STATISZTIKAI PRÓBÁK

(VALÓSZÍNŰSÉGI) VÁLTOZÓ

Absztrakcióval nyert fogalom
pl. SALARY

STATISZTIKAI PRÓBÁK

STATISZTIKAI MINTA

Absztrakcióval nyert fogalom:
A változóval azonos eloszlású, független változók halmaza

A változóra vonatkozó megfigyeléssorozat
pl. az adatmátrix egy oszlopa

STATISZTIKAI PRÓBÁK

MINTAREALIZÁCIÓ

A változóra vonatkozó megfigyeléssorozat
pl. az adatmátrix egy oszlopa

STATISZTIKAI PRÓBÁK

NULLHIPOTÉZIS (H_0 :)

Logikai állítás egy vagy több változóra vonatkozóan
pl. SALARY normális eloszlású

STATISZTIKAI PRÓBÁK

ALTERNATÍV HIPOTÉZIS (H_1 :)

Egy a H_0 -val ellentétes állítás, legtöbbször H_0 logikai tagadása
pl. SALARY nem normális eloszlású

STATISZTIKAI PRÓBÁK

ELŐZETES FELTEVÉSEK

A vizsgált változókra eleve elfogadott tulajdonságok
pl. SALARY normális eloszlású

STATISZTIKAI PRÓBÁK

PRÓBASTATISZTIKA

A mintából számolt változó, melynek eloszlása a H_0 feltételezése mellett pontosan (vagy legalább aszimptotikusan) megadható

STATISZTIKAI PRÓBÁK

SZÁMÍTOTT ÉRTÉK

A próbastatisztika felvett értéke az adott mintarealizáció mellett

STATISZTIKAI PRÓBÁK

KRITIKUS ÉRTÉK

A felhasználó által beállított korlát, amelytől a próbastatisztika igen nagy valószínűséggel kisebb értéket kell, hogy felvegyen, ha H_0 igaz$_{igaz}$

STATISZTIKAI PRÓBÁK

SZIGNIFIKANCIA-SZINT

Olyan 0 és 1 közé eső érték, amely a nullhipotézis elfogadhatóságának mértékét adja meg. Minél közelebb van értéke az 1-hez, annál bizonyosabb a nullhipotézis teljesedése. Megegyezik az elsőfajú hiba-valószínűséggel

