Moodle Tasks

Задача 1
Острата левкимия е една от най-смъртоносните форми на рак. Предишни изследвания показват, че времето на преживяване след първоначалното откриване на левкимия е нормално разпределена случайна величина с математическо очакване 13 месеца и стандартно отклонение 3 месеца. Въвежда се ново лечение, като се очаква то да удължи средното време на живот без да повлияе на дисперсията. Набюдавани са 16 пациента:

10.0, 13.6, 13.2, 11.6, 12.5, 14.2, 14.9, 14.5, 13.4, 8.6, 11.5, 16.0, 14.2, 16.8, 17.9, 17.0

Да се намери оценка за очакването. Да се построи $95\ \%$
доверителен интервал за средното време на живот на болните.

Решение
Доверителния интервал за средното на нормално разпределена случайна величина с известна дисперсия е:

```
> x <- c(10.0, 13.6, 13.2, 11.6, 12.5, 14.2, 14.9, 14.5, 13.4, 8.6, 11.5, 16.0, 14.2, 16.8,
17.9, 17.0)
> z.test <- function(x, sigma, conf.level = 0.95) {
+   n <- length(x)
+   xbar <- mean(x)
+   alpha <- 1 - conf.level
+   zstar <- qnorm(1 - alpha/2)
+   SE <- sigma/sqrt(n)
+   xbar + c(-zstar*SE,zstar*SE)
+ }
> z.test(x, sigma = 3)
[1] 12.27378 15.21372
```

или

```
> library(UsingR)
Warning: package 'UsingR' was built under R version 4.0.3
Loading required package: MASS
Loading required package: HistData
Loading required package: Hmisc
Loading required package: lattice
Loading required package: survival
Loading required package: Formula
Loading required package: ggplot2

Attaching package: 'Hmisc'
The following objects are masked from 'package:base':

    format.pval, units

Attaching package: 'UsingR'
The following object is masked from 'package:survival':

    cancer
```

```
> simple.z.test(x, sigma = 3)
[1] 12.27378 15.21372
```

Задача 2

Генерирайте 20 наблюдения над случайна величина, която е нормално разпределена с очакване 5, и дисперсия 4. Постройте $90\,\%$ процентен доверителен интервал за математическото очакване. Повторете опита 100 пъти. Проверете, в колко от случаите математическото очакване принадлежи на доверителния интервал.

Решение:

```
> x <- rnorm(20, mean = 5, sd = 2)
> alpha <- 0.10
> ci <- function(x, alpha = alpha) {
+   n <- length(x)
+   mean(x) + c(-qnorm(1 - alpha/2) * sd(x)/sqrt(n), qnorm(1 - alpha/2) * sd(x)/sqrt(n))
+ }
> ci(x, 0.10)
[1] 4.239536 5.466073
```

Да го повторим 100 пъти

```
> s <- 0
> for (k in 1:100){
+   x <- rnorm(20, mean = 5, sd = 2)
+   CI <- ci(x, 0.10)
+   if(5 > CI[1] && 5 < CI[2]) s = s + 1
+ }
> s
[1] 86
> s/100
[1] 0.86
```

Задача 3

Постройте $95\,\%$ доверителен интервал за средното време на живот на болните, ако и дисперсията е неизвестна.

Решение:

Доверителния интервал за средното на нормално разпределена случайна величина с неизвестна дисперсия е:

```
> x <- c(10.0, 13.6, 13.2, 11.6, 12.5, 14.2, 14.9, 14.5, 13.4, 8.6, 11.5, 16.0, 14.2, 16.8,
17.9, 17.0)
> tTest <- function(x, conf.level = 0.95) {
+   n <- length(x)
+   xbar <- mean(x)
+   sigma <- sd(x)
+   alpha <- 1 - conf.level
+   zstar <- qt(1 - alpha/2, n - 1)
```

```
+   SE <- sigma / sqrt(n)
+   xbar + c(-zstar*SE, zstar*SE)
+ }
> tTest(x)
[1] 12.39066 15.09684
```

или

```
> t.test(x, conf.level = 0.95)

    One Sample t-test

data:  x
t = 21.65, df = 15, p-value = 9.976e-13
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 12.39066 15.09684
sample estimates:
mean of x
 13.74375
```

