Biofizika I.
fogorvostan hallgatóknak

2. előadás:

Biostatisztika II.

2019. Szeptember 16.

Veres Dániel

Ismétlés: Alapsokaság és minta

(Következtető statisztika alap problémája)

Alapsokaság (populáció)

Az **alapsokaság** rendszerint olyan méretű, hogy az összes eleme nem vizsgálható meg.

„VÉLETLEN"

Minta

Emiatt az alapsokaságnak csak egy részhalmazát vizsgáljuk, ezt nevezzük mintának.

A minta elemein méréseket végzünk, majd az így keletkező adathalmazt (amit szintén mintának nevezünk) grafikusan és matematikailag jellemezzük

A minta jellemzői alapján az alapsokaságra vonatkozó következtetést vonhatunk le

BIZONYTALANSÁG!
(P és H)

Becslés

Alapsokaság (populáció)	Minta
Valódi érték	Becslés
Becslés	
Valószínűség	Relatív gyakoriság
Várható érték (populációs átlag)	Minta átlaga
Elméleti variancia (szórásnégyzet)	Minta variancia
Két várható érték különbsége	Mintaátlagok különbsége

Becslés

Alapsokaság (populáció) ← Minta

Valódi érték

Becslés

Becslés

Valószínűség

Relatív gyakoriság

Várható érték (populációs átlag)

Minta átlaga

Minta mediánja?

Elméleti variancia (szórásnégyzet)

Minta variancia - melyik?

$$v1 = \frac{\sum_{i=1}^{n} \left(\overline{x} - x_i\right)^2}{n}$$

$$v1 = \frac{\sum_{i=1}^{n} \left(\overline{x} - x_i\right)^2}{n-1}$$

...

Hiba

Alapsokaság (populáció)