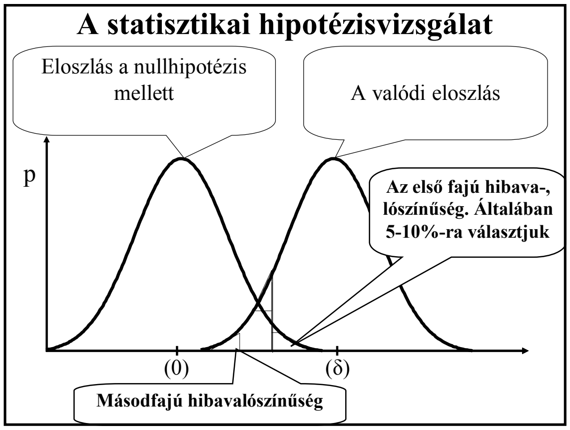
STATISZTIKAI PRÓBÁK

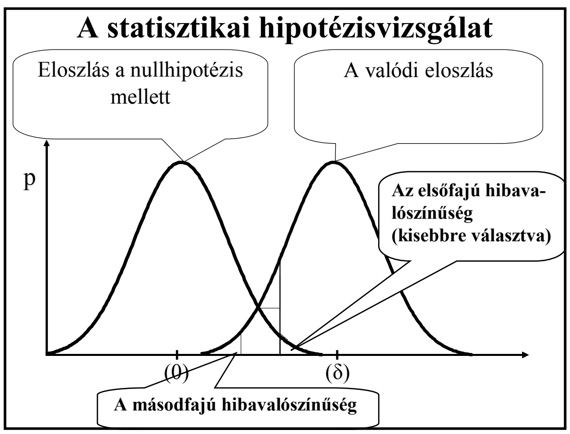
STATISZTIKAI DÖNTÉS

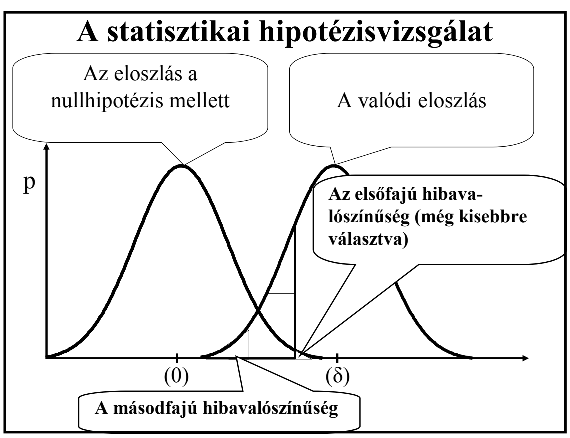
H_0 -t elfogadjuk, ha a számított érték kisebb a kritikus értéknél. Ellenkező esetben H_1 -t fogadjuk el

STATISZTIKAI DÖNTÉS

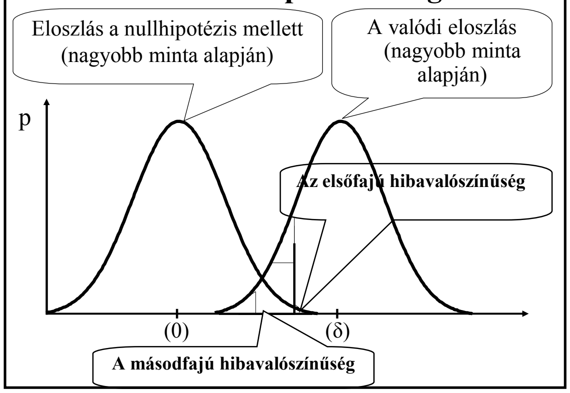
Valóság / Döntés	H_0 IGAZ	H_1 IGAZ
H_0-at Elfogad-juk	**HELYES DÖNTÉS**	**MÁSODFAJÚ HIBA**
H_1-et Fogad-juk el	**ELSŐFAJÚ HIBA**	**HELYES DÖNTÉS**


A statisztikai hipotézisvizsgálat
Eloszlás a nullhipotézis mellett
A valódi eloszlás
p
Az első fajú hibava-, lószínűség. Általában 5-10%-ra választjuk
(0)
(δ)
Másodfajú hibavalószínűség

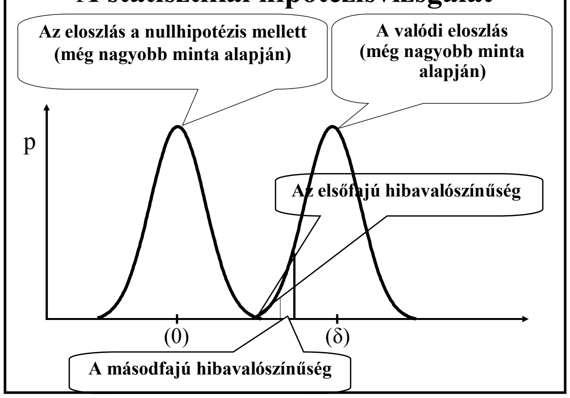

A statisztikai hipotézisvizsgálat
Eloszlás a nullhipotézis mellett
A valódi eloszlás
p
Az elsőfajú hibava- lószínűség (kisebbre választva)
(0)
(δ)
A másodfajú hibavalószínűség


A statisztikai hipotézisvizsgálat
Az eloszlás a nullhipotézis mellett
A valódi eloszlás
p
Az elsőfajú hibava- lószínűség (még kisebbre választva)
(0)
(δ)
A másodfajú hibavalószínűség

