Paraméteres eljárások, normalitásvizsgálat, t-eloszlás, t-próbák

Statisztika I., 2. alkalom

Paraméteres eljárások

- Becsülik a populáció egy paraméterét
- Alkalmazásuknak számos feltétele van (paraméterek és a változó eloszlása)
- Csak normál eloszlású, folytonos változók esetén alkalmazhatók!

Normalitásvizsgálat:

- Grafikus exploratív technikák, pl. hisztogram (szimmetria, egycsúcsúság)
- Normalitás tesztelése, pl. ferdeségi együttható, csúcsossági együttható, nemparaméteres eljárások

Amennyiben a változó nagy valószínűséggel nem normális elosztású, robusztus, nemparaméteres eljárások alkalmazása javasolt!

Normalitásvizsgálat

Normális eloszlású változók esetében a *ferdeségi együttható* értéke nulla.

A tapasztalati ferdeségi együttható:

$$g_1 = \left(\frac{1}{N} \sum_i (x_i - \overline{x})^3 \right) / s^3 = \frac{1}{N} \sum_i (z_i)^3$$

Ha az érték pozitív, akkor az eloszlás a jobb, ha negatív, akkor a bal oldalra nyúlik el.

Normális eloszlású változó esetében a *csúcsossági együttható* értéke nulla.

A tapasztalati csúcsossági együttható:

$$g_2 = \left\{ \left(\frac{1}{N} \sum_i (x_i - \overline{x})^4 \right) / s^4 \right\} - 3 = \frac{1}{N} \sum_i z^4 - 3$$

Normális eloszlású változó esetén, ha a mintanagyság elég nagy (n>30), mind a két együttható mintabeli eloszlása normális eloszlású, így a standardizált érték a z eloszlással összevethető.

g1 st.hibája:

$$\sigma_1 = \sqrt{\frac{6}{N}}$$

g2 st.hibája:

$$\sigma_2 = \sqrt{\frac{24}{N}}$$

Normalitásvizsgálat

Példa. (Vargha, 2000)

99 fiú 10 éves kori testsúlya:

$$g_1 = \left(\frac{1}{N} \sum_i (x_i - \overline{x})^3 \right) / s^3 = \frac{1}{N} (z_i)^3 = 1.84$$

$$\sigma_1 = \sqrt{\frac{6}{N}} = \sqrt{\frac{6}{99}} = 0.246$$

$$z_1 = 1.84 / 0.246 \approx 7.48$$

$$p_1 < 3.719 * 10^{-14}$$

$$g_2 = \left\{ \left(\frac{1}{N} \sum_i (x_i - \overline{x})^4 \right) / s^4 \right\} - 3 = \frac{1}{N} (z)^4 - 3 = 4.1$$

$$\sigma_2 = \sqrt{\frac{24}{N}} = \sqrt{\frac{24}{99}} = 0.492$$

$$z_2 = 4.1 / 0.492 \approx 8.33$$

$$p_2 < 0.001$$

Student t-eloszlás

A korábbiakban már használt tesz statisztikák :

z-eloszlás (standard normális eloszlás):

a különböző eloszlásokat összehasonlíthatóvá teszi, egyes értékek valószínűsége normál eloszlás esetén kiszámolható

$$z = \frac{x - \mu}{\sigma}$$

u-próba (z-próba): akkor használható, ha a populációbeli szórás ismert. Átlagok vizsgálatára használjuk.

$$u = \frac{\bar{x} - m}{\sigma / \sqrt{N}}$$

Student t-eloszlás: akkor használjuk, ha nincs elképzelésünk a populációbeli szórásról. A t–eloszlás zéró átlagú, nagyobb csúcsosságú, mint a anormál eloszlás. Nagy mintanagyság esetén jól közelíti a normál eloszlást.

$$t = \frac{\bar{x} - m}{s / \sqrt{N}}$$

Egymintás t-próba

W.S Gosset (1876-1936) Az ír Guinness sörgyár dolgozója, 1908-ban Student néven publikál.

egymintás t-próba (adott populációátlag lehetséges-e a a minta alapján)

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}} \qquad f=N-1$$

H0: A populáció átlaga egyenlő egy feltételezett értékkel.