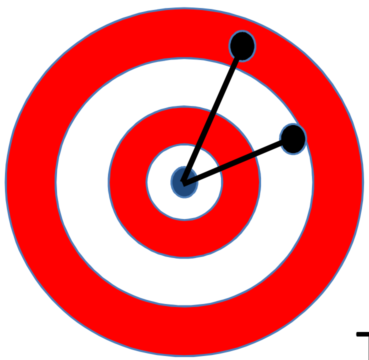
- Valódi érték

Minta

- Becslések

Becslés

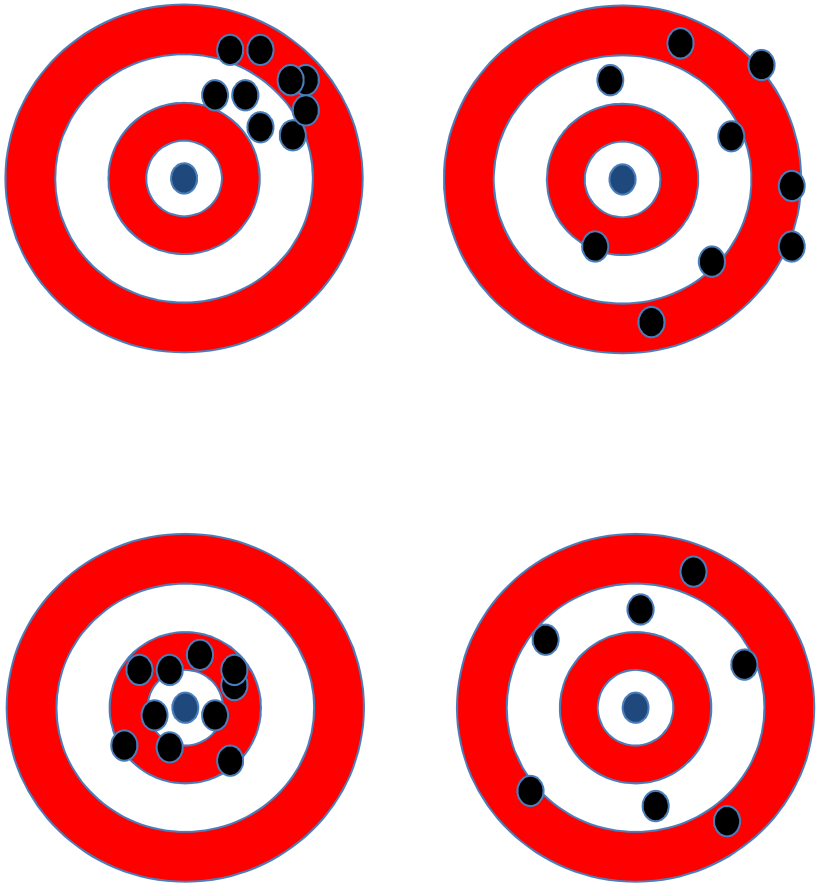
Hiba



Több véletlen mintavételt képzelünk el –becslések

Hiba – 2 dimenziója

„Átlagos eltérés” (torzítás)
Szisztematikus hiba



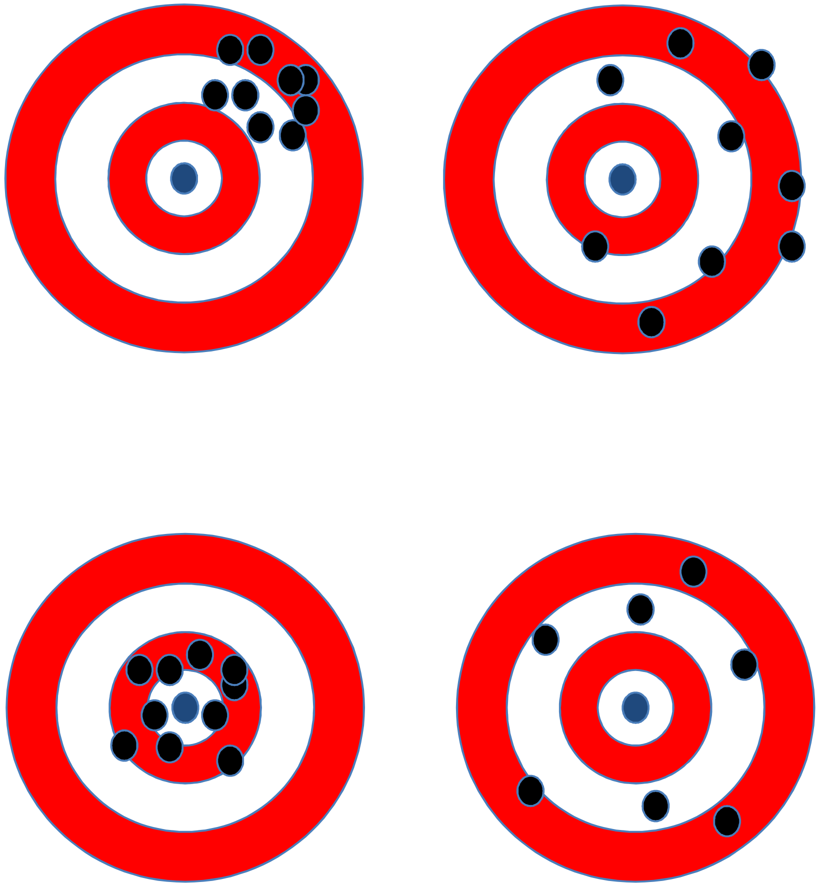
„Változékonyság” (szóródás);
Véletlen hiba (mintavételi hiba)

1. A becslések változékonysága

2. A becslések „közepének” eltérése a valódi értéktől

Hiba – 2 dimenziója

„Átlagos eltérés” (torzítás)
Szisztematikus hiba



„Változékonyság” (szóródás);
Véletlen hiba

Jó becslés, ha:

Torzítatlan:

A becslések „közepe” (várható értéke) a valódi érték

Hatásos:

A becslések változékonysága a lehető legkisebb

Konzisztens:

az elemszám növelésével csökken a becslés változékonysága

()

Néhány megjegyzés

- A várható érték becslésére az átlag torzítatlan, konzisztens és a leghatásosabb
- A szórás becslésénél az n-1-gyel való osztás torzítatlan, de n-nel osztva torzított

Konfidencia intervallumok

De nekünk csak 1 mintánk van, valódi értéket nem ismerjük!

– Akkor számítható egyáltalán a hiba?

IGEN, a véletlen hiba (mintavételi hiba): ez a becslés standard hibája - a mintából becsülhető!

Konfidencia intervallumok

De nekünk csak 1 mintánk van, valódi értéket nem ismerjük!

– Akkor számítható egyáltalán a hiba?

