

T-próbák



Szűcs Mónika, Griechisch Erika, Rárosi Ferenc
SZTE ÁOK-TTIK Orvosi Fizikai és Orvosi Informatikai Intézet

Utoljára frissítve: 2020. június 4.



Abban az esetben, ha a vizsgálandó állítás a populáció(k) átlagáról szól, leggyakrabban a t-próbák valamelyikét használjuk.

1. Egymintás t-próba

Célja annak vizsgálata, hogy a populációátlag eltér-e egy adott, hipotetikus értéktől. Továbbiakban ezt c -vel jelöljük.

Feltétele, hogy a vizsgált változó normális eloszlásból származik.

Hipotézisek:

- $H_0: \mu = c$, vagyis a populációátlag egyenlő az adott értékkel.
- $H_1: \mu \neq c$, vagyis a populációátlag eltér az adott értéktől.

Számolás és döntés:

I. konfidencia-intervallum alapján

$$(1 - \alpha) \cdot 100\% \text{ konfidencia-intervallum: } \left(\bar{x} - t_\alpha \cdot \frac{SD}{\sqrt{n}} ; \bar{x} + t_\alpha \cdot \frac{SD}{\sqrt{n}} \right)$$

ahol t_α a t-eloszlás α -hoz és az $n - 1$ szabadságfokhoz tartozó kritikus értéke.
($\bar{x}$ a mintaátlag; SD a minta szórása; n a minta elemszáma)

Döntési szabály:

- Ha a c benne van a $(1 - \alpha) \cdot 100\%$ szintű konfidencia-intervallumban, H_0 -t elfogadjuk. A populációátlag nem tér el szignifikánsan az adott értéktől α szinten, vagyis nincs elegendő okunk H_0 elvetéséhez.
- Ha a c nincs benne a $(1 - \alpha) \cdot 100\%$ szintű konfidencia-intervallumban, H_0 -t elvetjük. A populációátlag szignifikánsan különbözik az adott értéktől α szinten.

II. Próbat statisztika

$$t = \frac{\bar{x} - c}{\frac{SD}{\sqrt{n}}} = \frac{\bar{x} - c}{SE}$$

Ha H_0 igaz, a próbat statisztika (t) Student-féle t-eloszlást követ $n - 1$ szabadságfokkal.

($SE = SD/\sqrt{n}$ a standard hiba)

A t-eloszlás kritikus értéke (t_α) α nagyságától és a szabadságfoktól függ. α a t-eloszlás görbéje alatti terület a $(-t_\alpha, t_\alpha)$ intervallumon kívül.



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



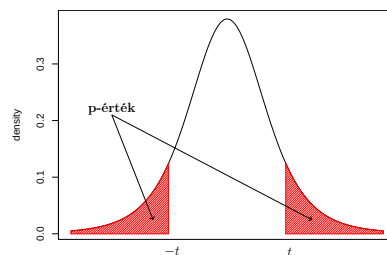
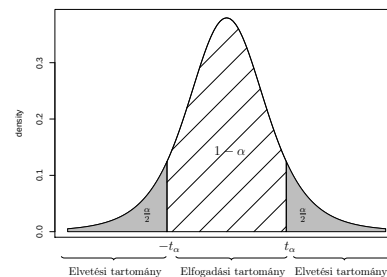
BEFEKTETÉS A JÖVŐBE

Döntési szabály:

- Ha $t \in (-t_\alpha, t_\alpha)$, azaz $|t| < t_\alpha$, akkor H_0 -t elfogadjuk.
- Ha $t \notin (-t_\alpha, t_\alpha)$, azaz $|t| > t_\alpha$, akkor H_0 -t elvetjük.

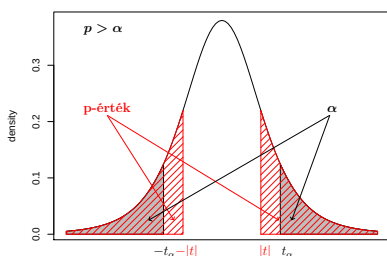
III. p-érték

A p -érték a sűrűségfüggvény $(-t, t)$ intervallumon kívül eső görbéje alatti terület. A p -érték annak a valószínűsége, hogy a nullhipotézis fennállása esetén legalább annyira ellentmondó mintát kapunk, mint a ténylegesen megfigyelt minta.

