Matematikai statisztika

Informatika alapszak, "B" spec.

Zempléni András

Valószínűségelméleti és Statisztika Tanszék
Matematikai Intézet
Természettudományi Kar
Eötvös Loránd Tudományegyetem

Honlap: `zempleni.elte.hu`

E-mail: `andras.zempleni@ttk.elte.hu`

Szoba: D 3-310

4. stat. előadás

Nevezetes paraméteres próbák – áttekintés

Próbák a normális eloszlás várható értékére vonatkozóan:

- Egymintás u-próba, egymintás t-próba
- Kétmintás próbák:

	a két minta független		a két minta párosított
σ_1 és σ_2 ismert	kétmintás u-próba		egymintás u-próba a különbségekre
σ_1 és σ_2 ismeretlen	előzetes F-próba		egymintás t-próba a különbségekre
	$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$	
	kétmintás t-próba	Welch-próba	

Próbák normális eloszlás szórásnégyzetére vonatkozóan:

- Egymintás próba: χ^2-próba
- Kétmintás próba: F-próba

Összefüggő (párosított) minták: X_i és Y_i ugyanahhoz, az i-edik személyhez, tárgyhoz, objektumhoz tartozó véletlen mennyiség, $i = 1, 2, \ldots$

Egymintás u**-próba** (z-próba)

$X_1, \ldots, X_n \sim N(m, \sigma^2)$, ahol σ ismert, m ismeretlen paraméter

Kétoldali: $H_0 : m = m_0$

$\underline{H_1 : m \neq m_0}$

Próbastatisztika: $T(\mathbf{X}) = u := \sqrt{n}\frac{\overline{X} - m_0}{\sigma} \overset{H_0 \text{ esetén}}{\sim} N(0,1)$

Kritikus tartomány: $\mathcal{X}_k = \{\mathrm{x} : |u| > u_{1-\alpha/2}\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak:	$H_1 : m > m_0$	$H_1 : m < m_0$
Kritikus tartomány:	$\mathcal{X}_k = \{\mathrm{x} : u > u_{1-\alpha}\}$	$\mathcal{X}_k = \{\mathrm{x} : u < -u_{1-\alpha}\}$

Áttekintés

Egymintás t-próba

$X_1, \ldots, X_n \sim N(m, \sigma^2)$, ahol m és σ ismeretlen paraméterek

Kétoldali: $H_0 : m = m_0$
$H_1 : m \neq m_0$

Próbastatisztika: $T(\mathbf{X}) = t := \sqrt{n}\frac{\overline{X} - m_0}{S_n^*} \overset{H_0 \text{ esetén}}{\sim} t_{n-1}$

Kritikus tartomány: $\mathcal{X}_k = \{\mathrm{x} : |t| > t_{n-1,1-\alpha/2}\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak:	$H_1 : m > m_0$	$H_1 : m < m_0$
Kritikus tartomány:	$\mathcal{X}_k = \{\mathrm{x} : t > t_{n-1,1-\alpha}\}$	$\mathcal{X}_k = \{\mathrm{x} : t < -t_{n-1,1-\alpha}\}$

Áttekintés

Kétmintás u-próba

$X_1, \ldots, X_n \sim N(m_1, \sigma_1^2)$
$Y_1, \ldots, Y_m \sim N(m_2, \sigma_2^2)$ független minták
ahol m_1, m_2 ismeretlen paraméterek, σ_1, σ_2 ismert

Kétoldali: $H_0 : m_1 = m_2$
$\underline{H_1 : m_1 \neq m_2}$

Próbastatisztika: $T(\mathbf{X}, \mathbf{Y}) = u := \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \overset{H_0 \text{ esetén}}{\sim} N(0,1)$

Kritikus tartomány: $\mathcal{X}_k = \{(\mathrm{x}, \mathrm{y}) : |u| > u_{1-\alpha/2}\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak:	$H_1 : m_1 > m_2$	$H_1 : m_1 < m_2$
Kritikus tartomány:	$\mathcal{X}_k = \{(\mathrm{x}, \mathrm{y}) : u > u_{1-\alpha}\}$	$\mathcal{X}_k = \{(\mathrm{x}, \mathrm{y}) : u < -u_{1-\alpha}\}$

Áttekintés

Kétmintás t-próba

$X_1, \ldots, X_n \sim N(m_1, \sigma_1^2)$
$Y_1, \ldots, Y_m \sim N(m_2, \sigma_2^2)$ független minták
ahol m_1, m_2, $\sigma_1 = \sigma_2$ ismeretlen paraméterek

Kétoldali: $H_0 : m_1 = m_2$
$H_1 : m_1 \neq m_2$

Próbastatisztika: $T(\mathbf{X}, \mathbf{Y}) = t := \sqrt{\frac{nm}{n+m}} \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n-1)(S_1^*)^2 + (m-1)(S_2^*)^2}{n+m-2}}} \overset{H_0 \text{ esetén}}{\sim} \mathrm{t}_{n+m-2}$

Kritikus tartomány: $\mathcal{X}_k = \{(\mathrm{x}, \mathrm{y}) : |t| > t_{n+m-2, 1-\alpha/2}\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak: $H_1 : m_1 > m_2$ $\quad$ $H_1 : m_1 < m_2$

Krit. tart. $\mathcal{X}_k = \{(\mathrm{x}, \mathrm{y}) : t > t_{n+m-2, 1-\alpha}\}$

