

Testovanie štatistických hypotéz

Príklad Meral sa percentuálny obsah cínu vo vzorkách rudy. Výsledky sú v tabuľke, pričom x_i je triedny znak:

x_i	30	35	40	45	50	55	60	65	70	75
n_i	1	3	4	10	15	20	11	5	3	2

Predpokladáme, že obsah cínu má normálne rozdelenie s disperziou $\sigma^2 = 85$. Na hladine významnosti $\alpha = 0,05$ testujeme hypotézu $H_0 : \mu = 52$ proti $H_1 : \mu \neq 52$.

Riešenie - spôsob A:

• Použijeme Y-test zhody strednej hodnoty μ so známou konštantou μ_0 , pričom σ poznáme. Budeme testovať nulovú hypotézu $H_0 : \mu = 52$ proti alternatívnej hypotéze $H_1 : \mu \neq 52$, kde

$\mu_0 = 52$. Použijeme testovaciu charakteristiku $Y = \frac{\bar{x} - \mu_0}{\sigma} \cdot \sqrt{n}$.

• $\alpha = 0,05$.

• Výpočet testovacej charakteristiky ilustrujeme použitím MATLABu, pričom výsledok uložíme do premennej Y:

```
>> x=[30,35*ones(1,3),40*ones(1,4),45*ones(1,10),50*ones(1,15),...  
55*ones(1,20),60*ones(1,11),65*ones(1,5),70*ones(1,3),75*ones(1,2)];  
n=length(x);Y=(mean(x)-52)*sqrt(n)/sqrt(85)
```

```
Y =  
1.1600
```

• Hypotézu H_0 nezamietame, keď bude platiť $Y \notin K_\alpha$, pričom pre kritickú oblasť v tomto prípade platí:

$$K_\alpha = \left(-\infty; -y_{1-\frac{\alpha}{2}} \right) \cup \left(y_{1-\frac{\alpha}{2}}; \infty \right).$$

Použijeme kvantil normovaného normálneho rozdelenia (napr. z tabuliek):

$$c = y_{1-\frac{\alpha}{2}} = y_{0,975} = 1,9600.$$

Pomocou MATLABu dostaneme:

```
>> alfa=.05;c=norminv(1-alfa/2)
```

```
c =  
1.9600
```

Teda kritická oblasť je $K_\alpha = (-\infty; -1,96) \cup (1,96; \infty)$.

• Pretože $Y \notin K_\alpha$, hypotézu H_0 nezamietame.

Riešenie - spôsob B:

Použijeme

```
[h,sig,ci,zval] = ztest(x,m,sig,alpha,tail)
```

```
>> sigma=sqrt(85)
```

sigma =
9.2195

>> [h,significance,ci,zval] = ztest(x,52,sigma,0.05,0) % alebo [h,significance,ci,zval] =
ztest(x,52,sigma,0.05,'both')

h =
0

significance =
0.2460 % significance is 0.2460, which means that by chance we would
have observed values of t more extreme than the one in this
example in only 2460 of 10,000 similar experiments!

ci =
51.1427 55.3438 % 95 %-ný obojstranný interval spoľahlivosti pre strednú hodnotu
zval =
1.1600 % hodnota testovacej charakteristiky (v A spôsobe je to Y)

Záver: Pretože $h = 0$ hypotézu H_0 nezamietame.

Príklad Bola meraná koncentrácia istej chemickej látky v roztoku [%], s týmito výsledkami:
30, 32, 34, 40, 36, 37, 36, 38, 35, 42. Na hladine významnosti $\alpha = 0,05$ testujme hypotézu
 $H_0 : \mu = 38$ proti $H_1 : \mu < 38$ za predpokladu, že výberový súbor má normálne rozdelenie.

Riešenie - spôsob A:

• Použijeme t-test zhody strednej hodnoty μ so známou konštantou μ_0 , pričom σ nepoznáme.
Budeme testovať nulovú hypotézu $H_0 : \mu = 38$ proti alternatívnej hypotéze $H_1 : \mu < 38$, kde

$\mu_0 = 38$. Použijeme testovaciu charakteristiku $t = \frac{\bar{x} - \mu_0}{s^*} \cdot \sqrt{n}$.

• $\alpha = 0,05$.

• Výpočet testovacej charakteristiky ilustrujeme použitím MATLABu, pričom výsledok uložíme do premennej t:

>> x=[30,32,34,40,36,37,36,38,35,42];n=length(x);t=(mean(x)-38)*sqrt(n)/std(x)

t =
-1.7770

• Hypotézu H_0 nezamietame, keď bude platiť $t \notin K_\alpha$, pričom pre kritickú oblasť v tomto
prípade platí: $K_\alpha = (-\infty; -t_{1-\alpha, n-1})$.