IGEN, a véletlen hiba (mintavételi hiba): ez a becslés standard hibája - a mintából becsülhető!

Példa:

az átlag standard hibája (röviden standard hiba, vagy átlag hibája, SEM)

Emlékeztető: Centrális határeloszlás tétele (mintavételi átlagokra):

ha egy adathalmazból *n* elemű mintákat veszünk, akkor elég általános feltételek teljesülése esetén a minták átlagai normál eloszlásúak lesznek, és az eloszlás varianciája az eredeti eloszlás varianciájának *n*-ed része lesz.

Konfidencia intervallumok

1. Tehát az átlag standard hibája (szóródásának mértéke): a minta varianciájának n-ed részének négyzetgyöke.
2. Megadható egy olyan tartomány, amely az átlagot adott valószínűséggel tartalmazza: ez az átlag konfidencia intervalluma.
   az adott valószínűség: konfidenciaszint

Az átlag esetében (a normál eloszlás miatt) pl:
Az átlag 95% szintű konfidencia intervallumának határai:

$$\sim \bar{x} \pm 2 * SEM$$

Más becslésre is megadható adott szintű konfidencia intervallum!
Mutatja a becslés értékét, annak hibáját, adott valószínűségű intervallumát

Hipotézisvizsgálatok
Mintavételi (véletlen hiba) másképp

Hipotézisvizsgálat **célja**: **eldöntendő kérdésre statisztikai választ** adjon

Hogyan?

Állítás *direkt bizonyítása*: minden lehetséges esetre belátom, hogy igaz.
pl: 1,2…n számok összege: (n+1)*(n/2) – teljes indukcióval

Állítás *indirekt bizonyítás*: felteszem a pontos ellenkezőjét és belátom, hogy nem igaz.

Indirekt bizonyítás

Egy dobozban van 100 üveggolyónk. Mindegyikük vagy piros, vagy fehér.

1. eset: A feltevésünk (*H*): mind fehér.
Kísérlet: véletlenül kiveszünk egy golyót a dobozból.
Megfigyelés: az első kivett golyó piros színű.
Következtetés: a megfigyelésünk valószínűsége, *ha* a H igaz, 0: a feltevésünk 100% biztonság-gal téves.

2. eset: A feltevésünk (*H*): 99 fehér és egy piros.
Kísérlet: véletlenül kiveszünk és visszarakunk egy golyót, ezt 5-ször végezzük el.
Megfigyelés: mind az 5 piros.
Következtetés: a H *majdnem* 100% biztonsággal téves: a megfigyelésünk valsége, *ha* a H igaz, $0.01^5 = 10^{-10}$: (gyakorlatilag lehetetlen).

3. eset: A feltevésünk (*H*): 50 fehér és 50 piros.
Kísérlet: véletlenül kiveszünk és visszarakunk egy golyót, ezt 5-ször végezzük el.
Megfigyelés: mind az 5 piros.
Következtetés: most nem olyan egyértelmű: a megfigyelésünk valószínűsége, *ha* a H igaz, $0.5^5 = 0.03125$: alacsony de azért nem túl valószínűtlen..

4. eset: A feltevésünk (*H*): mind piros.
Kísérlet: véletlenül kiveszünk és visszarakunk egy golyót, ezt 5-ször végezzük el.
Megfigyelés: mind az 5 piros.
Következtetés: a megfigyelésünk valószínűsége, *ha* a H igaz, $1^5 = 1$. Ez mit is jelent?

Indirekt bizonyítás

Egy dobozban van 100 üveggolyónk. Mindegyikük vagy piros, vagy fehér.

1. eset: A feltevésünk (*H*): mind fehér.
Kísérlet: véletlenül kiveszünk egy golyót a dobozból.
Megfigyelés: az első kivett golyó piros színű.
Következtetés: a megfigyelésünk valószínűsége, *ha* a H ig… a feltevésünk 100% biztonság-gal téves.

Lehetetlen esemény

✓

A feltevés hamis

A falszifikáció működik

A verifikáció nem működik

4. eset: A feltevé… d piros.
Kísérlet: véletlenül kiveszünk és visszarakunk egy golyót, ezt 5-ször végezzük el.
Megfigyelés… os.
Következtetés: a meg… valószínűsége, *ha* a H igaz, $1^5 = 1$. Ez mit is jelent?

