Testiranje hipotez

A. Blejec

16. januar 2019

Kazalo

1 Test hipotez z uporabo normalne porazdelitve

Ohlapno rečeno so vzorčne porazdelitve ocen in testnih izrazov normalne, če so vzorci dovolj veliki.

1.1 Test hipotez o povprečni vrednosti

Če poznamo σ ali pa je vzorec dovolj velik.

1. Hipoteze

$$H_0: \quad \mu = \mu_0$$

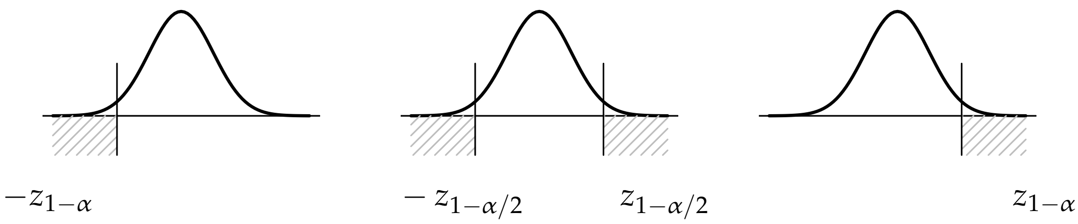
$$H_1: \quad \mu < \mu_0 \qquad\qquad H_1: \quad \mu \neq \mu_0 \qquad\qquad H_1: \quad \mu > \mu_0$$

2. Cenilka in testni izraz

$$\overline{x} = \frac{\sum x_i}{n}$$

$$Z = \frac{\overline{x} - \mu_0}{\sigma}\sqrt{n} \sim N(0,1)$$

3. Vzorčna porazdelitev in kritične vrednosti



4. Vzorec in račun testne vrednosti z.

5. H_0 zavrnem, če je testna vrednost z v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Povprečje v populaciji je statistično značilno [manjše | različno | večje] od μ_0 (s tveganjem $p < \alpha$).

1.2 Test hipotez o strukturnem deležu

1. Hipoteze

$$H_0 : \quad \pi = \pi_0$$

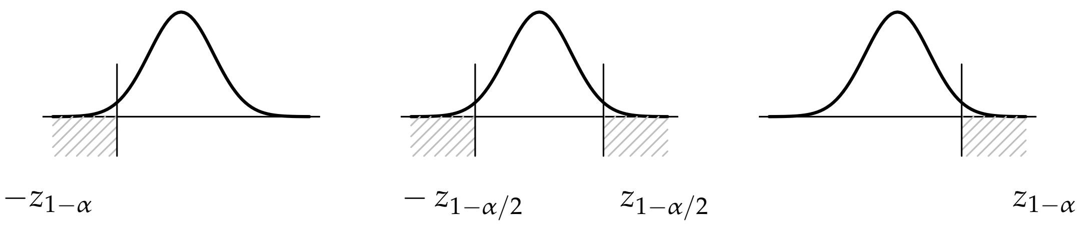
$$H_1 : \quad \pi < \pi_0 \qquad\qquad H_1 : \quad \pi \neq \pi_0 \qquad\qquad H_1 : \quad \pi > \pi_0$$

2. Testni izraz za k izidov izmed n opazovanih

$$p = k/n$$

$$Z = \frac{p - \pi_0}{\sqrt{p \cdot (1-p)}} \sqrt{n} \sim N(0,1)$$

3. Vzorčna porazdelitev in kritične vrednosti



4. Vzorec in račun testne vrednosti z.

5. H_0 zavrnem, če je testna vrednost z v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Strukturni delež v populaciji je statistično značilno [manjši | različen | večji] od π_0 (s tveganjem $p < \alpha$).

1.3 Test hipotez o razliki povprečij

Če poznamo standardne odklone ali pa sta vzorca dovolj velika.

1. Hipoteze

$$\begin{aligned} H_0: &\quad \mu_1 = \mu_2 + \Delta \\ H_0: &\quad \mu_1 - \mu_2 = \Delta \end{aligned}$$

$$H_1: \quad \mu_1 - \mu_2 < \Delta \qquad H_1: \quad \mu_1 - \mu_2 \neq \Delta \qquad H_1: \quad \mu_1 - \mu_2 > \Delta$$

