

# Valószínűségszámítás és Matematikai Statisztika

Miskolc, 2025.

Dr. Glavosits Tamás

## 12. előadás

# Statisztika III. Hipotézisvizsgálat

# 1. Hipotézisvizsgálat

# Alap és ellenhipotézis

## Definíció

Legyen  $\xi = (\xi_1, \xi_2, \dots, \xi_n)$  egy minta, a közös eloszlásfüggvény  $\mathbb{F}(\cdot, \vartheta)$ , ahol  $\vartheta \in \Theta$  az ismeretlen paraméter. A  $\Theta$  paraméterteret felbontjuk két részre  $\Theta_0, \Theta_1$  úgy, hogy  $\Theta_0 \cup \Theta_1 = \Theta$ .

$H_0$  **alaphipotézis**:  $\vartheta \in \Theta_0$ ;

$H_1$  **ellenhipotézis**:  $\vartheta \in \Theta_1$ .

Ha  $\vartheta \in \Theta_0 \cap \Theta_1$ , akkor véletlenszerűen döntünk.

## Definíció

Megőrizve a korábbi jelöléseket

$H_0$  **egyszerű alaphipotézis**, ha  $\Theta_0 = \{\vartheta_0\}$ , azaz  $\vartheta = \vartheta_0$ ;

$H_1$  **ellenhipotézis**, ha  $\Theta_1 = \Theta \setminus \{\vartheta_0\}$ , azaz  $\vartheta \neq \vartheta_0$ .

## 1.2. A döntés, szignifikancia szint, első és másodfajú hiba

## Kritikus tartomány

A  $\chi = \{(x_1, x_2, \dots, x_n) \mid x_1, x_2, \dots, x_n \text{ mintarealizáció}\} \subseteq \mathbb{R}^N$  mintateret két diszjunkt halmazra bontjuk  $\chi = C \cup \overline{C}$ , a  $C$  halmazt **kritikus tartomány**nak nevezzük.

- Ha  $(x_1, x_2, \dots, x_n) \in C$ , akkor  $H_0$ -t elvetjük,
- Ha  $(x_1, x_2, \dots, x_n) \in \overline{C}$ , akkor  $H_0$ -t elfogadjuk.

## Definíció

Egy próbát  $\alpha$  **szignifikancia szintűnek** nevezünk, ha

$$\mathbb{P}(\xi) \in C | H_0) \leq \alpha.$$

Ha döntést hozunk, elkövethetünk hibákat.

- **Elsőfajú hibának** nevezzük azt a valószínűséget, hogy  $H_0$ -t elvetjük pedig igaz, azaz az elsőfajú hiba a

$$\mathbb{P}(\xi) \in C | H_0)$$

valószínűséget.

- **Másodfajú hibának** nevezzük annak az eseménynek a valószínűségét, hogy  $H_0$ -et elfogadjuk, pedig hamis, azaz a másodfajú hiba a

$$\mathbb{P}(\xi) \in \bar{C} | H_1)$$

valószínűség.

# Erőfüggvény, konzisztens erőfüggvény

## Definíció

A  $W : \Theta \rightarrow [0, 1]$  ( $W = W(C, \vartheta, n)$ )

$$W(\vartheta) = \mathbb{P}(\{\xi \in C | \vartheta\}) \quad (\vartheta \in \Theta)$$

módon definiált függvényt a  $C$  **kritikus tartományhoz tartozó erőfüggvénynek** nevezzük, ahol  $\xi = (\xi_1, \xi_2, \dots, \xi_n)$  a független minta.