A feltevés igaz

✗

Biztos esemény

Indirekt bizonyítás

A 2. és 3. eset?

Matematikai logika:

Van egy feltevésünk (*H*).

Ha *H* igaz, az *E* esemény nem következhet be.

E bekövetkezik.

Tehát *H* hamis.

Amint azt az előzőekben láttuk, **egy feltevést csak elvetni tudunk**.

Statisztikai logika:

Van egy feltevésünk (*H*).

Ha *H* igaz, az *E* esemény bekövetkezése nagyon valószínűtlen.

E bekövetkezik.

Elvetjük *H*-t. De sohasem lehetünk 100%-ig bizonyosak, hogy *H* hamis.

Ez esetben **egy feltevést még elvetni sem tudunk százszázalékos bizonyossággal – mintavételi hiba lehetséges**. ENNEK VALÓSZÍNŰSÉGÉT KELLENE MEGBECSÜLNI

Milyen legyen a kérdés?

...legyen eldöntendő (igen/nem, dichotomikus).

- 50%-e a mielómások ötéves túlélési aránya (vagyis valsége)?
- Eltér-e a Cushing-kórosok vérkoleszterinszintje a normálisnak tartott 200 mg/dL értéktől? ✓

- Mekkora a mielómások ötéves túlélési aránya?
- Mi a Cushing-kórosok koleszterinszintjének várható értéke? ✗

...esetek halmazára és ne egyedi esetre vonatkozzon.
(És a kérdés mindig az alapsokaságra irányul, nem a mintára.)

- 50%-e a mielómások ötéves túlélési aránya? ✓

- Élni fog-e még ez a mielómás beteg 5 év múlva? ✗

...legalább az egyik válaszlehetőségnek egyértelműnek kell lennie.

- 50%-e a mielómások ötéves túlélési aránya? ✓

- Kevesebb-e a mielómások túlélési aránya, mint 50%? ✗

Milyen legyen a hipotézis?

Két válaszunk (feltevésünk) van a kérdésre:

A nullhipotézis (H_0)

- **Egyértelmű**: csak egyféleképpen teljesülhet. Valamilyen formában szerepel benne az = jel.
  A mielóma ötéves túlélési aránya 50%.

- A tudomány **jelenleg elfogadott álláspontját** tükrözi,
  A Cushing kórosok vérkoleszterinszintje megegyezik az egészségesek alapsokaságának átlagával.
  vagy valamit, ami „közhelyes”, kevés benne a megkötés (Occam borotvája).
  Az érmefeldobásban a fejek valsége 50%.
- **Nem** feltétlenül nemleges válasz a feltett kérdésre.

Az alternatív vagy ellenhipotézis (H_1)

- Jellemzően **többféleképpen is teljesülhet**.
  A mielóma ötéves túlélési aránya nem 50%.
  (lehet kissé több vagy sokkal kevesebb stb.)

- A tudományos konszenzusnak ellent-mondó **új megállapítást** fogalmaz meg,
  A Cushing kórosok vérkoleszterinszintje eltér az egészségesek alapsokaságának átlagától.
  vagy egy kevéssé közhelyes, megkötéseket tartalmazó állítást.
  Az érmefeldobásban a fejek valsége nem 50%.
- Legtöbbször a **H_0-sel komplementer** (vagyis annak tagadása).
  H_1 = nem H_0

Mintavételből származó („véletlen”) hiba

Hipotézisvizsgálat **célja**: **eldöntendő kérdésre statisztikai választ** adjon

Kiindulópont: az eldöntendő kérdés statisztikai átfogalmazása, majd :

H_0: nullhipotézis – „véletlen” hiba

H_a (vagy H_1): alternatív hipotézis (ellenhipotézis) – nem H_0

Döntés alapja: a „véletlen” szerepe a H_0 esetén (mintavételi hiba)

A minta az, amely esetleg megcáfolja H_0-t

Mintavételből származó („véletlen”) hiba

Hipotézisvizsgálat **célja**: **eldöntendő kérdésre statisztikai választ** adjon

Kiindulópont: az eldöntendő kérdés statisztikai átfogalmazása, majd :

H_0: nullhipotézis – „véletlen” hiba

H_a (vagy H_1): alternatív hipotézis (ellenhipotézis) – nem H_0

Döntés alapja: a „véletlen” szerepe a H_0 esetén (mintavételi hiba)

