

Hipotézis vizsgálatok

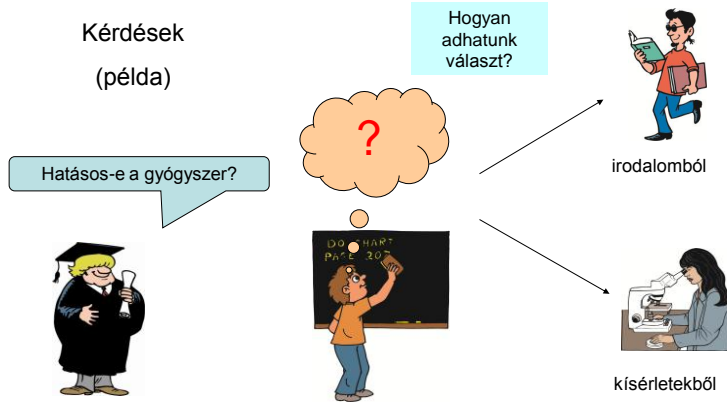
Kérdések
(példa)

Hatásos-e a gyógyszer?

Hogyan
adhatunk
választ?

irodalomból

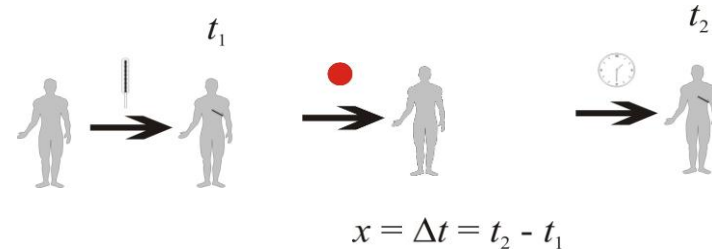
kísérletekből



Egy példa

Kérdés: Hatásos a lázcsillapító gyógyszer?

kísérlet



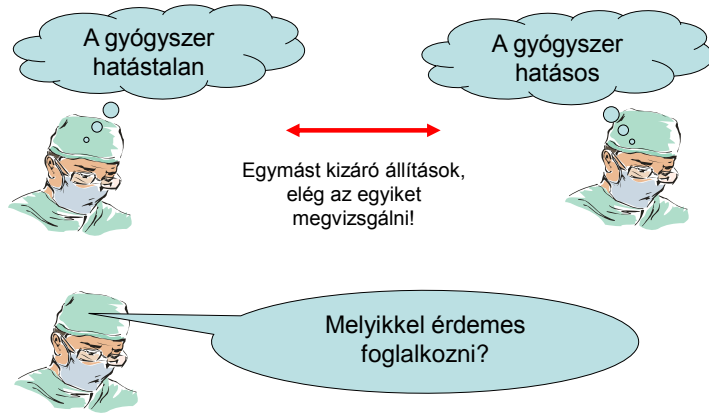
Hipotézisek

A gyógyszer
hatástalan

A gyógyszer
hatásos

Egymást kizáró állítások,
elég az egyiket
megvizsgálni!

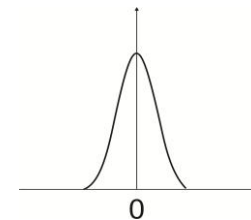
Melyikkel érdemes
foglalkozni?



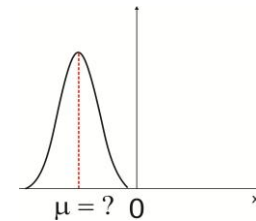
A megfigyelt változó eloszlása

A gyógyszer
hatástalan

A gyógyszer
hatásos



A véletlen hatások
eredője 0.



Mekkora a hatás?



Ha a populációt megismerhetnénk!!!