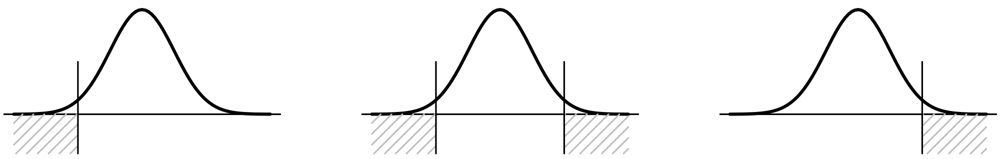
   Poseben primer: $\Delta = 0$

2. Cenilka in testni izraz

$$\overline{x}_1 - \overline{x}_2$$

$$Z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1)$$

3. Vzorčna porazdelitev in kritične vrednosti



$-z_{1-\alpha}$ $\quad$ $-z_{1-\alpha/2}$ $\quad$ $z_{1-\alpha/2}$ $\quad$ $z_{1-\alpha}$

4. Vzorec in račun testne vrednosti z.

5. H_0 zavrnem, če je testna vrednost z v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Razlika populacijskih povprečij je statistično značilno [manjša | različna | večja] od Δ (s tveganjem $p < \alpha$).

1.4 Test hipotez o razliki dveh strukturnih deležev

Če poznamo standardne odklone ali pa sta vzorca dovolj velika.

1. Hipoteze

$$H_0: \quad \pi_1 = \pi_2 + \Delta$$
$$H_0: \quad \pi_1 - \pi_2 = \Delta$$

$$H_1: \quad \pi_1 - \pi_2 < \Delta \qquad H_1: \quad \pi_1 - \pi_2 \neq \Delta \qquad H_1: \quad \pi_1 - \pi_2 > \Delta$$

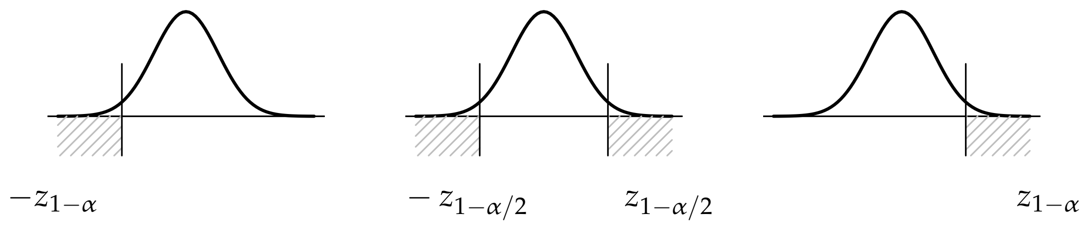
   Poseben primer: $\Delta = 0$

2. Cenilka in testni izraz

$$p_1 = k_1/n_1 \qquad\qquad p_2 = k_2/n_2$$

$$Z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}} \sim N(0,1)$$

3. Vzorčna porazdelitev in kritične vrednosti



4. Vzorec in račun testne vrednosti z.

5. H_0 zavrnem, če je testna vrednost z v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Razlika deležev v populacijah je statistično značilno [manjša | različna | večja] od Δ (s tveganjem $p < \alpha$).

2 Test hipotez z uporabo drugih porazdelitev

2.1 Studentov test hipotez o povprečni vrednosti

Če ne poznamo σ, vzorec majhen.

1. Hipoteze

$$H_0 : \quad \mu = \mu_0$$

$$H_1 : \quad \mu < \mu_0 \qquad\qquad H_1 : \quad \mu \neq \mu_0 \qquad\qquad H_1 : \quad \mu > \mu_0$$

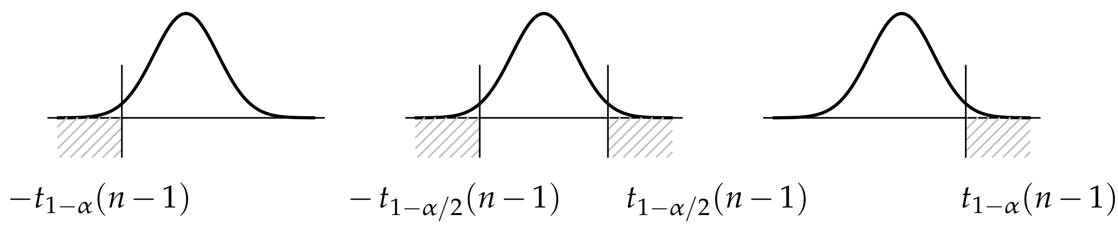
2. Cenilka in testni izraz

$$\overline{x} = \frac{\sum x_i}{n}$$

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

$$t = \frac{\overline{x} - \mu_0}{s}\sqrt{n} \sim t(n-1)$$

3. Vzorčna porazdelitev in kritične vrednosti



4. Vzorec in račun testne vrednosti t.

5. H_0 zavrnem, če je testna vrednost t v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Povprečje v populaciji je statistično značilno [manjše | različno | večje] od μ_0 (s tveganjem $p < \alpha$).