A minta az, amely esetleg megcáfolja H_0-t

		A populációban (a valóságban) a null hipotézis:	
		Igaz	Hamis
A döntés: a null hipotézist:	Megtartom (Nem vetem el)	Helyes döntés	Hiba (**másod fajú**) (β) (álnegatív eredmény)
	Elvetem	Hiba (**első fajú**) (α) (álpozitív eredmény)	Helyes döntés (**erő**) (1-β)

Egyszerűsített példa 1. lépések

Szituáció: Dobókockás társasjátékot játszunk – nem mi nyerünk...
„Rossz” a dobókocka?

Mi a kérdés??!! (és mire vonatkozik):

Minden oldala ugyanolyan valószínűségű-e? (ennek a kockának)
A hatos dobás valószínűsége eltér 1/6-tól, méghozzá nagyobb?
– most válasszuk ezt

Nullhipotézis ??!!:

A 6-os dobás valószínűsége 1/6 vagy kisebb – (*összetett hipotézis*)
Vegyük az alternatív hipotézis szempontjából „legrosszabb esetet”
(hiszen ha akárcsak egy eleme az összetett hipotézisnek „megfelelhet” a nullhipotézisnek, akár nincs bizonyíték az elvetésre a teljes tartományban):

H_0: A 6-os dobás valószínűsége 1/6.
H_a: nagyobb 1/6

Egyszerűsített példa további lépések

Mi a kérdés: A hatos dobás valószínűsége eltér 1/6-tól, méghozzá nagyobb?

Nullhipotézis: H_0: A 6-os dobás valószínűsége 1/6.

Mennyi bizonyíték kell, hogy eltérőnek tekintsem? - Szignifikancia szint ??!!:

Hatóság megadta… Adott terület szakirodalma --- **KIINDULÓPONT**, de kisebb/nagyobb, gyakoribb/ritkább mellékhatások?, olcsóbb/ drágább?...

itt: minden hétvégén játszunk vele, dobókockát nem drága lecserélni

– legyen 5% helyett 10% (kevesebb bizonyíték elég H_0 elvetéséhez)

Bizonyíték gyűjtése – a minta ??!! (hogyan, mennyi…kérdezze statisztikusát):

24 dobásból 6 darab 6-os.

Tényleg a kérdésre válaszolok?

Módosítanom/pontosítanom kell a kérdést? (mire vonatkozik)

(pl. itt: dobókockányi lukak vannak az asztalon…)

Egyszerűsített példa további lépések

Mi a kérdés: A hatos dobás valószínűsége eltér 1/6-tól, méghozzá nagyobb?

Nullhipotézis: H_0: A 6-os dobás valószínűsége 1/6.

Szignifikancia szint:10%

A minta: 24 dobásból 6 darab 6-os.

Lényeges a különbség egyáltalán – releváns ??!!:

24 dobásból 6 darab 6-ost dobtam, ez 6/24 = 1/4 ez **1,5**-szeres az 1/6-nak – igen, ez nekem lényeges

Mennyi a bizonyíték – p-érték ??!!:

Hogyan számoljak?, (torzítatlanul, hatásosan, konzisztensen…)

Lényeges: kérdés, változók mérési skálája, mérési körülmények

Számolás során kell:

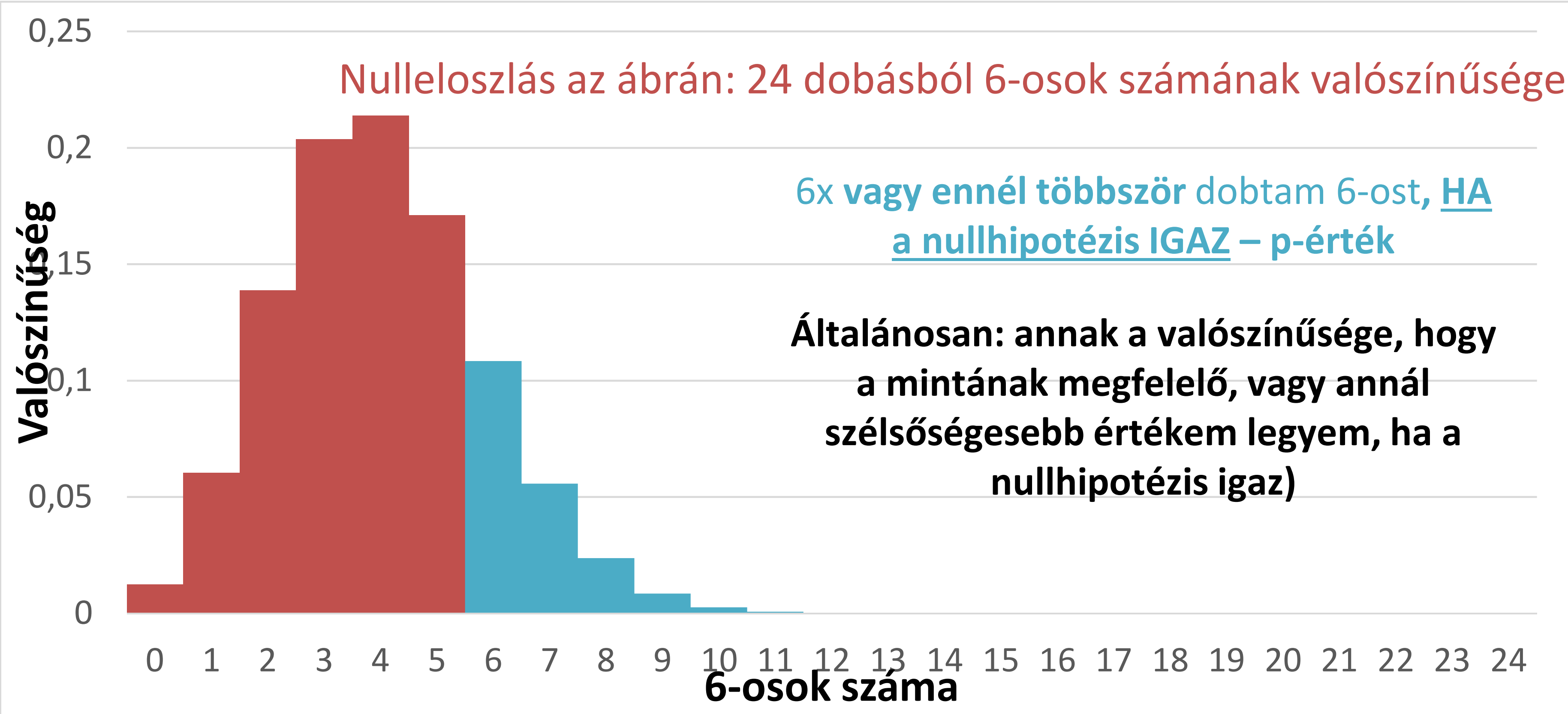
Nulleloszlás – ha a null hipotézis igaz, mekkora a valószínűsége az elméletileg lehetséges mintáknak, amit így vettünk?

A minta „helye” ebben az eloszlásban:

próbastatisztika (teszt-statisztika, pl: valószínűség, t-érték, p-érték…)

A számolás a „háttérben”

Binomiális
próbával!



Egyszerűsített példa további lépések

Mi a kérdés: A hatos dobás valószínűsége eltér 1/6-tól, méghozzá nagyobb?

Nullhipotézis: H_0: A 6-os dobás valószínűsége 1/6.

Szignifikancia szint:10%

A minta: 24 dobásból 6 darab 6-os.

Lényeges a különbség egyáltalán – releváns: minta alapján: 1/4 ez **1,5**-szeres az 1/6-nak

Mennyi a bizonyíték – p-érték: 0,1995

Döntés: nincs elég bizonyíték az elvetésre – megtartjuk a H_0-t

		A populációban (a valóságban) a null hipotézis:	
		Igaz	Hamis
A döntés: a null hipotézist:	Megtartom (Nem vetem el)	Helyes döntés	Hiba (**másod fajú**) (β) (álnegatív eredmény)
	Elvetem	Hiba (**első fajú**) (α) (álpozitív eredmény)	Helyes döntés (**erő**) (1-β)

Egymintás Student-féle t-próba

Mire vagyok kíváncsi

A minta várható értéke megegyezik-e egy ismert populációátlaggal

Változó típusa

1 számszerű és folytonos

Feltételek

Egymástól független megfigyelések

átlagok eloszlása legyen normál azaz:

normál eloszlású változó vagy nagy minta (CLT miatt)