Eredmény

$$\mu = 0$$



Következtetés

A gyógyszer hatástalan.

$$\mu < 0$$



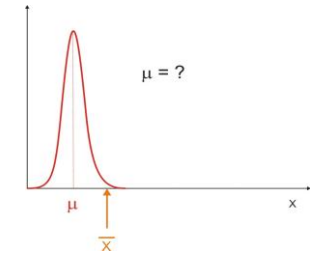
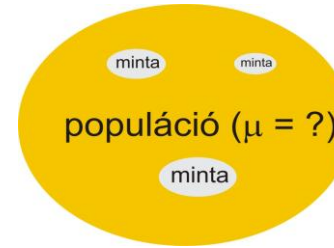
A gyógyszer hatásos, a hatás mértékére a μ jellemző.

A helyzet „fokozódik”

A populáció általában nem ismert.

A minta nem azonos a populációval!

pl. az átlagok ingadoznak a várható érték körül!



Mi az oka az eltérésnek?



Mintavételezés véletlen ingadozás.
(A feltevésünk helyes!)



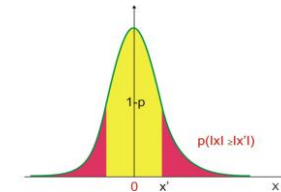
Az alapfeltevésünk (hipotézisünk) nem igaz (tévedtünk!). Az eltérés nem véletlen.



Mi alapján dönthetünk?

Mekkora az esélye, hogy a minta valóban az adott populációból származik?

Ehhez ismert paraméterű eloszlás szükséges!



Nullhipotézis: (H_0)

a minta/minták eltérése a választott populáció(k)tól a mintavételből származó **véletlen eltérés**. Gyakran egy tagadó válasz a feltett kérdésre. (példa: a gyógyszer nem hatásos.)



Alternatív hipotézis: (H_1)

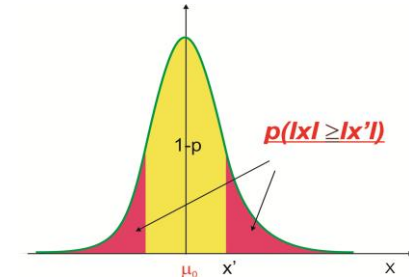
a minta/minták eltérése a választott populáció(k)tól **nem véletlen**. (példa: a gyógyszer hatásos)

Nullhipotézis

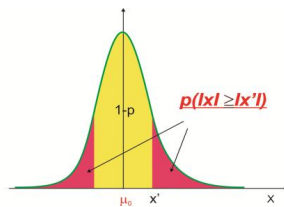


Ismert eloszlás esetében megadható!

(Az eloszlás alakja nem mindig ilyen, de ismert!)



Szignifikáns?



p annak a valószínűsége, hogy az eltérés **véletlen**!

Ha p elég nagy, lehet véletlen, ha p elég kicsi a különbséget **szignifikánsnak** tekintjük!



Szignifikancia szint



Válasszunk egy értéket, amelyet határnak tekintünk! Ez a szignifikancia szint.

Jelölése: α .
Orvosi gyakorlatban értéke igen gyakran 5%.



A döntés alapja

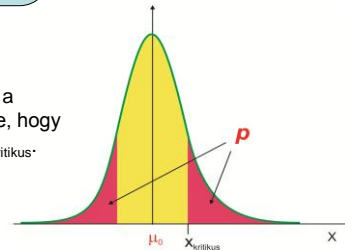
Ha a p elég kicsi, nagyobb az esélye, hogy a nullhipotézis nem igaz. Azaz inkább az alternatív hipotézis a valószínűbb.



p annak a valószínűsége, hogy $X_{\text{számolt}} \geq X_{\text{kritikus}}$.

X_{kritikus} : a szignifikancia szinthez tartozó érték

$X_{\text{számolt}}$: a mintá(k)ból számolt érték



A döntés

1. Ha a véletlen eltérés valószínűsége kicsi ($p(|x| \geq |x_{\text{krit}}|) \leq 5\%$) – **elvetjük** a nullhipotézist.
2. Ha a véletlen eltérés valószínűsége nagy ($p(|x| \geq |x_{\text{krit}}|) > 5\%$) – **megtartjuk** a nullhipotézist.

