Alkalmazott statisztika

Egymintás t-próba és páros t-próba

2016. szeptember 21.

Hipotézis vizsgálatok

A hipotézisvizsgálatok során arra keresünk választ, hogy egy sokasággal kapcsolatos bizonyos alapfeltevések (hipotézisek) elfogadhatóak-e egy rendelkezésre álló minta alapján.
Null-hipotézisnek (H_0) nevezzük azt a kitüntetett hipotézist, amiben hiszünk.
Ha a null-hipotézist nem fogadjuk el, akkor az ellenhipotézis (H_1) mellett döntünk.
Ha a nullhipotézist elvetjük, pedig igaz, akkor elsőfajú hibát követünk el, melynek mértékét a tesztelés során beállítjuk (α szignifikancia szint); ha ellenben elfogadjuk, pedig nem igaz, akkor másodfajú hibát követünk el.

Egymintás t-próba

Legyen ξ normális eloszlású valószínűségi változó ismeretlen μ várható értékkel és σ szórással. A hozzá tartozó minta: $\xi_1, \ldots, \xi_n$.
Rögzített μ_0 esetén a következő hipotéziseket vizsgáljuk:

$$H_0 : \mu = \mu_0, \; H_1 : \mu \neq \mu_0.$$

Konfidencia intervallum

Legyen θ ismeretlen paraméter (pl. várhatóérték, szórás). Az intervallum becslés lényege olyan intervallum konstruálása, amelybe θ nagy valószínűséggel (álltalában 0.95 vagy 0.99) beleesik.

Azt mondjuk, hogy (S_n, T_n) egy $1-\alpha$ megbízhatósági szintű konfidencia intervallum θ-ra, ha $P(S_n < \theta < T_n) = 1 - \alpha$.

Konfidencia intervallum szerkesztésének lépései:

- alakítsuk át θ-t $Z_n(\theta)$-vá, melynek ismerjük az eloszlását
- $Z_n(\theta)$-ra szerkesszünk intervallumot: $P(a < Z_n(\theta) < b) = 1 - \alpha$
- fejezzük ki θ-t $Z_n(\theta)$-ból: $P(S_n(a, b) < \theta < T_n(a, b)) = 1 - \alpha$

Konfidencia intervallum várható értékre

Legyen $\xi_1, \xi_2, \ldots \xi_n$ egy $N(\mu, \sigma^2)$-ből vett minta.

Tudjuk, hogy $E_n(\xi) \approx \frac{\xi_1 + \cdots + \xi_n}{n}$.

Alakítsuk át μ-t: $Z_n(\mu) = \frac{E_n(\xi) - \mu}{\sqrt{V_n^*(\xi)/n}} \approx t(n-1)$.

A $t(n-1)$ eloszlás szimmetriáját felhasználva a $P(-x_\alpha < Z_n(\mu) < x_\alpha) = 2\Phi_{n-1}(x_\alpha - 1) = 1 - \alpha$ összefüggésből adódik, hogy $x_\alpha = \Phi_{n-1}^{-1}(1 - \alpha/2)$. Ebből: $1 - \alpha = P(-x_\alpha < \frac{E_n(\xi) - \mu}{\sqrt{V_n^*(\xi)/n}} < x_\alpha) =$ $P(E_n(\xi) - x_\alpha \frac{D_n^*(\xi)}{\sqrt{n}} < \mu < E_n(\xi) + x_\alpha \frac{D_n^*(\xi)}{\sqrt{n}})$.

Döntés konfidencia intervallum alapján

A nullhipotézist nem utasítjuk el, ha a konfidencia intervallum tartalmazza μ_0-t. Tehát nem tudunk szignifikáns különbséget kimutatni a várható értékek között α szinten.

A nullhipotézist elvetjük, ha a konfidencia intervallum NEM tartalmazza μ_0-t. Tehát a várható értékek közötti különbség szignifikáns α szinten.

Döntés tesztstatisztika alapján

Legyen $t = \frac{E_n(\xi)-\mu}{\sqrt{V_n^*(\xi)/n}}$ tesztstatisztika és a hozzá tartozó kritikus érték: $t_\alpha = \Phi_{n-1}^{-1}(1-\alpha/2)$.
A nullhipotézist nem utasítjuk el, ha $|t| < t_\alpha$. Tehát nem tudunk szignifikáns különbséget kimutatni a várható értékek között α szinten.
Különben a nullhipotézist elvetjük. Tehát a várható értékek közötti különbség szignifikáns α szinten.

Döntés p érték alapján

Az empirikus szignifikancia szint (p érték) annak a valószínűsége, hogy a próbastatisztika a mintából kiszámított értéket veszi fel.

Másként fogalmazva: a p értéke az a legnagyobb szignifikancia szint, amely mellett a null-hipotézist még elfogadjuk.

A null-hipotézist nem utasítjuk el, ha $\alpha < p$. Tehát nem tudunk szignifikáns különbséget kimutatni a várható értékek között α szinten. Különben a nullhipotézist elvetjük. Tehát a várható értékek közötti különbség szignifikáns α szinten.

Páros t-próba

Legyenek $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ független azonos normális eloszlású megfigyelés párok. Legyen $\mu_1 = E(\xi)$ és $\mu_2 = E(\eta)$. A következő hipotéziseket vizsgáljuk:

$$H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 \neq \mu_2$$

A hipotézistesztelést visszavezetjük az egymintás t-próbára: legyen $\xi_1 - \eta_1, \ldots, \xi_n - \eta_n$ egy μ várható értékű normális eloszlású valószínűségi változóhoz tartozó minta. A következő hipotéziseket vizsgáljuk:

$$H_0 : \mu = 0,\ H_1 : \mu \neq 0$$