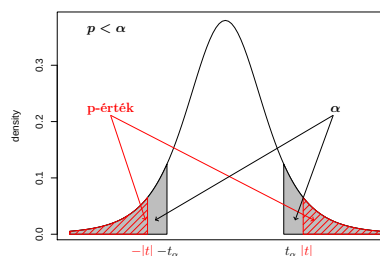


Döntési szabály:

Ha $p > \alpha$, H_0 -t elfogadjuk.



Ha $p < \alpha$, H_0 -t elvetjük.



Számolás R-rel:

```
> t.test(tomeg, mu=60, conf.level=0.95)
```

One Sample t-test

```
data: tomeg
t = 2.707, df = 66, p-value = 0.008636
alternative hypothesis: true mean is not equal to 60
95 percent confidence interval:
 61.0419 66.8984
sample estimates:
mean of x
 63.97015
```

2. Páros t-próba

Célja annak vizsgálata, hogy két összetartozó populáció átlaga eltér-e egymástól, *pl.: előtte - utána*.

Feltétele, hogy a két változó különbsége normális eloszlásból származik.

Hipotézisek:

- $H_0: \mu_1 = \mu_2$ ($\mu_{\text{diff}} = 0$), vagyis a két összetartozó populáció átlaga nem különbözik.
- $H_1: \mu_1 \neq \mu_2$ ($\mu_{\text{diff}} \neq 0$), vagyis a két összetartozó populáció átlaga eltér egymástól.

Számolás és döntés:

I. konfidencia-intervallum

$$(1 - \alpha) \cdot 100\% \text{ konfidencia-intervallum: } \left(\bar{x}_{\text{diff}} - t_{\alpha} \cdot \frac{SD_{\text{diff}}}{\sqrt{n}} ; \bar{x}_{\text{diff}} + t_{\alpha} \cdot \frac{SD_{\text{diff}}}{\sqrt{n}} \right)$$

ahol t_{α} a t-eloszlás α szignifikanciaszinthez és az $n - 1$ szabadságfokhoz tartozó kritikus értéke.
($\bar{x}_{\text{diff}}$ a különbségek átlaga; SD_{diff} a különbségek szórása; n a minta elemszáma)

Döntési szabály:

- Ha a 0 benne van a $(1 - \alpha) \cdot 100\%$ szintű konfidencia-intervallumban, H_0 -t elfogadjuk. Nincs elegendő okunk H_0 elvetéséhez; nincs szignifikáns különbség a két populációátlag között α szinten.
- Ha a 0 nincs benne a $(1 - \alpha) \cdot 100\%$ szintű konfidencia-intervallumban, H_0 -t elvetjük; a két populációátlag közötti különbség szignifikáns α szinten.

II. Próbastatisztika

$$t = \frac{\bar{x}_{\text{diff}}}{\frac{SD_{\text{diff}}}{\sqrt{n}}}$$

Ha H_0 igaz, a próbastatisztika (t) Student-féle t-eloszlást követ $n - 1$ szabadságfokkal.

Döntési szabály:

- Ha $t \in (-t_{\alpha}, t_{\alpha})$, vagyis $|t| < t_{\alpha}$, H_0 -t elfogadjuk.
- Ha $t \notin (-t_{\alpha}, t_{\alpha})$, vagyis $|t| > t_{\alpha}$, H_0 -t elvetjük.

ahol t_{α} az α szignifikanciaszinthez és az $n - 1$ szabadságfokhoz tartozó kritikus érték.

III. p-érték

Döntési szabály:

- Ha $p > \alpha$, H_0 -t elfogadjuk.
- Ha $p < \alpha$, H_0 -t elvetjük.

Számolás R-rel:

```
> t.test(tomeg_egyve, tomeg, paired=TRUE)
```

Paired t-test

```
data: tomeg_egyve and tomeg
t = -2.0122, df = 66, p-value = 0.04829
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.932775955 -0.007522552
sample estimates:
mean of the differences
 -0.9701493
```

3. Kétmintás t-próba (egyenlő varianciák esete)

Célja annak vizsgálata, hogy két független populáció átlaga eltér-e egymástól.

Feltétele, hogy a két populáció, melyekből a mintánk származik külön-külön normális eloszlást követ, és a két populáció varianciája egyenlő.

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

Hipotézisek:

- $H_0: \mu_1 = \mu_2$, vagyis a két populáció átlaga megegyezik.
- $H_1: \mu_1 \neq \mu_2$, vagyis a két populáció átlaga különböző.

Számolás és döntés:

I. Próbastatisztika

$$t = \frac{\bar{x} - \bar{y}}{s_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}, \text{ ahol } s_p = \sqrt{\frac{(n-1) \cdot s_x^2 + (m-1) \cdot s_y^2}{n+m-2}}$$

n, m a két minta elemszáma, s_x és s_y a két minta szórása.