A válasz sohasem igen - nem, vagy igaz - hamis!!!

A döntés „jósága”

döntés:
a nullhipotézist

megtartjuk

elvetjük

igaz

Helyes döntés

I. Típusú hiba (α)

tény:
a nullhipotézis

hamis

II. Típusú hiba (β)

Helyes döntés

Vizsgálat egy csoportban: (egymintás t-próba)

Kérdés: A minta alapján lehet-e a populáció jellemző értéke egy megadott érték?

A példa: Hatásos-e a lázcsillapító vagy sem?

Nullhipotézis: nem! $\mu_0 = 0$. De az átlag nem 0!

| minta | átlag |
|-------|---------|
| 1. | -0,2 °C |
| 2. | -1 °C |
| 3. | -1,5 °C |

Ha az eltérés nagyobb, biztosabbnak tűnik az alternatív hipotézis (a gyógyszer hatásos)

Mit jelent a nagy eltérés?

Mi a mértéke az eltérésnek?

Standard hiba: az átlagok átlagos eltérése a μ -tól.

$(\bar{x} \pm s_{\bar{x}})$

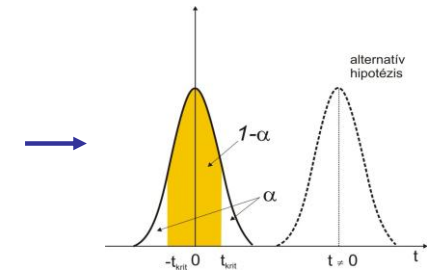
~ 68% - konfidencia intervallum.

A t -érték

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

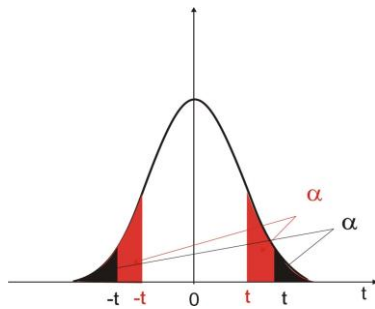
Viszonyítsuk az eltérést a standard hibához!
(μ_0 igen gyakran = 0)

Mivel az átlagok a μ_0 körül ingadoznak, a t -értékek a 0 körül.
(feltéve, hogy a nullhipotézis igaz!)



Miért alkalmasabb a t -érték?

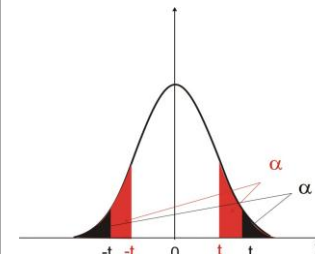
Képesek vagyunk kiszámolni ennek az eltérésnek a valószínűségét!!! (Student- vagy t -eloszlás)



Csak a t -értékek véletlen ingadozását írja le!

Az eloszlás alakja függ az elemszámtól.

A t -táblázat



t -táblázat

| d.f. | szignifikancia szint | | | |
|------|----------------------|------|------|------|
| | 0.1 | 0.05 | 0.02 | 0.01 |
| 1 | 6.31 | 12.7 | 31.8 | 63.7 |
| 2 | 2.92 | 4.3 | 6.96 | 9.92 |
| 3 | 2.35 | 3.18 | 4.54 | 5.84 |
| 4 | 2.13 | 2.78 | 3.75 | 4.6 |
| 5 | 2.02 | 2.57 | 3.37 | 4.03 |

Különböző t_{krit} értékek tartoznak a különböző valószínűség értékekhez.

szabadsági fok: (n-1)

A szabadsági fok

Gondoltam 3 számra! (minta)



3, 12, 8 vagy 5, 7, 11 stb.

A szabadsági fok = n

A 3 szám átlaga: 8! (információ!)



3, 12, 9 vagy 5, 7, 12 stb.