Задача 4

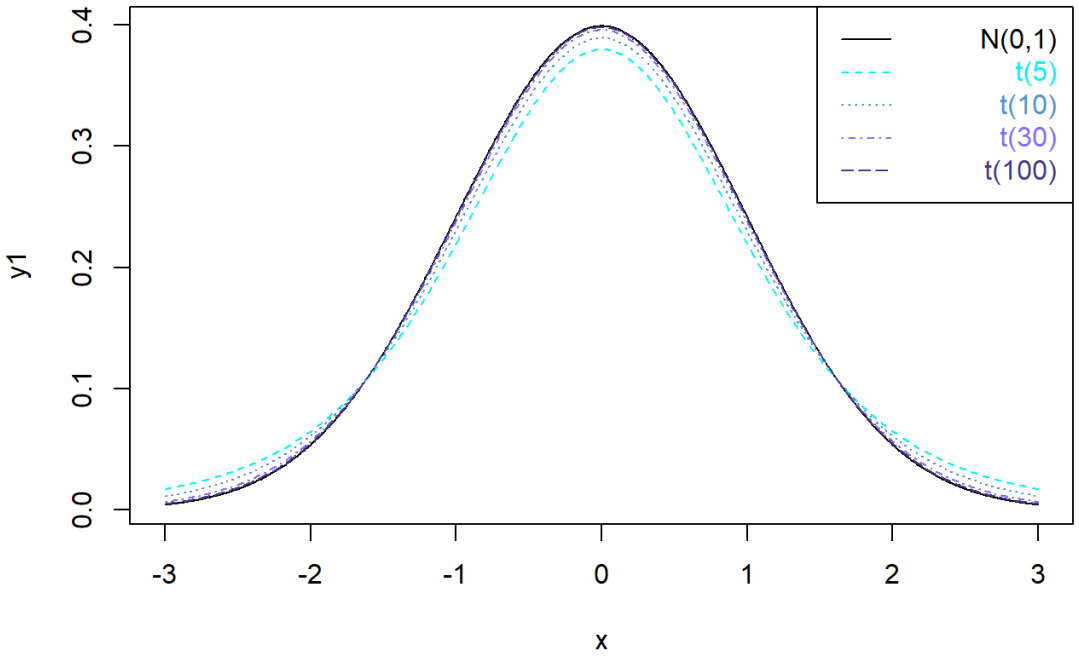
Направете графика с плътността на стандартно нормално разпределение и разпределение на Стюдънт с $5, 10, 30, 100$ степени на свобода.

Решение:

Колкото повече са степените на свобода на $t(n)$, толкова по-близко е до нормалното $N(0,1)$

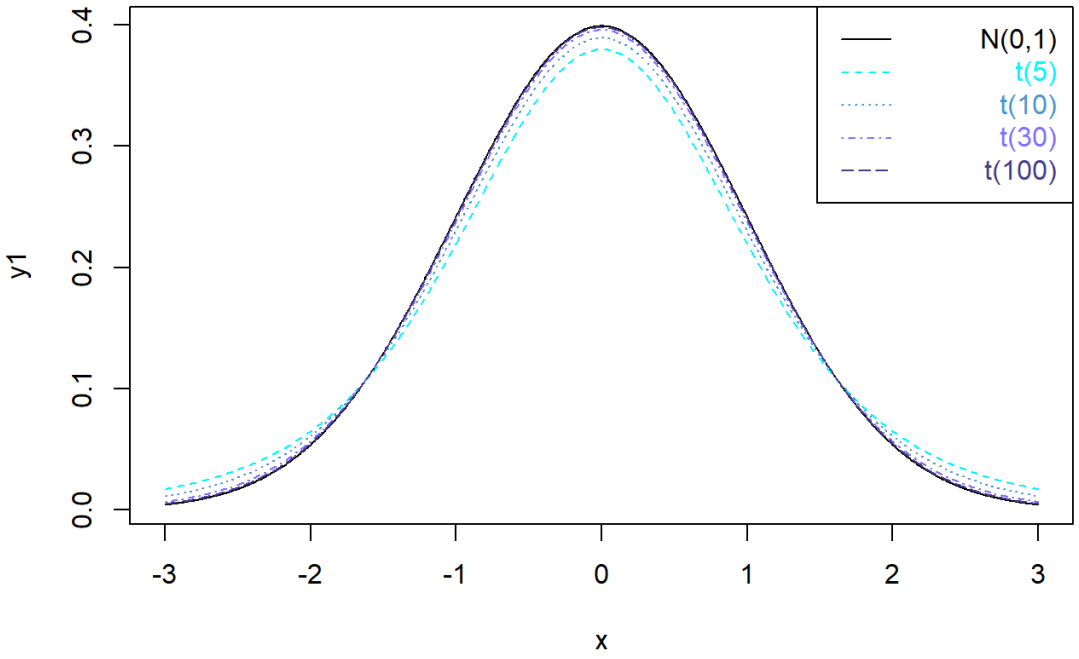
```
> x <- seq(from = -3, to = 3, by = 0.01)
> y1 <- dnorm(x)
> y2 <- dt(x, df = 5)
> y3 <- dt(x, df = 10)
> y4 <- dt(x, df = 30)
> y5 <- dt(x, df = 100)
> plot(x, y1, type = "l")
> lines(x, y2, lty = 2, col = "#00F5FF")
> lines(x, y3, lty = 3, col = "#4F94CD")
> lines(x, y4, lty = 4, col = "#836FFF")
> lines(x, y5, lty = 5, col = "#473C8B")
> temp = legend("topright",
+              legend = c(" "," "," "," "," "),
+              text.width = 1,
+              lty = 1:5,
+              xjust = -1,
+              yjust = 2,
+              col = c("black", "#00F5FF", "#4F94CD", "#836FFF", "#473C8B"))
> text(temp$rect$left + temp$rect$w,
+      temp$text$y,
```

```
+      c("N(0,1)", "t(5)", "t(10)", "t(30)", "t(100)"),
+      pos = 2,
+      col = c("black", "#00F5FF", "#4F94CD", "#836FFF", "#473C8B"))
```