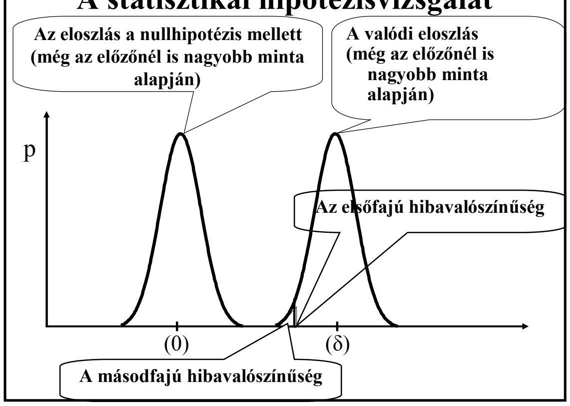
A statisztikai hipotézisvizsgálat



A statisztikai hipotézisvizsgálat



A statisztikai hipotézisvizsgálat



PARAMÉTERES PRÓBÁK

Az elemzett változó eloszlását ismertnek tételezzük fel, és a hipotézisünket az eloszlás paramétereire fogalmazzuk meg. Az SPSS-ben a normális eloszlás várható értékére és a varianciájára vonatkozó alábbi paraméteres próbák hajthatók végre:

várható értékre	**varianciára**
***t*-próbák**	**Bartlett-Box**
Egyszeres osztályozás	***F*-próba**

PARAMÉTERES PRÓBÁK

t-PRÓBÁK

A próbastatisztika Student- (t-) eloszlású, ha a nullhipotézis igaz

Egymintás t-próba

H_0 : A változó várható értéke megegyezik a tesztértékkel

Két független mintás t-próba

H_0 : A két változó várható értéke egyenlő

Egy változó eseteit két csoportba csoportosítjuk egy tördelő változó segítségével

Meg kell különböztetni azt a két alesetet, amikor a minták varianciái egyenlőek attól, amikor különböznek !

Két összetartozó mintás t-próba

H_0 : A két változó várható értéke egyenlő

PARAMÉTERES PRÓBÁK

Egymintás t-próba

$$H_0 : \text{A minta várható értéke } m_0$$

$$H_0 \Rightarrow t_{próba} = \frac{\bar{x}_n - m_0}{s_n^*}\sqrt{n} \in t_{n-1}$$

$$Prob(t_{n-1} < t_{krit}) = 1 - \varepsilon$$

DÖNTÉS: $$\left|t_{próba}\right| < t_{krit} \Rightarrow H_0$$

PARAMÉTERES PRÓBÁK

Kétmintás *t*-próba (független minták)

H_0 : A minták várható értékei egyenlőek

A minták szórásai egyenlőeknek tekintendők. Különben nem alkalmazható a próba. Ennek ellenőrzése *F*-próbával.

$$H_0 \Rightarrow t_{próba} = \frac{\bar{x}_n - \bar{y}_m}{\sqrt{(n-1)\left(s_{x,n}^*\right)^2 + (m-1)\left(s_{y,m}^*\right)^2}} \sqrt{\frac{nm(n+m-2)}{n+m}} \in t_{n+m-2}$$

$$Prob(t_{n+m-2} < t_{krit}) = 1 - \varepsilon$$

DÖNTÉS: $\left|t_{próba}\right| < t_{krit} \Rightarrow H_0$

PARAMÉTERES PRÓBÁK

Kétmintás *t*-próba (összetartozó minták)

H_0 : A minták várható értékei egyenlőek

$$H_0 \Rightarrow t_{próba} = \frac{\bar{x}_n - \bar{y}_n}{\sqrt{\left(s_{x,n}^*\right)^2 + \left(s_{y,n}^*\right)^2}} \sqrt{n} \in t_{2n-2}$$

$$Prob(t_{2n-2} < t_{krit}) = 1 - \varepsilon$$

DÖNTÉS: $\left|t_{próba}\right| < t_{krit} \Rightarrow H_0$

PARAMÉTERES PRÓBÁK

Egyszeres osztályozás (One-way ANOVA)

H_0 : A változók várható értékei azonosak

H_1 : Van két olyan változó, melyek várható értékei különböznek

Az egyes csoportok várható értékei különbségére konfidencia intervallum szerkeszthető. A különbség valódi értéke a beállított valószínűséggel ebbe az intervallumba esik.

A mintarealizációt egy tördelő változó értékei szerint kettőnél több csoportra osztjuk. A módszer a független mintás t-próba kiterjesztése kettőnél nagyobb esetre.

PARAMÉTERES PRÓBÁK

Fisher-féle *F*-próba

H_0 : A változók varianciái egyenlőek

A próbastatisztika eloszlása a nullhipotézis fennállása esetén *F*-eloszlású.