A szabadsági fok = $n-1$

Döntés t -táblázat alapján

t -táblázat

| d.f. | szignifikancia szint | | | |
|------|----------------------|------|------|------|
| | 0.1 | 0.05 | 0.02 | 0.01 |
| 1 | 6.31 | 12.7 | 31.8 | 63.7 |
| 2 | 2.92 | 4.3 | 6.96 | 9.92 |
| 3 | 2.35 | 3.18 | 4.54 | 5.84 |
| 4 | 2.13 | 2.78 | 3.75 | 4.6 |
| 5 | 2.02 | 2.57 | 3.37 | 4.03 |

szabadsági fok: $(n-1)$

Kiválasztunk egy alkalmas szignifikancia szintet!

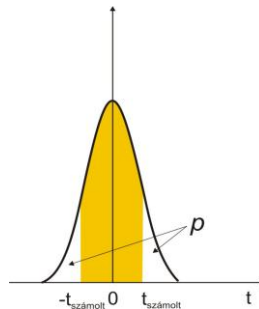


Döntés számítógép segítségével

Én tudok integrálni!!!



p : annak a valószínűsége, hogy véletlenül ilyen nagy a $t_{\text{számolt}}$



A döntés

- 1. Ha a véletlen eltérés valószínűsége kicsi ($p(|t| \geq t_{\text{krit}}) \leq 5\%$) – **elvetjük** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p(|t| \geq t_{\text{krit}}) > 5\%$) – **megtartjuk** a nullhipotézist.

Az egymintás t-próba feltétele

- A feladat: egy minta alapján döntés a μ értékéről.
- A változó **normális eloszlású** legyen.



Vizsgálat két csoportban

Kérdés: A két minta származhat-e azonos populációból, vagy a két populáció paraméterei azonosak?

$$\mu_1 = \mu_2 ?$$

Nullhipotézis: $\mu_1 = \mu_2$

(általában $\bar{x}_1 \neq \bar{x}_2$)

kétmintás t-próba

kétmintás t-próba

$$\bar{x}_1 \neq \bar{x}_2$$



Ismert eloszlású változóra van szükség!



$$t = \frac{\bar{x}_1 - \bar{x}_2}{s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

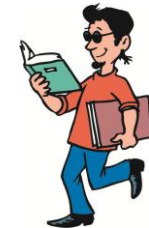
$$s^* = \sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 2}}$$

A próba

A t-érték az t-érték!



Akkor meg tudom csinálni!
Pardon, mennyi a szabadsági fokok száma?



$$\text{sz.f.} = n_1 + n_2 - 2 \\ ((n_1 - 1) + (n_2 - 1))$$

A kétmintás t-próba feltétele

- A feladat: két egymástól **független** csoport összehasonlítása.
- A változó **normális eloszlású** legyen.
- A **szórás** a két csoportban **azonos**nak tekinthető.



Ez utóbbi új!
Hogyan állapítható meg?

A szórások vizsgálata

Hogyan fogjunk hozzá?

Nullhipotézis: a két szórás azonos, az eltérés véletlen (mintavétel).



De hiszen ez olyan, mint egy hipotézis-vizsgálat!

Az F-próba

$$F = \frac{s_1^2}{s_2^2}$$

A nullhipotézishez tartozik egy ún. F-eloszlás.

De melyik variancia legyen a számlálóban?



A számlálóban mindig a nagyobb variancia van! ($F \geq 1$)



Döntés

- 1. Ha a véletlen eltérés valószínűsége kicsi ($p(F \geq F_{\text{krit}}) \leq 5\%$) – **elvetjük** a nullhipotézist.
- 2. Ha a véletlen eltérés valószínűsége nagy ($p(F \geq F_{\text{krit}}) > 5\%$) – **megtartjuk** a nullhipotézist.

2 vagy több változó

Korreláció és regresszió

Kapcsolat két változó között.

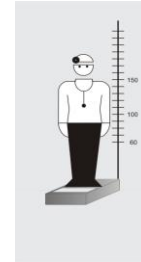
Függvényszerű leírás.