2.2 Studentov test hipotez o razliki povprečij (Studentov t-test)

Privzeto: varianci σ_1^2 in σ_2^2 sta neznani, a enaki , vzorca pa sta majhna. Predhodno je morda potreben test hipoteze o enakosti varianc, varianci ne smeta biti statistično značilno različni.

1. Hipoteze

$$H_0: \quad \mu_1 = \mu_2 + \Delta$$
$$H_0: \quad \mu_1 - \mu_2 = \Delta$$

$$H_1: \quad \mu_1 - \mu_2 < \Delta \qquad H_1: \quad \mu_1 - \mu_2 \neq \Delta \qquad H_1: \quad \mu_1 - \mu_2 > \Delta$$

   Poseben primer: $\Delta = 0$

2. Cenilka in testni izraz

   Velikosti vzorcev sta n_1 in n_2

$$\overline{x}_1 - \overline{x}_2$$

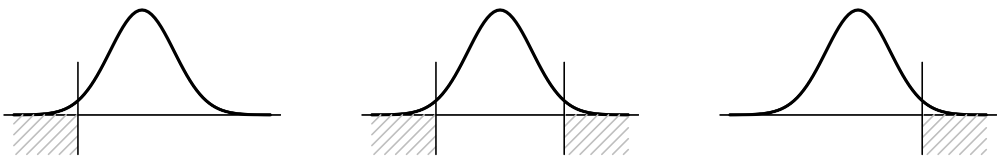
$$s_1^2 = \frac{\sum (x_{1,i} - \overline{x}_1)^2}{n_1 - 1} \qquad s_2^2 = \frac{\sum (x_{2,i} - \overline{x}_2)^2}{n_2 - 1}$$

   Združena varianca (pooled variance)

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{s} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \sim t(n_1 + n_2 - 2)$$

3. Vzorčna porazdelitev in kritične vrednosti (stopinje prostosti: $df = n_1 + n_2 - 2$)



$-t_{1-\alpha}(df)$ $\qquad - t_{1-\alpha/2}(df) \quad t_{1-\alpha/2}(df)$ $\qquad t_{1-\alpha}(df)$

4. Vzorec in račun testne vrednosti t.

5. H_0 zavrnem, če je testna vrednost t v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Razlika populacijskih povprečij je statistično značilno [manjša | različna | večja] od Δ (s tveganjem $p < \alpha$).

2.3 Studentov test hipotez o vrednosti standardnega odklona

Če ne poznamo σ, vzorec majhen.

1. Hipoteze

$$H_0: \quad \sigma = \sigma_0$$

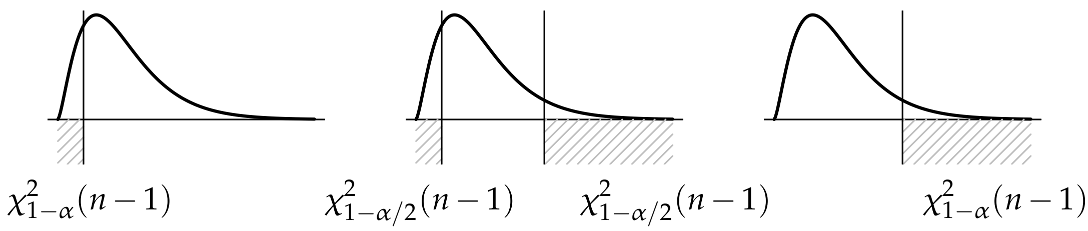
$$H_1: \quad \sigma < \sigma_0 \qquad\qquad H_1: \quad \sigma \neq \sigma_0 \qquad\qquad H_1: \quad \sigma > \sigma_0$$

2. Cenilka in testni izraz

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$$

$$c = \frac{(n-1)s}{\sigma_0} \sim \chi^2(n-1)$$

3. Vzorčna porazdelitev in kritične vrednosti



4. Vzorec in račun testne vrednosti t.

5. H_0 zavrnem, če je testna vrednost t v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Standardni odklon v populaciji je statistično značilno [manjši | različen | večji] od σ_0 (s tveganjem $p < \alpha$).