или

```
> n <- c(5, 10, 30, 100)
> col <- c("#00F5FF", "#4F94CD", "#836FFF", "#473C8B")
> plot(x, y1, type = "l")
> for (i in 1:4){
+   y <- dt(x, df = n[i])
+   lines(x, y, lty = i+1, col = col[i])
+ }
> temp <- legend("topright",
+                legend = c(" ", " ", " ", " ", " "),
+                text.width = 1,
+                lty = 1:5,
+                xjust = -1,
+                yjust = 2,
+                col = c("black", col))
> text(temp$rect$left + temp$rect$w,
+      temp$text$y,
+      c("N(0,1)", "t(5)", "t(10)", "t(30)", "t(100)"),
+      pos = 2,
+      col = c("black", col))
```



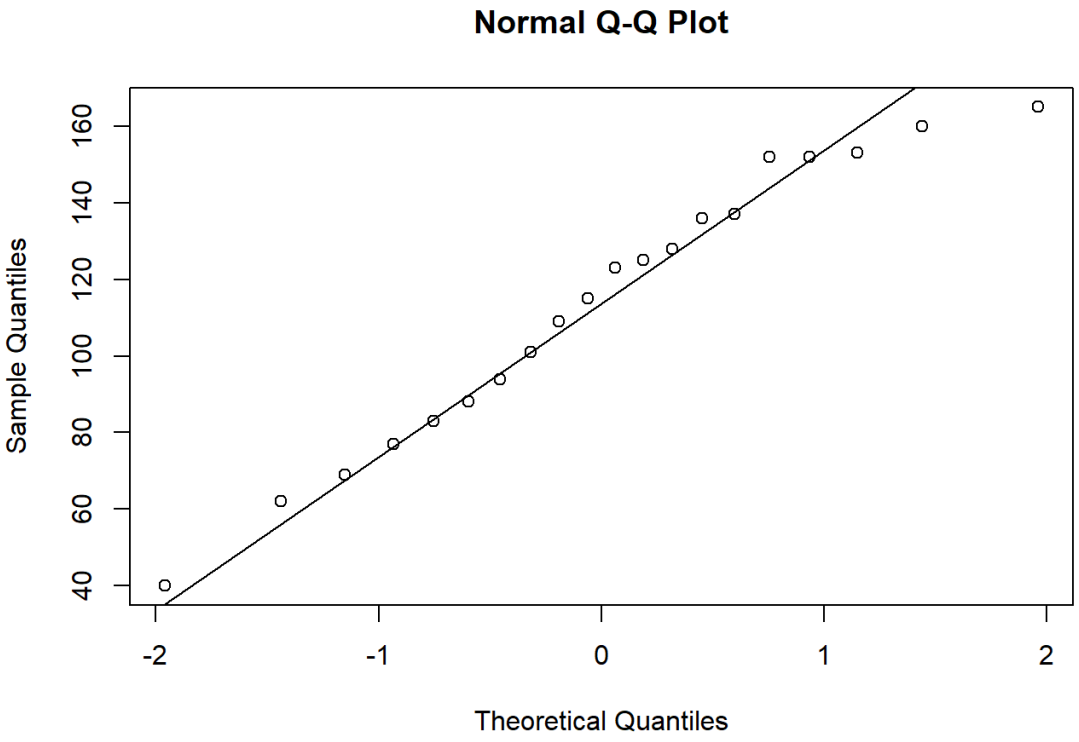
Задача 5

За данните rat от пакета UsingR постройте $96\,\%$ доверителен интервал за очакването.