Korreláció

Példa:

Van-e kapcsolat a testsúly és a testmagasság között?

kísérlet:



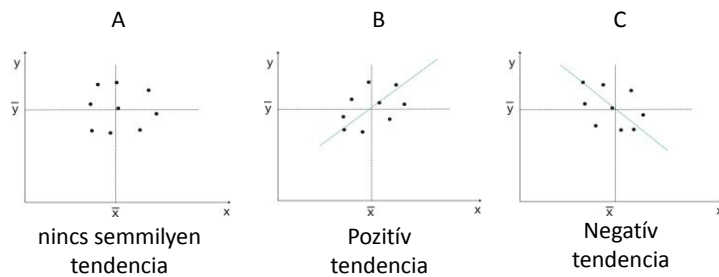
adatok:

| n | magasság (cm) | súly (kg) |
|---|---------------|-----------|
| 1 | 150 | 61 |
| 2 | 170 | 70 |
| 3 | 166 | 75 |
| 4 | 174 | 70 |
| 5 | 180 | 72 |
| 6 | 155 | 50 |
| 7 | 172 | 65 |
| 8 | 161 | 59 |
| 9 | 177 | 81 |

Ábrázolás

például: x a magasság és y a súly.

lehetséges esetek:



Pearson-féle korrelációs együttható

$$r = \frac{\text{cov}(x, y)}{s_x \cdot s_y} = \frac{Q_{xy}}{\sqrt{Q_x \cdot Q_y}}$$

Az r lehetséges értékei:

$$-1 \leq r \leq 1$$

$$Q_{xy} = \sum_i [(x_i - \bar{x})(y_i - \bar{y})]$$

$$Q_x = \sum_i (x_i - \bar{x})^2$$

$$Q_y = \sum_i (y_i - \bar{y})^2$$

A populációban:

$r = 0$ nincs korreláció,

$r \neq 0$ van! (mértéke arányos az r abszolút értékével.)

Determinációs együttható

$$r^2$$

Megadja, hogy milyen erős a kapcsolat.
Az y változásainak mekkora része értelmezhető az x váltoásaival.

Korrelációs t-teszt

A számolt r csak becslése az r populációbeli értékének. A számolt érték az elmélet r körül ingadozik.
(pl. $r_{\text{számolt}} = 0,1$?)

$$H_0: r = 0! \longrightarrow t = r \sqrt{\frac{n-2}{1-r^2}} \longrightarrow \text{sz.f.: } n-2$$

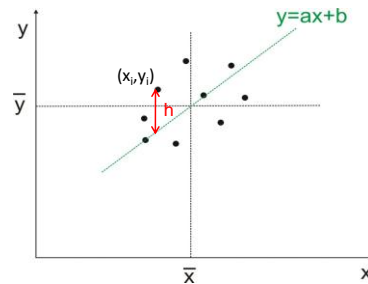
Döntés: a t-érték alapján. Lásd előző példákat!

Feltétele: Legalább az egyik változó normális eloszlású.

Lineáris regresszió

Ha a változók normális eloszlásúak, a kapcsolat közöttük lineáris jellegű.

$$y_i = ax_i + b + h_i$$



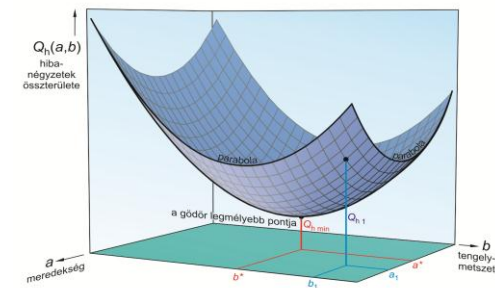
y : függő változó
 x : független változó
 h_i : hibahatás = $y_i - (ax_i + b)$.

(A különbség a megfigyelt és a feltételezett érték között)

A legkisebb négyzetek módszere

$$Q_h = \sum_i h_i^2 = \sum_i (y_i - (ax_i + b))^2$$

x_i és y_i mért értékek.
 a és b az ismeretlen!