PARAMÉTERES PRÓBÁK

Bartlett-Box-próba

H_0 : A kettőnél több változó varianciái egyenlőek

A próbastatisztika eloszlása a nullhipotézis fennállása esetén *F*-eloszlású. Az F-próba kiterjesztése.

PARAMÉTERES PRÓBÁK

Levene-próba

H_0 : A kettőnél több változó varianciái egyenlőek

Ez nem paraméteres próba! Nincs előzetes feltevés a változók normalitására vonatkozóan!

NEMPARAMÉTERES PRÓBÁK

• **ILLESZKEDÉSVIZSGÁLAT**

H_0 : Az elemzett változó eloszlása megegyezik a hipotetikussal

χ^2-próba, egymintás Kolmogorov-Szmirnov, P-P grafikon

• **FÜGGETLENSÉVIZSGÁLAT**

H_0 : Az elemzett változók függetlenek

χ^2-próba, nominális változókra, ordinális változókra

• **HOMOGENITÁSVIZSGÁLAT**

H_0 : Az elemzett változók eloszlása azonos

χ^2-próba, kétmintás Kolmogorov-Szmirnov, Wilcoxon, McNemar, Kruskal-Wallis, Friedmann

NEMPARAMÉTERES PRÓBÁK

Egymintás Kolmogorov-Szmirnov próba

H_0 : A minta eloszlásfüggvénye $F(x)$

Most illeszkedésvizsgálatról van szó!

$$t_{próba} = \sqrt{n} \sup_{x \in IR} \left| F_{emp}(x) - F(x) \right|$$

$F_{emp}(x) = \frac{k}{n}$ **ahol** $k = \#\{x_i;\, x_i < x\}$ vagy $x_k^* < x \le x_{k+1}^*$

Ha a nullhipotézis igaz, a próbastatisztika aszimptotikusan Kolmogorov-eloszlást követ. A kritikus értéket ez alapján az eloszlás alapján határozzuk meg a szignifikancia szinthez.

DÖNTÉS $t_{próba} < t_{krit} \Rightarrow H_0$

NEMPARAMÉTERES PRÓBÁK

Kétmintás Kolmogorov-Szmirnov próba

H_0 : A minták eloszlásfüggvénye azonos

Most homogenitásvizsgálatról van szó!

$$t_{próba} = \sqrt{\frac{n+m}{2}} \sup_{x \in IR} \left| F_{emp}(x) - G_{emp}(x) \right|$$

Ha a nullhipotézis igaz, a próbastatisztika most is aszimptotikusan Kolmogorov-eloszlást követ. A kritikus értéket ez alapján az eloszlás alapján határozzuk meg a szignifikancia szinthez.

DÖNTÉS $t_{próba} < t_{krit} \Rightarrow H_0$

PARAMÉTERES PRÓBÁK

Egymintás t-próba (one sample t-test)

Egy adott normálisnak tekintett minta várható értéke a beállított hipotetikus érték-e?

Két független mintás t-próba (independent-samples t-test)

Két adott normálisnak tekintett minta várható értékei Egyenlőknek tekinthetők-e? (A szórások egyeznek…)

Párosított mintás t-próba (paired-samples t-test)

Két adott normálisnak tekintett összetartozó minta várható értékei egyenlőnek tekinthetők-e?

Egyszerű csoportosítás (one-way ANOVA)

Több normálisnak tekintett minta várható értékei egyenlőnek tekinthetők-e?

NEMPARAMÉTERES PRÓBÁK

Chi-négyzet próba (Chi-square)

Diszkrét változók illeszkedését ellenőrizhetjük

Binomiális próba (Binomial)

Dichotóm változók illeszkedését ellenőrizhetjük

Véletlenszerűségi próba (Runs)

Dichotóm változók értékeinek véletlenszerűségét ellenőrizhetjük

Egymintás Kolmogorov-Szmirnov (1-sample K-S)

Diszkrét és folytonos változók illeszkedését tesztelhetjük

NEMPARAMÉTERES PRÓBÁK

Kétmintás homogenitás (two-independent samples test)
Mann-Whitney, kétmintás K-S, Moses, Wald-Wolfowitz

Független minták azonos eloszlását ellenőrizhetjük

Többmintás homogenitás (k-independent samples test)
Kruskal-Wallis H, Median, Jonckheere-Terpstra