## Definíció

A  $C$  kritikus tartományhoz tartozó  $W$  **erőfüggvényt konzisztensnek** nevezzük, ha

$$\lim_{n \rightarrow \infty} W(\vartheta) = 1 \quad (\vartheta \in \Theta_1)$$

# Torzítatlan próba $\alpha$ szinten, erősebb próbák

## Definíció

Egy próbát **torzítatlannak** nevezünk  $\alpha$  szinten, ha

$$\mathbb{P}(\xi \in C|\vartheta) \leq \alpha \quad (\vartheta \in \Theta_0),$$

$$\mathbb{P}(\xi \in C|\vartheta) \geq \alpha \quad (\vartheta \in \Theta_1),$$

## Definíció

Az  $\alpha$  szintű  $C_1$  és  $C_2$  kritikus tartománnyal definiált próbák közül a  $C_1$ -et **erősebbnek** nevezzük a  $C_2$ -nél, ha

$$\mathbb{P}(\xi \in C_1|\vartheta) \geq \mathbb{P}(\xi \in C_2|\vartheta) \quad (\vartheta \in \Theta),$$

azaz minden  $\vartheta \in \Theta$  esetén a  $C_1$ -hez tartozó erőfüggvény nagyobb vagy egyenlő, mint a  $C_2$ -höz tartozó erőfüggvény.

## 2. Nevezetes várható értékere vonatkozó próbák

# Nevezetes várható értékere vonatkozó próbák

- egymintás u-próba;
- egymintás t-próba;
- kétmintás u-próba;
- kétmintás t-próba (Ehhez a két ismeretlen elméleti szórásnégyzetnek meg kell egyezniük, ezt a tényt az F-próbával ellenőrizzük.)

## 2.1. Egymintás u-próba

# Egymintás u-próba

Legyen  $\xi_1, \dots, \xi_n \sim \mathcal{N}(\mu, \sigma^2)$  független minta, ahol  $\sigma^2$  ismert,  $\mu_0 \in \mathbb{R}$  egy adott szám:

**Az egyszerű alap és ellenhipotézis:**

$$H_0 : \mu = \mu_0;$$

$$H_1 : \mu \neq \mu_0.$$

**A döntés:**

1. Úgy döntünk, hogy  $\mu = \mu_0$  feltételezésével

$$\frac{\bar{\xi} - \mu_0}{\sigma} \sqrt{n} \sim \mathcal{N}(0, 1)$$

2. A konfidencia intervallumok szerkesztésénél megismert módon meg kell határozunk az  $u$  értéket, melyre

$$\mathbb{P} \left( -u < \frac{\bar{\xi} - \mu}{\sigma} \sqrt{n} < u, \mid \mu = \mu_0 \right) = 1 - \alpha$$

azaz a megfelelő  $u$  érték az

$$\Phi(u) = 1 - \frac{\alpha}{2}$$

összefüggés alapján számolható.

3. A minta alapján kiszámítjuk az

$$u_{\text{obs}} := \frac{\bar{\xi} - \mu_0}{\sigma} \sqrt{n}$$

értéket.

4. Döntés: Ha  $|u_{\text{obs}}| < u$ , (obs=observed)akkor a  $H_0$  alaphipotézist elfogadjuk, ellenkező esetben elutasítjuk.

## 2.2. Egymintás t-próba

# Egymintás t-próba

Legyen  $\xi_1, \dots, \xi_n \sim \mathcal{N}(\mu, \sigma^2)$  független minta, ahol  $\sigma^2$  ismeretlen,  $\mu_0 \in \mathbb{R}$  egy adott szám.

Az egyszerű alap és ellenhipotézis:

$$H_0 : \mu = \mu_0;$$

$$H_1 : \mu \neq \mu_0.$$

**A döntés:**

1. Úgy döntünk, hogy tudjuk, hogy  $\mu = \mu_0$  feltételezésével

$$\frac{\bar{\xi} - \mu_0}{s_n^*} \sqrt{n} \sim t_{n-1}$$

## Egymintás t-próba

2. A konfidencia intervallumok szerkesztésével megismert módon meg kell határozni azt a  $t$  értéket, amelyre

$$\mathbb{P} \left( -t \frac{\bar{\xi} - \mu_0}{s_n^*} \sqrt{n} < t, \mid \mu = \mu_0 \right) = 1 - \alpha,$$

azaz ha  $\mathbb{F}$  jelöli egy  $t_{n-1}$  eloszlású valószínűségi változó eloszlásfüggvényét, akkor keressük azt a  $t$  értéket, amelyre

$$\mathbb{F}(t) = 1 - \frac{\alpha}{2} \quad \left( \text{illetve } G(t) = \frac{\alpha}{2} \right).$$

3. A minta alapján kiszámoljuk a

$$t_{\text{obs}} := \frac{\bar{\xi} - \mu_0}{s_n^*} \sqrt{n}$$

értéket (obs=observed).