Melyik a legjobban illeszkedő egyenes?

Q_h minimális! $\rightarrow a^* = \frac{Q_{xy}}{Q_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$ $b^* = \bar{y} - a^* \bar{x}$

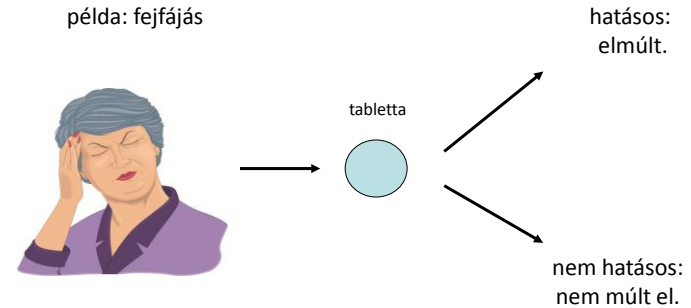
Kapcsolat az inzulin érzékenység és a BMI között.

r^2 : determinációs koefficiens.

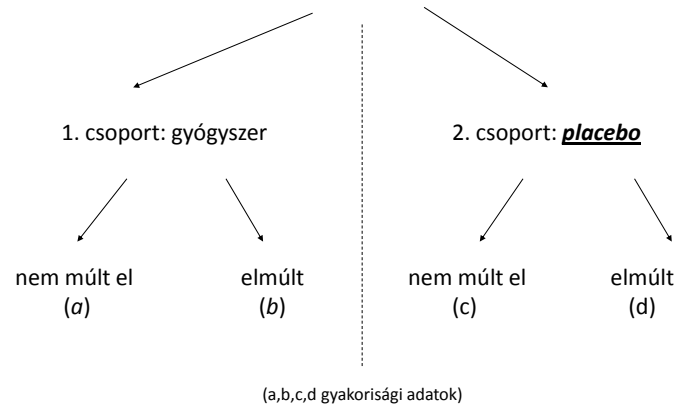
| független | regressziós együttható | st. hiba | t | p | döntés |
|-----------|------------------------|----------|-------|--------|--------------|
| BMI | -0,077 | 0,018 | -4,25 | 0,0011 | szignifikáns |
| r^2 | 0,6 | | | | |

Khi-négyzet teszt (gyakorisági adatok elemzése)

példa: fejfájás



Kísérlet



Kontingencia tábla

| | Nem múlt el | elmúlt | Összes |
|------------|-------------|--------|--------|
| 1. csoport | a | b | a+b |
| 2. csoport | c | d | c+d |
| összes | a+c | b+d | n |

2 x 2 tábla.

Nullhipotézis

Ha a hatás független a gyógyszertől,
azt várjuk, hogy:

$$\frac{a}{b} = \frac{c}{d} \longrightarrow a \times d = b \times c$$

Nullhipotézis: a hatás független a gyógyszertől,
csupán placebo hatás.

khi-négyzet teszt (függetlenség).

χ^2 -eloszlás

Képlet 2 x2 táblához:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Nullhipotézis: $\chi^2 = 0$, a különbség csupán mintavételi hiba.

χ^2 -eloszlás: megadja a χ^2 -érték véletlen eltéréseit.

Döntés

Hasonló a t-eloszlás esetében
megbeszéltekhez. A különbség: a χ^2 -eloszlást
használjuk.

A várható érték = 0, ha a nullhipotézis igaz.

ha $\chi^2_{\text{számolt}} \geq \chi^2_{\text{krit}}$ - elvetjük ellenkező esetben megtartjuk a nullhipotézist.

vagy $p(\chi^2 \geq \chi^2_{\text{számolt}}) \leq 5\%$ - elvetjük ellenkező esetben megtartjuk a nullhipotézist.

szabadsági fokok száma: ebben a speciális esetben = 1.

általában:

sz.f. = $(s-1)(o-1)$, ahol s – a sorok száma
 o – az oszlopok száma