Több független minta azonos eloszlását ellenőrizhetjük

Két összetartozó minta homogenitása (2-related samples test)
Wilcoxon, Sign, McNemar, Marginal Homogenity

Két összetartozó minta azonos eloszlását ellenőrizhetjük

Több összetartozó minta homogenitása (k-related samples test)
Friedman, Kendall W, Cohran Q

Több összetartozó minta azonos eloszlását ellenőrizhetjük

A próba neve	H_0	A próbastatisztika	Eloszlás ha H_0 igaz
Egymintás t	$\mathbf{E}X=m_0$	$\frac{\overline{X}-\mu_0}{s_n^*}\sqrt{n}$	t_{n-1}
Független kétmintás t	$\mathbf{E}X_i=\mathbf{E}X_j$	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{(n_1-1)s_1^{*2}+(n_2-1)s_2^{*2}}}\sqrt{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}}$	$t_{n_1+n_2-2}$
Összetartozó kétmintás t	$\mathbf{E}X_i=\mathbf{E}X_j$	$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{(n-1)s_1^{*2}+(n-1)s_2^{*2}}}\sqrt{n(n-1)}$	t_{2n-2}
Egyszerű csoportosítás	$\mathbf{E}X_i=\mathbf{E}X_j=\ldots=\mathbf{E}X_m$	$Q_k=\sum_{i=1}^{L} n_i\left(\overline{x}^{(i)}-\overline{x}\right)^2,\ Q_b=\sum_{i=1}^{L}\sum_{j=1}^{n_i}(x_j^{(i)}-\overline{x}^{(i)})^2,\ \frac{\frac{Q_k}{L-1}}{\frac{Q_b}{n-L}}$	$F_{L-1,n-L}$
F - próba	$\mathbf{D}^2X_i=\mathbf{D}^2X_j$	$\frac{s_{n_1}^{*2}}{s_{n_2}^{*2}}$	F_{n_1-1,n_2-1}
Bartlett-próba	$\mathbf{D}^2X_i=\mathbf{D}^2X_j=\ldots=\mathbf{D}^2X_n$	$f_i=n_i-1,\ f=\sum_{i=1}^{p} n_i-p=n-p,\ c=1+\frac{1}{3(p-1)}\cdot\left(\sum_{i=1}^{p}\frac{1}{f_i}-\frac{1}{f}\right)$ $z=\frac{1}{f}\cdot\sum_{i=1}^{p} f_i\cdot z_i\ \ B=\frac{2.3026}{c}\cdot\left(f\cdot\log z-\sum_{i=1}^{p} f_i\cdot\log z_i\right)$	χ_{p-1}^2

A próba neve	H_0	A próbastatisztika	Eloszlás ha H_0 igaz
χ^2 illeszkedés	$P(X<t)=F_0(t)$	$\sum_{i=1}^{r}\frac{\upsilon_i^2}{np_i}-n$	χ_{r-1}^2
Egymintás Kolmogorov-Szmirnov	$P(X<t)=F_0(t)$	$\sqrt{n}\cdot\sup_{x\in IR}\lvert F_n(x)-F_0(x)\rvert$	Kolmogorov
χ^2 függetlenség	$P(X_1<t,X_2<v)=P(X_1<t)P(X_2<v)$	$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(\upsilon_{ij}-n\cdot p_{i\cdot}\cdot p_{\cdot j})^2}{n\cdot p_{i\cdot}\cdot p_{\cdot j}}$	$\chi_{(r-1)(s-1)}^2$
χ^2 homogenitás	$P(X_1<t)=P(X_2<t)$	$n_1\cdot n_2\sum_{i=1}^{r}\frac{\left(\frac{\upsilon_i}{n_1}-\frac{\lambda_i}{n_2}\right)^2}{\upsilon_i+\lambda_i}$	χ_{r-1}^2
Kétmintás Kolmogorov-Szmirnov	$P(X_1<t)=P(X_2<t)$	$\sqrt{\frac{n_1\cdot n_2}{n_1+n_2}}\sup_{x\in R}\lvert F_{n_1}(x)-F_{n_2}(x)\rvert$	Kolmogorov