4. A döntést úgy hozzuk meg, hogyha  $|t_{\text{obs}}| < t$ , akkor a  $H_0$  alaphipotézist elfogadjuk, ellenkező esetben elvetjük.

## 2.3 Kétmintás u-próba

## Kétmintás u-próba

Adottak  $\xi_1, \xi_2, \dots, \xi_m \sim \mathcal{N}(\mu_1, \eta_1 \sigma^2)$  és  $\eta_1, \dots, \eta_n \sim \mathcal{N}(\mu_2, \sigma^2)$  független minták. (A  $\xi_i$  és  $\eta_j$  egymástól is függetlenek.)

Az egyszerű alap és ellenhipotézis:

$$H_0 : \mu_1 = \mu_2;$$

$$H_1 : \mu_1 \neq \mu_2.$$

**A döntés:**

1. Tudjuk, hogy  $\mu = \mu_0$  feltételezésével

$$\frac{\bar{\xi} - \bar{\eta}}{\sigma} \sqrt{\frac{mn}{m+n}} \sim \mathcal{N}(0, 1).$$

## Kétmintás $u$ -próba

2. Megkeressük azt az  $u$  értéket, amelyre

$$\mathbb{P}\left(-u < \frac{\bar{\xi} - \bar{\eta}}{\sigma} \sqrt{\frac{mn}{m+n}} < u \mid \mu = \mu_0\right) = 1 - \alpha,$$

azaz

$$\Phi(u) = 1 - \frac{\alpha}{2}.$$

3. A minta alapján kiszámoljuk a

$$u_{\text{obs}} := \frac{\bar{\xi} - \bar{\eta}}{\sigma} \sqrt{\frac{mn}{m+n}}$$

értéket.

4. A döntést úgy hozzuk, hogy  $|u_{\text{obs}}| < u$  esetén a  $H_0$  alaphipotézist elfogadjuk, ellenkező esetben elvetjük.

## 2.4 Kétmintás t-próba

## Kétmintás t-próba

Adottak  $\xi_1, \dots, \xi_m \sim \mathcal{N}(\mu_1, \sigma^2)$  és  $\eta_1, \dots, \eta_n \sim \mathcal{N}(\mu_2, \sigma^2)$  független minták (ahol  $\xi_i$  és  $\eta_j$  egymástól is függetlenek). (Fontos, hogy a minták közös szórásnégyzetű eloszlásokból származnak.)

Az egyszerű alap és ellenhipotézisek:

$$H_0 : \mu_1 = \mu_2;$$

$$H_1 : \mu_1 \neq \mu_2.$$

**A döntés:**

1. Ha  $\mu_1 = \mu_2$ , akkor

$$\frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{(m-1)s_m^* + (n-1)s_n^{*2}}{m+n-2}}} \sqrt{\frac{mn}{m+n}} \sim t_{m+n-2}.$$

## Kétmintás t-próba

2. Keressük azt a  $t$  értéket, amelyre

$$\mathbb{F}(t) = 1 - \frac{\alpha}{2}, \quad \left( \text{illetve } G(t) = \frac{\alpha}{2} \right).$$

3. A minta alapján számoljuk a

$$t_{\text{obs}} := \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{(m-1)s_m^* + (n-1)s_n^{*2}}{m+n-2}}} \sqrt{\frac{mn}{m+n}}$$

értéket.