Megjegyzés: Számolás:

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

normalitást NE teszteljük másik hipotézisvizsgálattal! Előzetes tudásunk, ábrázolás alapján

(+teszt növeli az elsőfajú hibát, „multiplicitás”)

Párosított t-próba

Mire vagyok kíváncsi

Két várható érték megegyezik-e - párosított csoportokban

Változó típusa

1 számszerű és folytonos, valamint 1 bináris („csoportok”)

Feltételek

Egymástól független megfigyelések a csoportokban, párosított csoportok

átlagok különbségének eloszlása legyen normál azaz:

normál eloszlású különbség vagy nagy minta

Megjegyzés:

párosított próbák ereje általában nagyobb, mint nem párosítotté

egyéb helyparaméterek (kvantilisek) összehasonlítására is jó

átlagok különbsége = különbségek átlaga

Student-féle 2 mintás t-próba

Mire vagyok kíváncsi

Két várható érték megegyezik-e

Változó típusa

1 számszerű és folytonos, valamint 1 bináris („csoportok”)

Feltételek

Csoportonként és egymástól is független megfigyelések

átlagok eloszlása legyen normál mindkét csoportban azaz:

normál eloszlás a csoportokban vagy nagy minta

szórások azonosak legyenek a csoportokban

Megjegyzés:

egyéb helyparaméterek (kvantilisek) összehasonlítására is jó

szórásokat általában nem ismerjük, ne teszteljük! (multiplicitás)

Welch próbát használjunk!

Welch-féle t-próba

Mire vagyok kíváncsi

Két várható megegyezik-e

Változó típusa

1 számszerű és folytonos, valamint 1 bináris („csoportok”)

Feltételek

Csoportonként és egymástól is független megfigyelések

átlagok eloszlása legyen normál mindkét csoportban azaz:

normál eloszlás a csoportokban vagy nagy minta

Megjegyzés:

egyéb helyparaméterek (kvantilisek) összehasonlítására is jó

nem érzékeny az eltérő varianciákra (robosztus a varianciák eltérésére)

Khi-négyzet függetlenség próba

Mire vagyok kíváncsi

Ismeretlen eloszlás, illetve gyakoriságok és ismert eloszlás megegyezik-e

Másképp: a 2 változó független-e

Változó típusa

2 kategoriális változó

Feltételek

Egymástól független megfigyelések

Az összes „elvárt” gyakoriság nagyobb, mint 1 és maximum 20%-uk kisebb, mint 5.

„Korrelációs” t-próba (Pearson lineáris regresszió)

Mire vagyok kíváncsi
- 2 változó (lineárisan) függ-e egymástól

Változó típusa
- 2 számszerű változó (X és Y)

Feltételek
- Egymástól független megfigyelések (x és y párok)
- Lineáris összefüggést feltételezünk
- x értékei mérési hiba nélküliek
- minden x értékre y-ok eloszlása normális
- minden x értékre y-ok szorása azonos

Megjegyzés:
- 2 értéket becslünk: a lineáris meredekségét és tengelymetszetét
  - a meredekség a lényeges, ezt teszteljük

Releváns, de mégsem szignifikáns

Okai lehetnek:

kicsi statisztikai erő:

kicsi elemszám (+klinikai problémák: pénz, etikai kérdések)

nagy a változatosság a vizsgált paraméterben, illetve alanyokban

kicsi erejű próbát használunk

(megsértjük a próba feltételeit)

nem mérünk elég precízen

Tervezzünk előre!!

más hibák (lásd mindjárt)

Pechünk volt (mintavételi hiba, véletlen)

-Kérdezze meg statisztikusát...

