
boxplot(speed, horizontal = T)
```

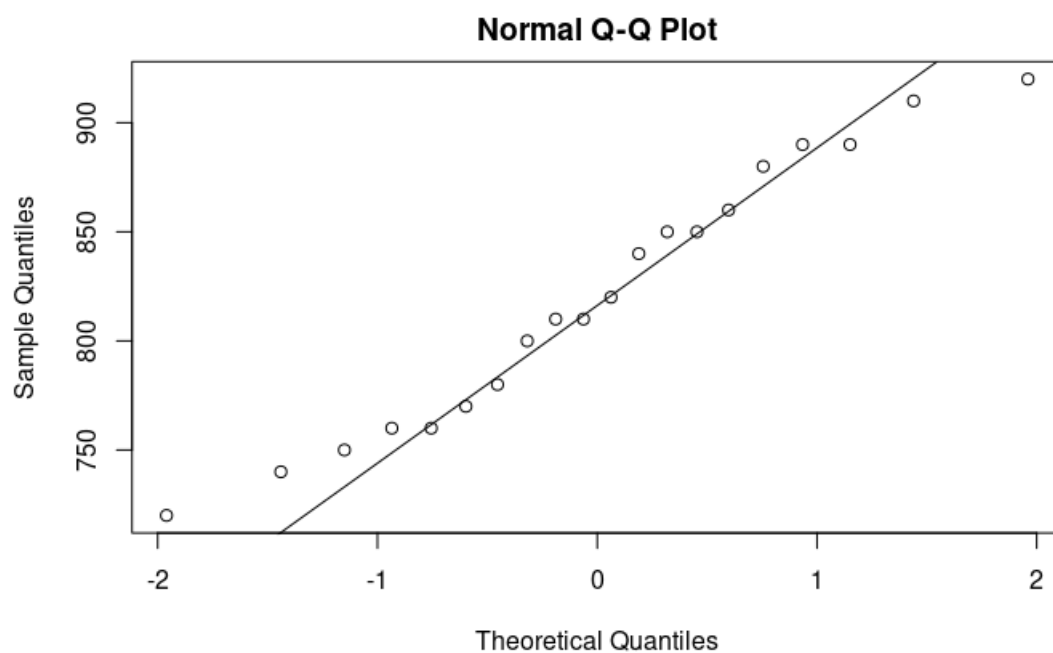
[Hide](#)

```
hist(speed)
```



Hide

```
qqnorm(speed)
qqline(speed)
```



Данните изглеждат нормално разпредени, ще направим Shapiro-Wilk тест за нормалност

Hide

```
shapiro.test(speed)
```

```
Shapiro-Wilk normality test

data:  speed
W = 0.96113, p-value = 0.5667
```

p-value > 0.5. Приемаме, че данните са нормално разпределени

...

[Hide](#)

```
t.test(speed, conf.int = TRUE, conf.level=0.97)
```

One Sample t-test

```
data: speed
t = 61.114, df = 19, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
97 percent confidence interval:
 789.008 851.992
sample estimates:
mean of x
 820.5
```

97% Confidence Interval: [789.008 - 851.992]

Зад. 2 Тества се ново лекарство за отслабване. Наблюдавани са 10 мишки, теглото им преди е било: 255, 258, 256, 201, 215, 207, 213, 215, 242, 216, след употребата на лекарството е: 250, 263, 254, 200, 213, 200, 210, 211, 232, 210. Вярно ли е, че лекарството води до отслабване?

[Hide](#)

```
x = c(255, 258, 256, 201, 215, 207, 213, 215, 242, 216)
y = c(250, 263, 254, 200, 213, 200, 210, 211, 232, 210)
shapiro.test(x)
```

Shapiro-Wilk normality test

```
data: x
W = 0.83443, p-value = 0.03783
```

[Hide](#)

```
shapiro.test(y)
```

Shapiro-Wilk normality test

```
data: y
W = 0.85288, p-value = 0.06286
```

Двете извадки не са нормално разпределени.

X - първоначалното тегло на мишките Y - теглото на мишките след приемане на лекарството

H0: Y = X

H1: Y < X

[Hide](#)

```
wilcox.test(y, x, alternative = 'less', paired = TRUE)
```

```
cannot compute exact p-value with ties
```

Wilcoxon signed rank test with continuity correction