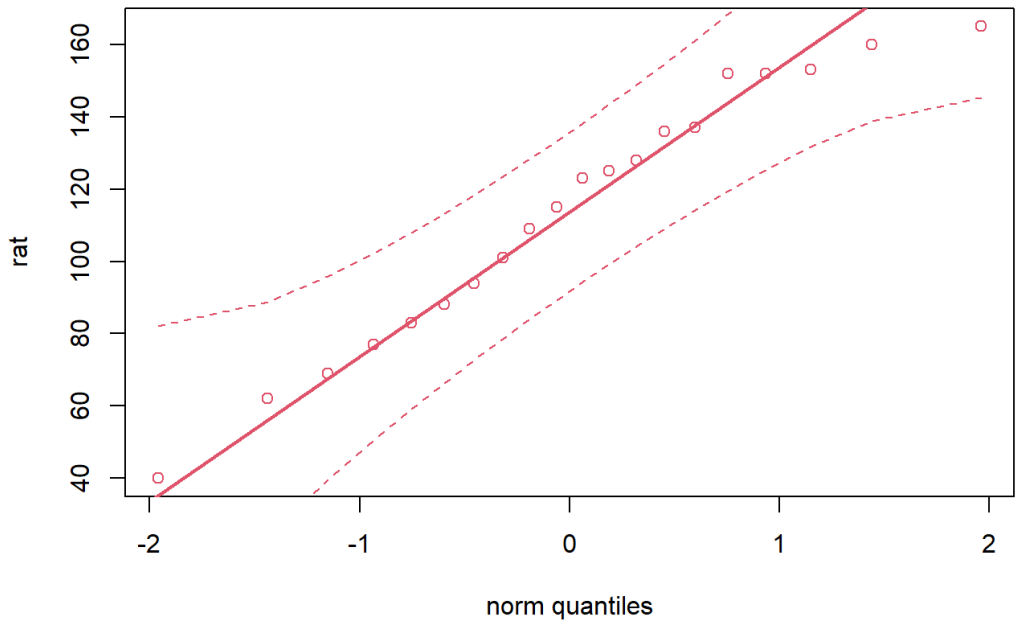
Решение:

Първо трябва да проверим дали данните са номрално разпределени

```
> qqnorm(rat)
> qqline(rat)
```




```
> library(StatDA)
Warning: package 'StatDA' was built under R version 4.0.3
Loading required package: sgeostat
Warning: package 'sgeostat' was built under R version 4.0.3
Registered S3 method overwritten by 'geoR':
  method         from
  plot.variogram sgeostat
> qqplot.das(rat, "norm")
```



```
> shapiro.test(rat)

    Shapiro-Wilk normality test

data:  rat
W = 0.96134, p-value = 0.571
```

Както виждаме от графиките и от $p-value = 0.571 > 0.05 = \alpha$ можем да допуснем, че данните ни са нормално разпредлени.
Данните са ни много малко.

```
> length(rat)
[1] 20
```

Имаме само 20 наблюдения и σ е неизвестно, поради това можем да използваме t.test

```
> mean(rat)
[1] 113.45
> t.test(rat, conf.level = 0.96)

    One Sample t-test

data:  rat
t = 14.176, df = 19, p-value = 1.48e-11
alternative hypothesis: true mean is not equal to 0
96 percent confidence interval:
  95.80624 131.09376
sample estimates:
mean of x
   113.45
```

Задача 6

При провеждане на анкета 87 от 150 анкетирани са отговорили, че са използвали даден продукт. Постройте $92\,\%$ доверителен интервал за броя на хората използвали продукта.

Решение:

```
> n <- 150; k <- 87
> alpha <- 0.08
> phat <- k/n
> SE <- sqrt((phat * (1 - phat)) / n)
> MaxE <- qnorm(1 - alpha/2) * SE
> phat + c(-MaxE, MaxE)
[1] 0.5094493 0.6505507
> ci <- prop.test(87, 150, conf.level = 0.92)
> ci$conf.int[1]*150
[1] 75.77986
> ci$conf.int[2]*150
[1] 97.7171
```

Sources

[1] Monika Petkova's notes on R programming language @ FMI, Sofia University