Példa (Vargha, 2000)
Súlycsökkenés:
-2,10,10,11,12,13,13,14,16,18,18,35
Lehet-e, hogy a kúra átlagosan 10 kg súlycsökkenést eredményez?

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}} = \frac{14 - 10}{\sqrt{Q / (N-1)} / \sqrt{N}} = \frac{14 - 10}{\sqrt{780 / (12-1)} / \sqrt{12}} = 1.646$$

f=11, α=0.05, t=1.646, p<0.128.
r-parancs: t.test(adat, mu=10)

Páros t-próba

Páros vagy összetartozó mintás t-próba.

Két normális eloszlású mintáról van szó. A két minta minden eleme egyértelműen összepárosítható.

H0: A populációk átlaga azonos.

A különbségeket úgy kezelhetjük, mintha egyetlen mintával állnánk szemben. Gyakorlatilag azt vizsgáljuk, hogy a különbség átlaga nulla-e.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

Példa (Schafer, 2002)

Egypetéjű ikreket vizsgálnak mágneses rezonancia vizsgálattal, olyanokat, amelyeknek egyike skizofrén. A bal oldali hippocampus nagyságát állapítják meg. 15 ikerpárt vizsgálnak.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{N}} = \frac{0.199}{.238 / \sqrt{15}} = \frac{0.199}{0.0615} = 3.236$$

f=14, p<0.006

r-parancs: t.test(adat1, adat2, paired=T)

Kétmintás t-próba

Kétmintás, független mintás t-próba

Két normális eloszlású mintáról van szó. Párok nem képezhetőek egyértelműen.
H0: A populációk átlaga azonos.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma_1 / \sqrt{N_1} + \sigma_2 / \sqrt{N_2}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\sigma_1^{\,2} / N_1 + \sigma_2^{\,2} / N_2}}$$

$f=N_1+N_2-2$

A szórás becslése annak megfelelően alakul, hogy feltételezhetjük-e, hogy a populációbeli varianciák azonosak. Ennek vizsgálatára a Levene-teszt szolgál. A varianciák elméleti átlagos abszolút eltérést vizsgálja. F eloszlású a teszt statisztikája.

Kétmintás t-próba

Kétmintás, független mintás t-próba

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma_1 / \sqrt{N_1} + \sigma_2 / \sqrt{N_2}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\sigma_1^{\ 2} / N_1 + \sigma_2^{\ 2} / N_2}}$$

Ha a populációbeli varianciáról feltételezzük, hogy azonos a két csoportban, így a összevont becslése a szórásnak:

$$s_p = \sqrt{\frac{(N_1 - 1)s_1^{\ 2} + (N_2 - 1)s_2^{\ 2}}{(N_1 + N_2 - 2)}}$$, így $$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

Példa (Oláh, 1985)

kreatív mérnökök 47 fős minta, 18.72 az originalitás átlaga, 6.02 a szórása

nem kreatív mérnökök 50 fős minta, 16.27 az originalitás átlaga, 5.78 a szórása

f=47+50-2=95   t(95)=1.987

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} = \frac{(18.71 - 16.27)}{\sqrt{34.78}\sqrt{\frac{1}{47} + \frac{1}{50}}} = 2.045 \qquad s_p = \sqrt{\frac{(47-1)6.02^2 + (50-1)5.78^2}{(47+50-2)}} = \sqrt{34.78}$$

r-parancs: t.test(adat1, adat2, paired=F)

Kétmintás t-próba

Kétmintás, független mintás t-próba

Ha a populációbeli varianciáról nem feltételezhetjük, hogy azonos a két csoportban, akkor a Welch féle d-próbát alkalmazhatjuk.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma_1 / \sqrt{N_1} + \sigma_2 / \sqrt{N_2}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\sigma_1^2 / N_1 + \sigma_2^2 / N_2}}$$

$$d = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s_1^2 / N_1 + s_2^2 / N_2}}$$

, ahol $f = \dfrac{(a+b)^2}{\dfrac{a^2}{N_1 - 1} + \dfrac{b^2}{N_2 - 1}}$ és $a = \dfrac{s^2}{N_1}$ $b = \dfrac{s^2}{N_2}$