```
data: y and x
V = 6.5, p-value = 0.01821
alternative hypothesis: true location shift is less than 0
```

p-value < 0.05 => Приемаме H1 и лекарството води до отслабване.

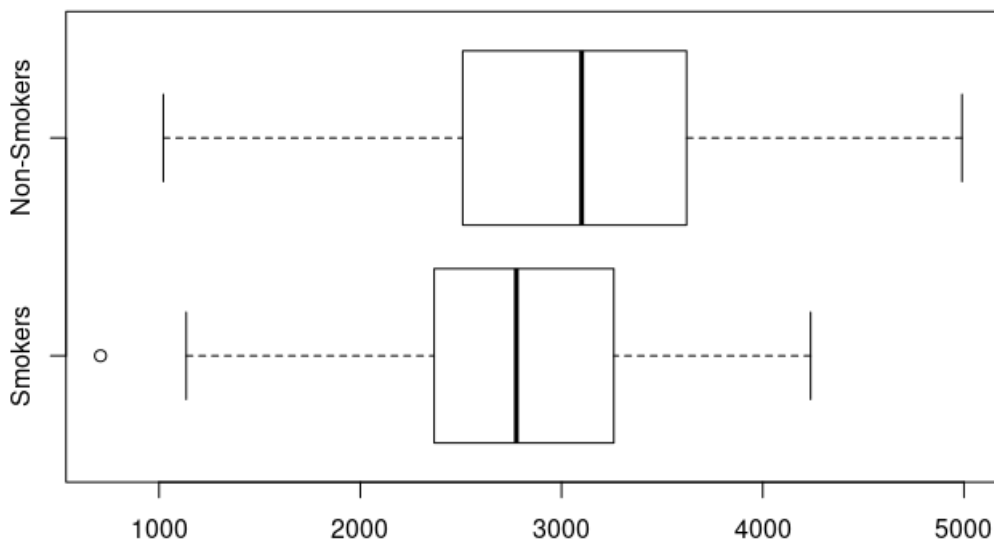
Зад. 3 По данните 'birthwt' от пакета 'MASS', можем ли да твърдим, че има връзка между тютюнопушенето и раждането на бебе под нормалните (2.5) кг? Проверете графично. Формулирайте и проверете хипотеза за това твърдение.

Hide

```
birthwt
```

Hide

```
smoker_babies = bwt[smoke == 1]
non_smoker_babies = bwt[smoke == 0]
boxplot(smoker_babies, non_smoker_babies, horizontal = TRUE, names=c('Smokers', 'Non-Smokers'))
```

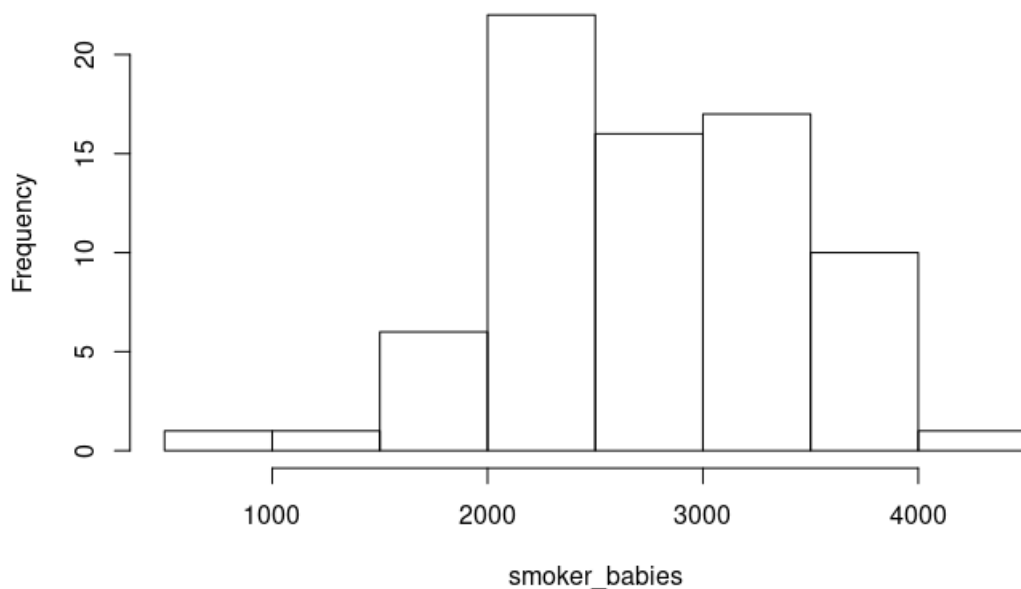


Забелязва се, че средното тегло на децата на майките пушачи е по-ниско от това на непушачите. Също максималните стойности при майките непушачи достигат 5кг, докато при майките пушачи това е само 4.2кг.

Hide

