Wnioskowanie statystyczne

№1. 26 lutego 2018

Hipoteza o wartości średniej

Statystyka testowa:

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim \mathrm{N}(0,1).$$

Przedział ufności:

$$\left(\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

Sposób klasyczny, przedział ufności, wartość_p.

```
dane <- read.csv("ww0101.csv")
head(dane)
nrow(dane)
summary(dane)

dane <- as.matrix(dane)
mean(dane)
var(dane)
sd(dane)
se <- function(x) sqrt(var(x)/length(x))
se(dane)

# Przedział ufności
# Tutaj też metoda klasyczna
(half <- qnorm(0.975)*sd(dane)/sqrt(nrow(dane)))
(left <- mean(dane) - half)
(right <- mean(dane) + half)

# Metoda klasyczna
mu0 <- 17.5
z <- (mean(dane) - mu0) / ( sd(dane)/sqrt(length(dane)) )
z
qnorm(0.975)

# Wartość_p
2* pnorm(z)
```

Hipoteza o wartości średniej – mała próbka

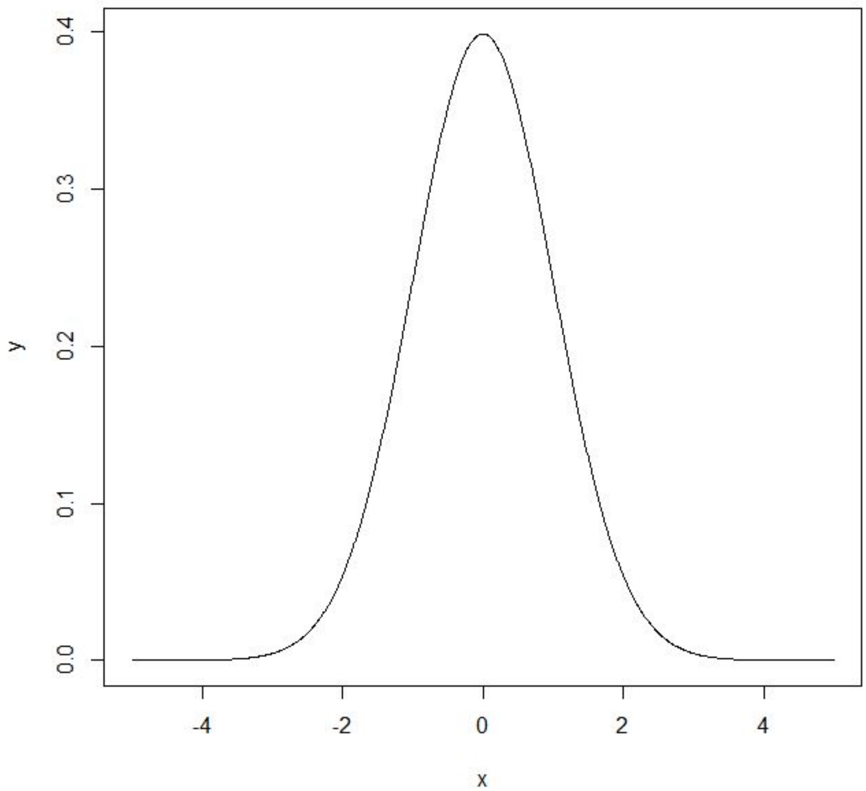
Statystyka testowa:

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim \mathrm{t}(n-1).$$

Przedział ufności:

$$\left(\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

Sposób klasyczny, przedział ufności, wartość_p.




```
x <- seq(-5, 5, 0.05)
y <- dnorm(x)
plot(y~x, type="l")
# Wykres powyżej

# Poniżej, polecenia dla
#     trzech rozkładów t-Studenta
#     na wspólnym wykresie
x <- seq(0, 3, 0.01)
y2 <- dt(x, 2)
y5 <- dt(x, 5)
y10 <- dt(x, 10)
df <- data.frame(x,y2,y5,y10)
require(ggplot2)
g <- ggplot(df, aes(x))
g <- g + geom_line(aes(y=y2), colour="red")
g <- g + geom_line(aes(y=y5), colour="green")
g <- g + geom_line(aes(y=y10), colour="blue")
g
```

Hipoteza o wskaźniku struktury

Statystyka testowa:

$$Z = \frac{\hat{x} - p_0}{\sqrt{p_0 q_0 / n}} \sim \mathrm{N}(0,1), \quad (np_0 > 5,\ nq_0 > 5).$$

Przedział ufności:

$$\left(\hat{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \right)$$

Plik	**Opis**
ww0101.csv	tygodniowe wydatki na słodycze
ww0102.csv	czas życia sprzętu elektr.

Lista zadań №1, 28 lutego 2018

1. Wykonać test dla hipotezy $H_0 : \mu_0 = 18$ $(\alpha = 0.05,\ \mu \neq 18)$.
2. Wyznaczyć wartość_p dla hipotezy $H_0 : \mu_0 = 17.5$, $(\mu \neq 18)$.
3. Dla danych z pliku w0102.csv
   - (a) wyznaczyć przedział ufności dla wartości średniej,
   - (b) obliczyć wartość_p.
4. Na przedziale $(0; 3)$ zilustrować gęstości rozkładu $\mathrm{N}(0,1), t(3), t(25)$.
5. W ankiecie przeprowadzonej wśród 350 osób – 110 osób wskazało, że przynajmniej raz w tygodniu zjada kebab.
   - (a) Wyznaczyć przedział ufności dla wskaźnika (procentu) osób znajomych z kebabem.
   - (b) Wyznaczyć wartość_p dla wskaźnika struktury (spożycia).
   - (c) Niech $\alpha = 0.05$. Czy prawdziwą jest hipoteza, że $p_0 = \frac{1}{3}$?

Witold Karczewski