$\mathcal{X}_k = \{(\mathrm{x}, \mathrm{y}) : t < -t_{n+m-2, 1-\alpha}\}$

Áttekintés

Welch-próba

$X_1,\dots,X_n \sim N(m_1,\sigma_1^2)$ és $Y_1,\dots,Y_m \sim N(m_2,\sigma_2^2)$ független minták
ahol m_1, m_2, $\sigma_1 \neq \sigma_2$ ismeretlen paraméterek

Kétoldali: $H_0 : m_1 = m_2$
$\underline{H_1 : m_1 \neq m_2}$

Próbastatisztika: $T(\mathbf{X},\mathbf{Y}) = t' := \dfrac{\overline{X}-\overline{Y}}{\sqrt{\frac{(S_1^*)^2}{n}+\frac{(S_2^*)^2}{m}}} \overset{H_0 \text{ esetén}}{\sim} \mathrm{t}_f$, ahol

$\frac{1}{f} = \frac{c^2}{n-1} + \frac{(1-c)^2}{m-1}$, $\quad c = \dfrac{\frac{(s_1^*)^2}{n}}{\frac{(s_1^*)^2}{n}+\frac{(s_2^*)^2}{m}}$, ha $s_1^* > s_2^*$ (így csináljuk)

Kritikus tartomány: $\mathcal{X}_k = \{(\mathrm{x},\mathrm{y}) : |t| > t_{f,\alpha/2}\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak:	$H_1 : m_1 > m_2$	$H_1 : m_1 < m_2$
Krit. tartomány:	$\mathcal{X}_k = \{(\mathrm{x},\mathrm{y}) : t > t_{f,\alpha}\}$	$\mathcal{X}_k = \{(\mathrm{x},\mathrm{y}) : t < -t_{f,\alpha}\}$

Áttekintés

χ^2-**próba** (normális eloszlás szórására)

$X_1, \ldots, X_n \sim N(m, \sigma^2)$, ahol m és σ ismeretlen paraméterek

Kétoldali: $H_0 : \sigma = \sigma_0$
$\underline{H_1 : \sigma \neq \sigma_0}$

Próbastatisztika: $T(\mathbf{X}) = h := \frac{(n-1)(S_n^*)^2}{\sigma_0^2} \overset{H_0 \text{ esetén}}{\sim} \chi^2_{n-1}$

Kritikus tartomány: $\mathcal{X}_k = \left\{ \mathrm{x} : h < \chi^2_{n-1,\alpha/2} \text{ vagy } h > \chi^2_{n-1,1-\alpha/2} \right\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak:	$H_1 : \sigma > \sigma_0$	$H_1 : \sigma < \sigma_0$
Kritikus tartomány:	$\mathcal{X}_k = \{\mathrm{x} : h > \chi^2_{n-1,1-\alpha}\}$	$\mathcal{X}_k = \{\mathrm{x} : h < \chi^2_{n-1,\alpha}\}$

Áttekintés

F-próba

$X_1, \dots, X_n \sim N(m_1, \sigma_1^2)$
$Y_1, \dots, Y_m \sim N(m_2, \sigma_2^2)$ független minták
ahol m_1, m_2, σ_1, σ_2 ismeretlen paraméterek

Kétoldali: $H_0 : \sigma_1 = \sigma_2$
$\underline{H_1 : \sigma_1 \neq \sigma_2}$

Próbastatisztika: $T(\mathbf{X}, \mathbf{Y}) = F = \frac{(S_1^*)^2}{(S_2^*)^2} \overset{H_0 \text{ esetén}}{\sim} F_{n-1,m-1}$

Kritikus tartomány:

$\mathcal{X}_k = \{(x, y) : F < F_{n-1,m-1,\alpha/2} \text{ vagy } F > F_{n-1,m-1,1-\alpha/2}\}$

Egyoldali próbák esetén H_0 és a próbastatisztika ugyanaz marad, csak H_1, és ezáltal a kritikus tartomány változik.

Egyoldaliak:	$H_1 : \sigma_1 > \sigma_2$	$H_1 : \sigma_1 < \sigma_2$
Krit. tart.:	$\{(x, y) : F > F_{n-1,m-1,1-\alpha}\}$	$\{(x, y) : F < F_{n-1,m-1,\alpha}\}$

Áttekintés

A χ^2-próba

Legyen $A_1, \ldots, A_r$ teljes eseményrendszer.
Végezzünk n darab független megfigyelést, jelölje az i-edik esemény bekövetkezési gyakoriságát N_i $(i = 1, \ldots, r)$. A megfigyelések egyes eredményei segítségével definiálható az X_i valószínűségi változó, ami vegyen fel olyan értéket, amelyik számú esemény a teljes eseményrendszerből bekövetkezett. Ezáltal formálisan $N_i = \sum_{j=1}^{n} I(X_j = i)$ és $\sum_{i=1}^{r} N_i = n$

H_0: $P(A_i) = p_i$, $i = 1, \ldots, r$ $\rightsquigarrow$ tfh. $p_i > 0\ \forall i$, $p_1 + \ldots + p_r = 1$
H_1: a nullhipotézis tagadása

Próbastatisztika: $T_n(\mathbf{X}) := \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{np_i} \xrightarrow[n\to\infty]{H_0 \text{ esetén}} \chi^2_{r-1}$ eloszlásban
Kritikus tartomány: $\mathcal{X}_k = \left\{ \mathrm{x} : T_n(\mathbf{X}) > \chi^2_{r-1;1-\alpha} \right\}$