Použijeme kvantil Studentovho t-rozdelenia: $c = t_{1-\alpha, n-1} = t_{0,95;9} = 1,8331$.

Pomocou MATLABu dostaneme:

>> alfa=.05;c=tinv(1-alfa,n-1)

c =
1.8331

Teda kritická oblasť je $K_\alpha = (-\infty; -1,8331)$.

• Pretože $t \notin K_\alpha$, hypotézu H_0 nezamietame.

Riešenie - spôsob B:

Použijeme

`[h,sig,ci] = ttest(x,m,alpha,tail)`

`>> m=mean(x);[h,significance,ci] = ttest(x,m,0.05,-1)`

`h =`

`0`

`significance =`

`0.5000`

`ci =`

`-Inf 38.0631` % 95 %-ný jednostranný interval spoľahlivosti pre strednú hodnotu

Záver: Pretože $h = 0$ hypotézu H_0 nezamietame.

Príklad Na 32 súčiastkach sme merali priemer hriadeľa a zistili sme modifikovaný výberový rozptyl $s^{*2} = 0,3$. Predpokladáme, že sledovaný znak má normálne rozdelenie. Na hladine významnosti $\alpha = 0,05$ testujeme hypotézu $H_0 : \sigma^2 = 0,2$ proti $H_1 : \sigma^2 > 0,2$.

Riešenie:

• Použijeme χ^2 -test zhody rozptylu σ^2 so známou konštantou σ_0^2 .

Budeme testovať nulovú hypotézu $H_0 : \sigma^2 = 0,2$ proti alternatívnej hypotéze $H_1 : \sigma^2 > 0,2$,

kde $\sigma_0^2 = 0,2$. Použijeme testovaciu charakteristiku $\chi^2 = \frac{n-1}{\sigma_0^2} \cdot s^{*2}$.

• $\alpha = 0,05$.

• Výpočet testovacej charakteristiky: $\chi^2 = \frac{31 \cdot 0,3}{0,2} = 46,5$.

• Hypotézu H_0 nezamietame, keď bude platiť $\chi^2 \notin K_\alpha$, pričom pre kritickú oblasť v tomto prípade platí: $K_\alpha = (\chi_{1-\alpha, n-1}^2; \infty)$.

Použijeme kvantil χ^2 rozdelenia: $c = \chi_{1-\alpha, n-1}^2 = \chi_{0,95;31}^2 = 44,9853$.

Pomocou MATLABu dostaneme:

`>> c=chi2inv(0.95,31)`

`c =`

`44.9853`

Teda kritická oblasť je $K_\alpha = (44,9853; \infty)$.

• Pretože $\chi^2 \in K_\alpha$, hypotézu H_0 zamietame v prospech alternatívy H_1 .

Príklad Vzorky chemickej látky sme analyzovali dvoma metódami: a) polarografickou, b) titračnou metódou. Obdržali sme výsledky:

a)	38,2	36,4	37,7	36,1	37,9	37,8	–	–
b)	39,5	38,7	37,8	38,6	39,2	39,1	38,9	39,2

Na hladine významnosti $\alpha = 0,05$ overme hypotézu o rovnosti stredných hodnôt vzoriek chemickej látky oboch metód, pričom predpokladáme normálne rozdelenie pri oboch výberových súboroch.

Riešenie - spôsob A:

- Použijeme t-test zhody dvoch stredných hodnôt μ_1 a μ_2 , pričom σ_1 a σ_2 nepoznáme.

Budeme testovať nulovú hypotézu $H_0 : \mu_1 = \mu_2$ proti alternatívnej hypotéze $H_1 : \mu_1 \neq \mu_2$, kde μ_1 a μ_2 sú stredné hodnoty vzoriek chemickej látky oboch metód. Pri oboch výberových súboroch predpokladáme normálne rozdelenie s neznámymi, ale rovnakými smerodajnými odchýlkami, t.j. $\sigma_1 = \sigma_2$. Použijeme testovaciu charakteristiku:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}{\sqrt{(n_1 - 1) \cdot s_1^{*2} + (n_2 - 1) \cdot s_2^{*2}}} \cdot \sqrt{n_1 + n_2 - 2}.$$

- $\alpha = 0,05$.
- Výpočet ilustrujeme použitím MATLABu, pričom testovaciu charakteristiku uložíme do premennej t:

```
>> x1=[382,364,377,361,379,378]/10;x2=[395,387,378,386,392,391,389,392]/10;
n1=length(x1);n2=length(x2);
t=(mean(x1)-mean(x2))*sqrt(n1*n2*(n1+n2-2)/(n1+n2)/((n1-1)*var(x1)+...
(n2-1)*var(x2)))
```

```
t =
-4.0864
```