2.4 Test hipotez o enakosti varianc

1. Hipoteze

$$H_0: \quad \sigma_1^2 = \sigma_2^2$$
$$H_0: \quad \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_1: \quad \frac{\sigma_1^2}{\sigma_2^2} > 1$$

2. Cenilka in testni izraz

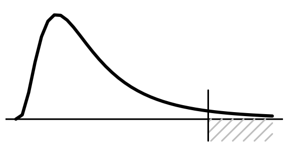
   Velikosti vzorcev sta n_1 in n_2, $s_1^2 > s_2^2$

$$s_1^2 = \frac{\sum (x_{1,i} - \overline{x}_1)^2}{n_1 - 1}$$

$$s_2^2 = \frac{\sum (x_{2,i} - \overline{x}_2)^2}{n_2 - 1}$$

$$F = \frac{s_1^2}{s_2^2} \sim F(n_1 - 1, n_2 - 1)$$

3. Vzorčna porazdelitev in kritične vrednosti
   (stopinje prostosti: števec: $n_1 - 1$, imenovalec: $n_2 - 1$)



$F_{1-\alpha}(n_1 - 1, n_2 - 1)$

4. Vzorec in račun testne vrednosti F.

5. H_0 zavrnem, če je testna vrednost F v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Varianci populacij sta statistično značilno različni (s tveganjem $p < \alpha$).

2.5 Test hipotez o povprečni razliki

Podatki izmerjeni na istih enotah (npr. pred in po postopku, katerega vpliv proučujemo).

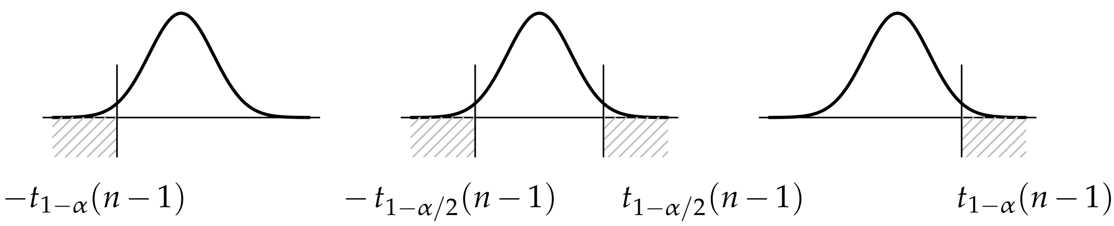
$$\begin{aligned} X_1 &= \mu + \varepsilon & &\sim N(\mu, \sigma^2) \\ X_2 &= \mu + \delta + \varepsilon & &\sim N(\mu + \delta, \sigma^2) \end{aligned}$$

Zanima nas velikost spremembe $\delta = X_2 - X_1$.
Če za podatke vzamemo razlike $d_i = x_{2,i} - x_{1,i}$, lahko uporabimo test za povprečno vrednost.

1. Hipoteze
$$H_0 : \delta = \delta_0$$
Poseben primer: $\delta_0 = 0$.
$$H_1 : \delta < \delta_0 \qquad H_1 : \delta \neq \delta_0 \qquad H_1 : \delta > \delta_0$$

2. Cenilka in testni izraz
$$d_i = x_{2,i} - x_{1,i}$$
$$\overline{d} = \frac{\sum d_i}{n}$$
$$s^2 = \frac{\sum (d_i - \overline{d})^2}{n-1}$$
$$t = \frac{\overline{d} - \delta_0}{s}\sqrt{n} \sim t(n-1)$$

3. Vzorčna porazdelitev in kritične vrednosti



4. Vzorec in račun testne vrednosti t.

5. H_0 zavrnem, če je testna vrednost t v kritičnem območju (šrafirano), ter sprejmemo H_1.

   Popvrečna razlika je statistično značilno [manjša | različna | večja] od δ_0 (s tveganjem $p < \alpha$).

SessionInfo

Windows 10 x64 (build 17134)

- R version 3.3.1 (2016-06-21), `x86_64-w64-mingw32`
- Locale: `LC_COLLATE=Slovenian_Slovenia.1250`, `LC_CTYPE=Slovenian_Slovenia.1250`, `LC_MONETARY=Slovenian_Slovenia.1250`, `LC_NUMERIC=C`, `LC_TIME=Slovenian_Slovenia.1250`
- Base packages: base, datasets, graphics, grDevices, stats, utils
- Other packages: knitr 1.20
- Loaded via a namespace (and not attached): evaluate 0.11, magrittr 1.5, methods 3.3.1, stringi 1.1.7, stringr 1.3.1, tools 3.3.1

Project path: [link]
`D:/`
`_Y/`
`R/`
`Stat-V/`

Main file : `../doc/Test-hipotez.Rnw`
Project file: [link]

View as vignette

Project files can be viewed by pasting this code to R console:

```
projectName <-"Stat-V";  mainFile <-"Test-hipotez"
 commandArgs()
 library(tkWidgets)
 openPDF(file.path(dirname(getwd()),"doc",
 paste(mainFile,"PDF",sep=".")))
 viewVignette("viewVignette", projectName,
 file.path("../doc",paste(mainFile,"Rnw",sep=".")))
```