Wnioskowanie statystyczne

№1. 19 marca 2018

Lista zadań №1, 28 lutego 2018

1. Wykonać test dla hipotezy $H_0 : \mu_0 = 18$ ($\alpha = 0.05,\ \mu \neq 18$).

   Statystyka testowa:

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

```
> dane <- read.csv("ww0101.csv")
> head(dane)
  Wydatki
1      15
2       7
3      29
...
> nrow(dane)
[1] 300
> summary(dane)
    Wydatki
 Min.   : 0.00    1st Qu.: 8.00
 Median :17.00    Mean   :16.09
 3rd Qu.:24.00    Max.   :39.00
> dane <- as.matrix(dane)
> mu0 <- 18.0
> z <- (mean(dane) - mu0) / ( sd(dane)/sqrt(length(dane)) )
> z
[1] -3.633029
> qnorm(0.975)
[1] 1.959964
```

2. Wyznaczyć wartość_p dla hipotezy $H_0 : \mu_0 = 2700$ ($\mu \neq 2700$).

   Statystyka testowa:

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

```
> dane <- read.csv("ww0102.csv")
> head(dane)
  Czas
1 3150
2 2669
3 2860
...
> nrow(dane)
[1] 15
> summary(dane)
      Czas
 Min.   :2364    1st Qu.:2622
 Median :2860    Mean   :2837
 3rd Qu.:3078    Max.   :3161
> dane <- as.matrix(dane)
```

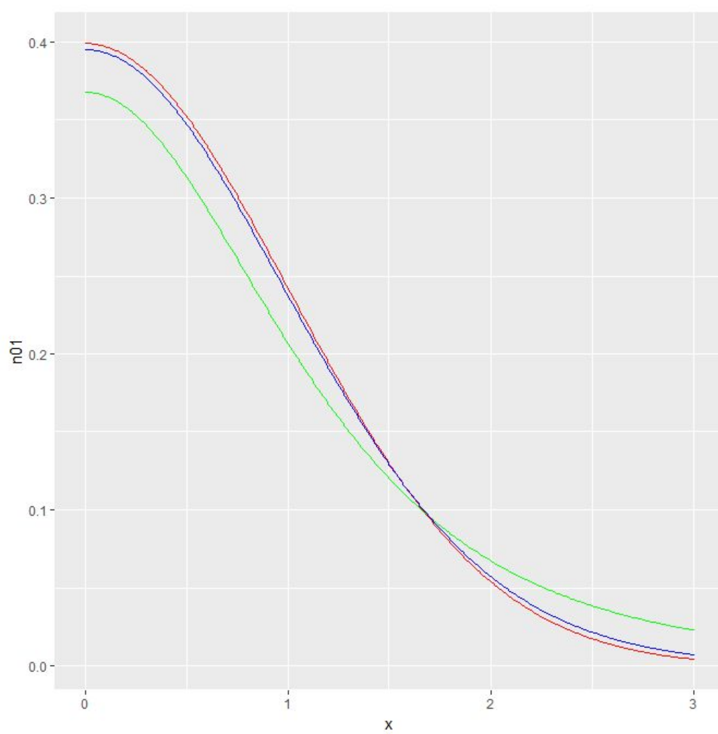
```
> mu0 <- 2700
> z <- (mean(dane) - mu0) / ( sd(dane)/sqrt(length(dane)) )
> z
[1] 1.992946
> qnorm(0.975)
[1] 1.959964
> qt(0.975,15)
[1] 2.13145
> # Wartość_p
> 2* (1-pnorm(z))
[1] 0.04626739
> 2* (1-pt(z,15))
[1] 0.06478532
```

3. Dla danych z pliku w0102.csv wyznaczyć przedział ufności dla wartości średniej,

Przedział ufności:

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

```
> # Przedział ufności
> (half <- qt(0.975,15)*sd(dane)/sqrt(nrow(dane)))
[1] 146.4498
> (left <- mean(dane) - half)
[1] 2690.484
> (right <- mean(dane) + half)
[1] 2983.383
```



4. Na przedziale $(0;3)$ zilustrować gęstości rozkładu $\mathrm{N}(0,1), t(3), t(25)$.

```
# Poniżej, polecenia:
#     trzy rozkłady 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
```

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 ($p_0 = \frac{1}{3}$).
   (c) Niech $\alpha = 0.05$. Czy prawdziwą jest hipoteza, że $p_0 = \frac{1}{3}$?

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:

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

```
> # Przedział ufności
> (p <- 110/350)                 [1] 0.3142857
> (q <- 1-p)   [1] 0.6857143
> (half <- qnorm(0.975)*sqrt(p*q/nrow(dane)))   [1] 0.2349289
> (left <- p - half)             [1] 0.07935685
> (right <- p + half)            [1] 0.5492146

> # Wartość_p
> (z <- (p-1/3)/sqrt(p*q/nrow(dane)))
[1] -0.1589104
> pnorm(z)
[1] 0.4368697
> # Nie ma podstaw do odrzucenia hipotezy.
```

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

Witold Karczewski

Wnioskowanie statystyczne

Lista zadań nr 4, 19 marca 2018.

1. Dane zawierają dzienną liczbę ogłoszeń w gazetach A, B oraz C. Testujemy hipotezę: **średnia liczba ogłoszeń w trzech gazetach jest taka sama**.

2. Dane: dzienna liczba zgłoszeń w czterech działach (A, B, C, D) firmy. Testowana hipoteza: **liczba zgłoszeń jest średnio taka sama we wszystkich oddziałach**.

3. W czterech punktach obserwacyjnych notujemy liczbę wstrząsów. Testujemy hipotezę: **punkty położone są w obrębie tej samej struktury geologicznej**.

4. Dane: wytrzymałość trzech rodzajów opon pochodzących od pięciu producentów. Testowane hipotezy: **średnia wytrzymałość opony jest niezależna od producenta** oraz **różne rodzaje opon mają różną wytrzymałość**.

5. Dane: czynnik A – normalny (1) lub podwyższony poziom cukru (2); czynnik B – waga normalna (1) lub powyżej normy (2). Mierzymy ciśnienie. **czy istnieje związek między poziomem cukru a ciśnieniem**, **czy istnieje związek między wagą a ciśnieniem** oraz **czy istnieje interakcja wagi z poziomem cukru**?

6. Czynnik A to środek reklamy, czynnik B to główny temat reklamy, dane to procent pozytywnych reakcji na reklamę. **Czy istnieje związek między nośnikiem reklamy a skutecznością reklamy**, **który typ reklamy jest najskuteczniejszy** oraz **czy istnieje związek między rodzajem reklamy a jej nośnikiem**?

Witold Karczewski