- Hypotézu H_0 nezamietame, keď bude platiť $t \notin K_\alpha$, pričom pre kritickú oblasť v tomto prípade platí:

$$K_\alpha = \left(-\infty; -t_{1-\frac{\alpha}{2}, n_1+n_2-2} \right) \cup \left(t_{1-\frac{\alpha}{2}, n_1+n_2-2}; \infty \right).$$

Použijeme kvantil Studentovho t -rozdelenia: $c = t_{1-\frac{\alpha}{2}, n_1+n_2-2} = t_{0,975;12} = 2,1788$.

Použitím MATLABu dostaneme:

```
>> alfa=.05;c=tinv(1-alfa/2,n1+n2-2)
```

```
c =
2.1788
```

Teda kritická oblasť je $K_\alpha = (-\infty; -2,1788) \cup (2,1788; \infty)$.

- Pretože $t \in K_\alpha$, hypotézu H_0 zamietame v prospech alternatívy H_1 .

Riešenie - spôsob B:

Použijeme

[h,significance,ci,stats] = ttest2(x1,x2,alpha,tail)

(allows specification of one- or two-tailed tests, where tail is a flag that specifies one of three alternative hypotheses: tail = 0 specifies the alternative (default) tail = 1 specifies the alternative tail = -1 specifies the alternative)

```
>> x1=[38.2,36.4,37.7,36.1,37.9,37.8];
>> x2=[39.5,38.7,37.8,38.6,39.2,39.1,38.9,39.2];
>> [h,significance,ci,stats] = ttest2(x1,x2,0.05,0)
h = 1
significance = 0.0015
ci = -2.3381 -0.7119
stats =
```

tstat: -4.0864 % hodnota testovacej charakteristiky (v A spôsobe je to t)

df: 12 % počet stupňov voľnosti $df = n_1 + n_2 - 2 = 12$

Záver: Hypotézu H_0 zamietame.

Príklad Základnú surovinu pre výrobu na automatizovanej výrobnjej linke dodávajú dva výrobné podniky. Pre plynulosť a kvalitu výroby je dôležité, aby základná surovina mala rovnomernú hrúbku. Z dodávky suroviny od prvého výrobcu sa uskutočnilo 25 meraní hrúbky suroviny a jej variabilitu charakterizuje modifikovaný výberový rozptyl $s_1^{*2} = 0,6940$. Pri 31 analogických meraniach z dodávky od druhého výrobcu sa zistil modifikovaný výberový rozptyl hrúbky $s_2^{*2} = 0,6110$. Na hladine významnosti $\alpha = 0,01$ testujte hypotézu o tom, že pokiaľ ide o rovnomernosť hrúbky suroviny, je kvalita dodávok od obidvoch dodávateľov rovnaká. Predpokladáme normálne rozdelenie u oboch výberových súborov.

Riešenie:

- Použijeme F-test zhody dvoch rozptylov σ_1^2 a σ_2^2 .

Budeme testovať nulovú hypotézu $H_0 : \sigma_1^2 = \sigma_2^2$ proti $H_1 : \sigma_1^2 \neq \sigma_2^2$, kde σ_1^2 a σ_2^2 sú rozptyly

hrúbky suroviny od oboch dodávateľov. Použijeme testovaciu charakteristiku $F = \frac{s_1^{*2}}{s_2^{*2}}$.

- $\alpha = 0,01$.

- Výpočet testovacej charakteristiky: $F = \frac{0,6940}{0,6110}$.

- Hypotézu H_0 nezamietame, keď bude platiť $F \notin K_\alpha$, pričom pre kritickú oblasť v tomto prípade platí:

$$K_\alpha = \left(0; F_{\frac{\alpha}{2}, n_1-1, n_2-1} \right) \cup \left(F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}; \infty \right).$$

Použijeme kvantily Fisherovho-Snedecorovho rozdelenia, ktoré vypočítame použitím MATLABu takto:

```
>> alfa=.01;n1=25;n2=31;cl=finv(alfa/2,n1-1,n2-1),cp=finv(1-alfa/2,n1-1,n2-1)
```

cl =

0.3487

cp =

2.7272

Teda kritická oblasť bude $K_\alpha = (0 ; 0,3487) \cup (2,7272 ; \infty)$.

- Pretože $F \notin K_\alpha$, hypotézu H_0 nezamietame, t.j. rozdiel vo variabilite hrúbky dodávanej suroviny od oboch dodávateľov považujeme za štatisticky nevýznamný.