```
hist(smoker_babies)
```

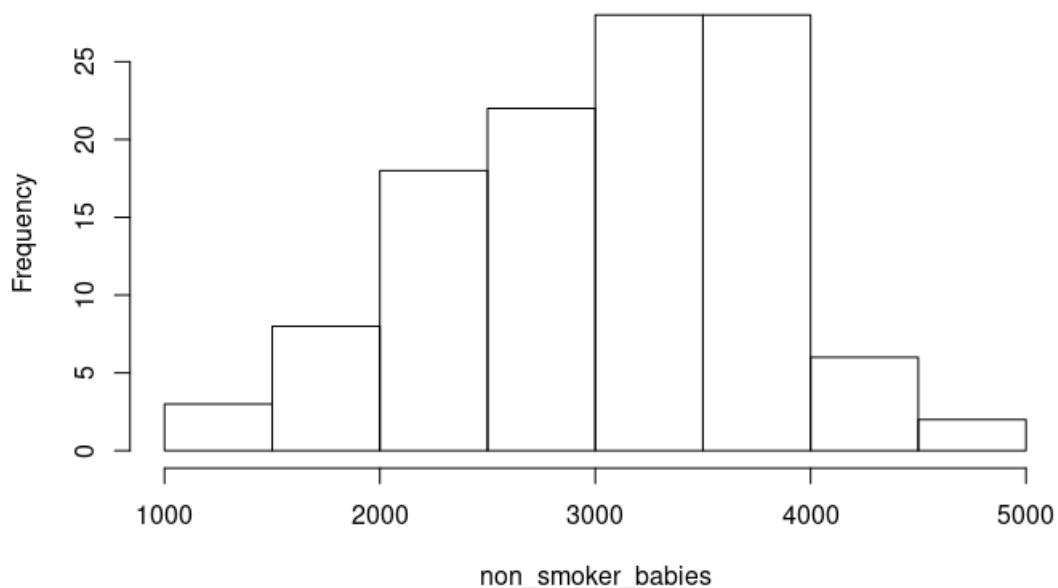
Histogram of smoker_babies



Hide

```
hist(non_smoker_babies)
```

Histogram of non_smoker_babies



Потвърждава се наблюдението, че най-слабите деца са родени от майки пушачи. Не се забелязва обаче по-голяма пропорция за бебета под нормалните килограми.

Съставяме следната хипотеза:

Вероятността майка пушач да роди бебе под 2.5кг е равна на вероятността майка непушач да роди бебе под 2.5кг

Алтернативата гласи, че вероятността за майките пушачи е по-голяма

$H_0: P_x = P_y$

$H_1: P_x > P_y$

Hide

```
smoker_all = length(bwt[smoke == TRUE])
smoker_underweight = length(bwt[smoke == TRUE & bwt < 2500])
non_smoker_all = length(bwt[smoke == FALSE])
non_smoker_underweight = length(bwt[smoke == FALSE & bwt < 2500])
prop.test(c(smoker_underweight, non_smoker_underweight), c(smoker_all, non_smoker_all), alternative = 'greater')
```

2-sample test for equality of proportions with continuity correction

```
data: c(smoker_underweight, non_smoker_underweight) out of c(smoker_all, non_smoker_all)
X-squared = 4.2359, df = 1, p-value = 0.01979
alternative hypothesis: greater
95 percent confidence interval:
 0.02701885 1.00000000
sample estimates:
 prop 1      prop 2 
0.4054054 0.2521739
```

$p\text{-value} < 0.05 \Rightarrow$ приемаме H_1 и можем да твърдим, че за майките пушачи е по-вероятно да родят бебе под 2.5кг