4. Ha  $|t_{\text{obs}}| < t$ , akkor a  $H_0$  alaphipotézist elfogadjuk, ellenkező esetben elvetjük.

## 2.5. F-próba

## F-próba

Legyen  $\xi_1, \xi_2, \dots, \xi_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$  és  $\eta_1, \eta_2, \dots, \eta_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$ .

$$H_0 : \sigma_1^2 = \sigma_2^2;$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2.$$

**Statisztika:**

$$f = \frac{ns_n^2}{ms_m^2} \sim F_{m-1, n-1} \quad \text{ha} \quad ns_n^2 > ms_m^2.$$

Jelölje  $\mathbb{F}$  az  $F_{m-1, n-1}$  eloszlás eloszlásfüggvényét. Az  $x_a, x_f$  értékeket úgy határozzuk meg, hogy

$$\mathbb{F}(x_f) - \mathbb{F}(x_a) = 1 - \alpha,$$

ami nehezen kivitelezhető, helyette szorítkozhatunk a

$$\mathbb{F}(x_f) = 1 - \frac{\alpha}{2}, \quad \mathbb{F}(x_a) = \frac{\alpha}{2}$$

feltételekre, illetve (táblázattól függően) a  $G(x_f) = \frac{\alpha}{2}$ ,  $G(x_a) = \frac{\alpha}{2}$  értékekre. Ha  $a_a < f_{sz} < x_b$ , akkor  $H_0$ -t elfogadjuk, ellenkező esetben elvetjük.

### 3. Szórásnégyzetre vonatkozó próba

## $\chi^2$ próba

Adott  $\xi_1, \xi_2, \dots, \xi_n \sim \mathcal{N}(\mu, \sigma^2)$  független minta.

$H_0 : \sigma = \sigma_0;$

$H_1 : \sigma \neq \sigma_0.$

**Statisztika:**

$$\frac{ns_n^2}{\sigma_0^2} \sim \chi_m^2, \quad \text{feltéve, hogy } \sigma = \sigma_0.$$

Jelölje  $\mathbb{F}$  a  $\chi_{n-1}^2$  eloszlásfüggvényét. Meghatározzuk az  $x_a, x_f$  értékeket úgy, hogy

$$\mathbb{F}(x_f) - \mathbb{F}(x_a) = 1 - \alpha,$$

amihez elég, ha

$$\mathbb{F}(x_f) = 1 - \frac{\alpha}{2}, \quad \mathbb{F}(x_a) = \frac{\alpha}{2}.$$

A táblázatban  $\mathbb{F}(x) + G(x) = 1$ , így  $G(x_f) = \frac{\alpha}{2}$ ,  $G(x_a) = 1 - \frac{\alpha}{2}$ .

**Döntés:**

Ha  $x_a < \chi < x_f$ , akkor  $H_0$ -t elfogadjuk, ellenkező esetben elvetjük.

## 4. Illeszkedésvizsgálat

# Illeszkedésvizsgálat

Legyen  $\xi_1, \xi_2, \dots, \xi_n$  egy független minta.  $\mathbb{F}$  eloszlásfüggvénnyel.

$$H_0 : \mathbb{F} = \mathbb{F}_0;$$

$$H_1 : \mathbb{F} \neq \mathbb{F}_0.$$

Felosztjuk a számegeyenest  $k$  részre osztópontokkal

$$-\infty = a_0 < a_1 < a_2 < \dots < a_{k-1} < a_k = +\infty$$

$$\mathbb{P}(a_{i-1} \leq \xi < a_i) = \mathbb{F}_0(a_i) - \mathbb{F}_0(a_{i-1}) = p_i \quad (i = 1, 2, \dots, k)$$

A minta alapján tudjuk, hogy az  $i$ -edik intervallumba  $\eta_i$  mintaelem esik bele. Ekkor

$$\chi = \mathbb{F}(1 - \alpha) = G(\alpha);$$

ahol  $\mathbb{F}$  jelöli egy  $k - 1$  szabadságfokú Kszi-négyzet eloszlás eloszlásfüggvényét, illetve  $G(x) = 1 - \mathbb{F}(x)$  ( $x \in \mathbb{R}$ ) módon van definiálva.

# Illeszkedésvizsgálat

A

$$\chi_{\text{obs}} := \sum_{i=1}^k \frac{(\eta_i - np_i)^2}{np_i}.$$

**Döntés:**

Elfogadjuk  $\chi < \chi_{\text{obs}}$  esetén  $H_0$ -t, ellenkező esetben elvetjük.

Vége az 12. előadásnak