Ha H_0 igaz, a próbastatisztika (t) Student-féle t-eloszlást követ $n + m - 2$ szabadságfokkal.

Döntési szabály:

- Ha $t \in (-t_\alpha, t_\alpha)$, vagyis $|t| < t_\alpha$, H_0 -t elfogadjuk.
 - Ha $t \notin (-t_\alpha, t_\alpha)$, vagyis $|t| > t_\alpha$, H_0 -t elvetjük.
- ahol t_α a t-eloszlás α szignifikanciaszinthez és az $n + m - 2$ szabadságfokhoz tartozó kritikus értéke.

II. p-érték

Döntési szabály:

- Ha $p > \alpha$, H_0 -t elfogadjuk.
- Ha $p < \alpha$, H_0 -t elvetjük.

Számolás R-rel:

```
> t.test(BMI_fiu, BMI_lany, var.equal=TRUE)
```

Two Sample t-test

```
data: BMI_fiu and BMI_lany
t = 3.2315, df = 65, p-value = 0.001935
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 0.9002415 3.8134303
sample estimates:
mean of x mean of y
 22.98733  20.63050
```

```
> t.test(BMI ~ nem, var.equal=TRUE)
```

Two Sample t-test

```
data: BMI by nem
t = 3.2315, df = 65, p-value = 0.001935
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 0.9002415 3.8134303
sample estimates:
mean in group fiu mean in group lany
 22.98733          20.63050
```

4. Welch féle t-próba

Célja annak vizsgálata, hogy két független populáció átlaga eltér-e egymástól.

Feltétele, hogy a minták olyan két populációból származnak, melyek külön-külön normális eloszlást követnek.

A Welch-féle t-próbát akkor érdemes használni a kétmintás t-próba helyett, ha két populáció varianciája különböző, vagy nem tudunk semmit a varianciák egyenlőségéről.

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE

Hipotézisek:

- $H_0: \mu_1 = \mu_2$, vagyis a két populáció átlaga megegyezik.
- $H_1: \mu_1 \neq \mu_2$, vagyis a két populáció átlaga különböző.

Számolás és döntés:

I. Próbat statisztika

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

ahol n, m a mintaelemszámok, s_x és s_y a két minta szórása.

Ha H_0 igaz, t közelítőleg Student-féle t-eloszlást követ $df = \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{(s_x^2/n)^2}{n-1} + \frac{(s_y^2/m)^2}{m-1}}$ szabadságfokkal.

Döntési szabály:

- Ha $t \in (-t_\alpha, t_\alpha)$, vagyis $|t| < t_\alpha$, H_0 -t elfogadjuk.
- Ha $t \notin (-t_\alpha, t_\alpha)$, vagyis $|t| > t_\alpha$, H_0 -t elvetjük.

ahol t_α az α szignifikanciaszinthez és a df szabadságfokhoz tartozó kritikus érték.

II. p-érték

Döntési szabály:

- Ha $p > \alpha$, H_0 -t elfogadjuk.
- Ha $p < \alpha$, H_0 -t elvetjük.

Számolás R-rel:

```
> t.test(BMI_fiu, BMI_lany)
```

Welch Two Sample t-test

```
data: BMI_fiu and BMI_lany
t = 2.814, df = 30.234, p-value = 0.008523
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 0.6469025 4.0667692
sample estimates:
mean of x mean of y
 22.98733  20.63050
```

```
> t.test(BMI ~ nem)
```

Welch Two Sample t-test

```
data: BMI by nem
t = 2.814, df = 30.234, p-value = 0.008523
alternative hypothesis: true difference in means is not
equal to 0
95 percent confidence interval:
 0.6469025 4.0667692
sample estimates:
mean in group fiu mean in group lany
 22.98733          20.63050
```

Felhasznált irodalom

- Reiczgel Jenő, Harnos Andrea, Solymosi Norbert: Biostatisztika nem statisztikusoknak, Pars Kft. (2014)
- Reiczgel Jenő: Válogatott fejezetek a biostatistikából, SZIE ÁOTK (2005)
<http://www2.univet.hu/users/jreiczgel/valfej/val-fej-jegyzet-2005-02-05.pdf> (2019.05.21.)

Jelen tananyag a Szegedi Tudományegyetemen készült
az Európai Unió támogatásával.
Projekt azonosító: EFOP-3.4.3-16-2016-00014

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE