

Cvičenie 8x – Intervalový odhad strednej hodnoty a rozptylu

Príklady

z teoretickej časti (v inom poradí).

1. (Príklad 2a)

Náhodná veličina má normálne rozdelenie $N(m_i, 22)$ s neznámym parametrom m_i .

Pri 10 pokusoch dostaneme dátový vektor

$$\mathbf{x} = [3 \ 4 \ 7 \ 8 \ 9 \ 11 \ 12 \ 13 \ 16 \ 17]$$

Nájdite obojstranný intervalový odhad strednej hodnoty so spoľahlivosťou 0.9 (alfa = 0.1) a jednostranný odhad so spoľahlivosťou 0.95 (alfa = 0.05).

Riešenie:

Načítame zadané hodnoty a naložíme do známych vzorcov. Kvantilová funkcia (tj. F^{-1}) pre normálne rozdelenie je `qnorm`. ISO je obojstranný interval spoľahlivosti pre alfa=0.9, ISL a ISP sú jednostranné intervaly spoľahlivosti pre alfa = 0.95.

Hodnoty premenných vidíme vpravo hore na záložke Environment, ale ak ich chceme mať presnejšie, treba si ich dať vypísať v Console.

```
x <- c(3,4,7,8,9,11,12,13,16,17)
n <- length(x)
xm <- mean(x)
s2 <- 22
s1 <- sqrt(22)
u <- qnorm(0.95)
d <- u*s1/sqrt(10)
ISO <- xm + c(-1,1)*d
ISL <- c(-Inf,xm+d)
ISP <- c(xm-d,Inf)
```

Na výpočet obojstranného intervalu spoľahlivosti vieme použiť aj pokročilejšie nástroje – funkciu `ZTest`. Tá dokáže veľa vecí, ale dnes sa obmedzíme iba na odhad intervalu spoľahlivosti.

!!! Najprv si treba nainštalovať a načítať knižnicu **DescTools**.

```
ZTest(x,sd_pop = sqrt(22),conf.level = 0.9)$conf.int
```

Vysvetlivky – syntax

`x` je skúmaný vektor dát

`sd_pop` vysvetlí `ZTestu`, že hodnotu smerodajnej odchýlky má považovať za presnú

`conf.level` je spoľahlivosť; ak ne zadáme tento parameter, default hodnota je 0.95

`$conf.int` vytiahne z výsledkov iba interval spoľahlivosti, ušetrí nás zatiaľ od iných výsledkov

Jednostranné odhady – default spoľahlivosť 0.95, nastavenie cez parametre `g`, `l` (*greater/lesser*)

```
ZTest(x,sd_pop = sqrt(22),alternative = "g")$conf.int
```

```
ZTest(x,sd_pop = sqrt(22),alternative = "l")$conf.int
```

2. (Príklad 1a)

Náhodná veličina má normálne rozdelenie $N(\mu, \sigma^2)$ s neznámymi parametrami. Pri 10 pokusoch dostaneme dátový vektor x (ten istý ako vyššie spomenutý).

Nájdite intervalový odhad strednej hodnoty so spoľahlivosťou 0.9.

Riešenie:

Základný postup je podobný ako v predošlom príklade, musíme však použiť Studentovo rozdelenie, ktorého kvantilová funkcia je `qt`.

```
s1 <- sd(x)
u <- qt(0.95,n-1)
d <- u*s1/sqrt(n)
ISO <- xm + c(-1,1)*d
ISL <- c(-Inf,xm+d)
ISP <- c(xm-d,Inf)
```

Na výpočet obojstranného intervalu spoľahlivosti opäť vieme použiť pokročilejšie nástroje – funkciu `t.test`, pri ktorej sa takisto obmedzíme iba na odhad intervalu spoľahlivosti. Funkcia `t.test` je súčasťou balíka `stats`, ktorý by mal byť pripravený na použitie bez inštalovania a načítania.

```
t.test(x,conf.level = 0.9)$conf.int
```

Vysvetlivky

`x` je skúmaný vektor dát

`conf.level` je spoľahlivosť; ak nezádame tento parameter, default hodnota je 0.95

`$conf.int` vytiahne z výsledkov iba interval spoľahlivosti

Jednostranné odhady – default spoľahlivosť 0.95, nastavenie cez parametre `g`, `l` (*greater/lesser*)

```
t.test(x,alternative = "l")$conf.int
```

```
t.test(x,alternative = "g")$conf.int
```

3. (Príklad 1b)

Počítajme intervalový odhad rozptylu pre zadanie predošlého príkladu, tj. pri neznámom mi a so spoľahlivosťou 0.9 (obojustranne), 0.95 (jednostranne).

Riešenie:

Začneme klasickým postupom. Kvantilová funkcia pre chí-kvadrát je `qchisq`.

```
s2 <- var(x)
u2 <- qchisq(0.95,n-1)
u1 <- qchisq(0.05,n-1)
isox <- s2*(n-1)/c(u2,u1)
```

Jednostranné odhady:

```
c(0,isox[2])
c(isox[1],Inf)
```

Na výpočet je možné použiť pokročilejšie nástroje – funkciu `varTest` (nemýliť si s `var.test`).

!! Najprv je potrebné si nainštalovať a načítať knižnicu **EnvStats** (nemýliť si s `envstat`).

Obojustranný odhad:

```
varTest(x, conf.level = 0.9)$conf.int
```

Jednostranné odhady:

```
varTest(x, alternative = "g", conf.level = 0.95)$conf.int
varTest(x, alternative = "l", conf.level = 0.95)$conf.int
```

4. (Príklad 2b)

Počítajme intervalový odhad rozptylu pre zadanie s $mi=10$ presným(!) a spoľahlivosťou 0.9.

Riešenie:

Začneme klasickým postupom.

```
u2 <- qchisq(0.95,n)
u1 <- qchisq(0.05,n)
s2*n/c(u2,u1)
```

Cez funkciu `varTest` – ??

Ďalšie úlohy

1. Náhodná veličina má normálne rozdelenie $N(\mu, \sigma^2)$ s neznámymi parametrami. V sérii experimentov sa prejavila vektorom dát, ktoré predstavujú stĺpec Age v tabulke ciska.csv (pracovalo sa s ňou minulý týždeň).

Pomôcka

```
library(readr)
ciska <- read_csv("ciska.csv")
age <- ciska$Age
```

Nájdite s využitím testovacích funkcií (predstavených na tomto cvičení) intervalové odhady strednej hodnoty a rozptylu so spoľahlivosťou 0.95 a 0.99.

2. Základný postup riešenia príkladov 1-4 zostručnite (zredukujte počet priradení, v rámci možnosti všetko dosadzujte rovno do výsledného vzorca na výpočet IS). Aplikujte postup na iné vektory dát a iné úrovne spoľahlivosti.

3. Náhodná veličina má normálne rozdelenie $N(\mu, \sigma^2)$ s neznámymi parametrami. Pri 1000 pokusoch dostaneme dátový vektor, z ktorého určíme bodové odhady

$$m = 11, \quad s^2 = 49$$

Nájdite (klasickou cestou) intervalové odhady strednej hodnoty a rozptylu so spoľahlivosťou 0.95 a 0.99.