(☺ pl: https://www.youtube.com/watch?v=PbODigCZqL8)

Mintabeli érték

<u>Populációbeli (valódi) érték</u>

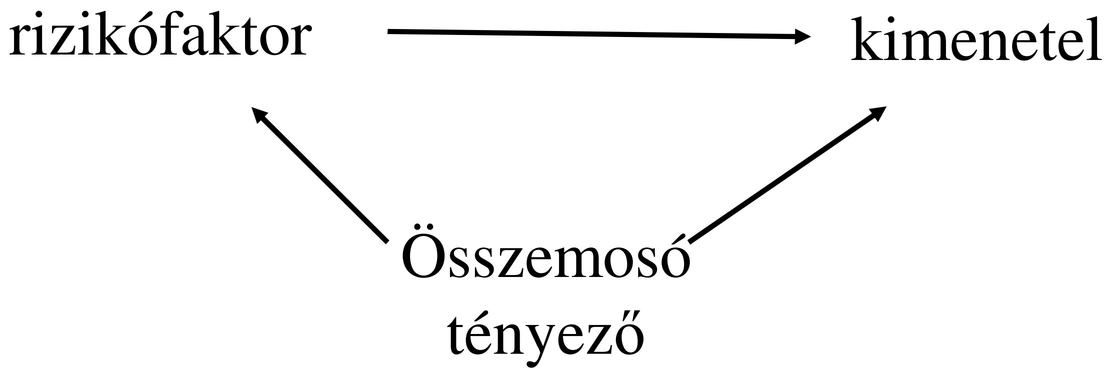
Hibák

Torzítás (szisztémás hiba) (bias):
Kiválasztási hiba (Selection bias)
Információs hiba (Information bias)
Összemosó tényező (Confounding bias)

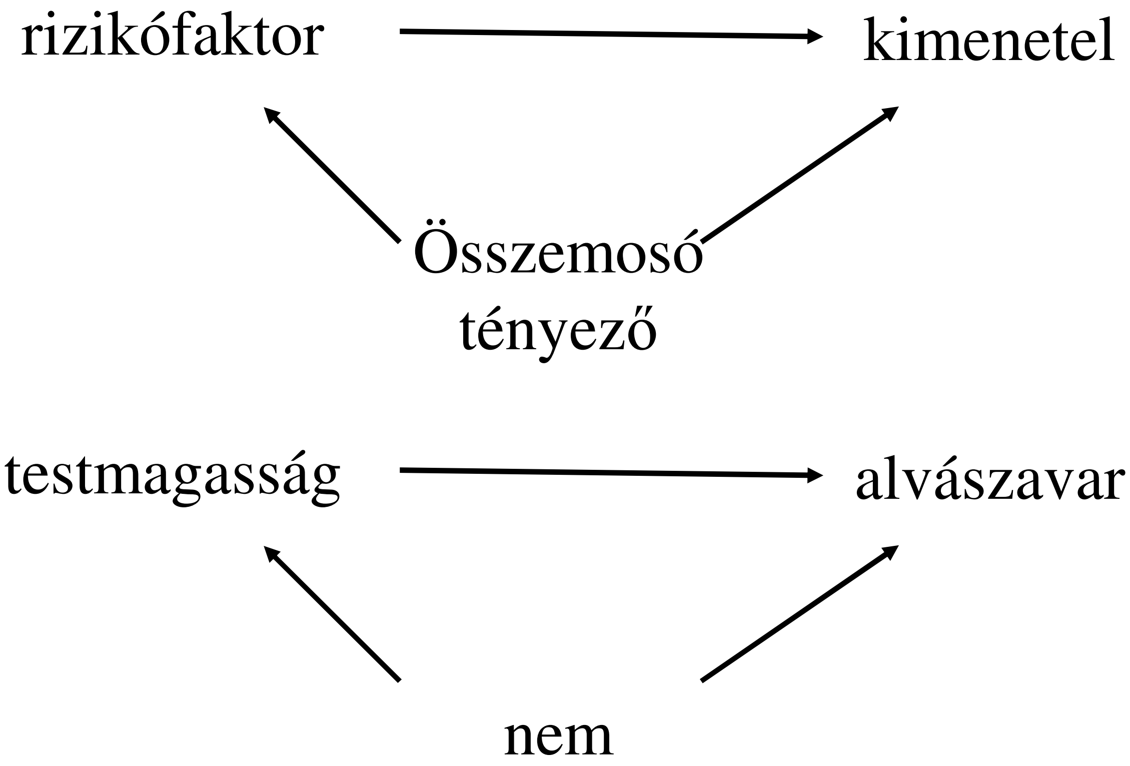
Mintavételből származó
(„véletlen”) hiba

(elemszám növelésével csökken)

Összemosó (megzavaró) hiba (confounding)



Összemosó hiba



Leggyakoribbak: nem, életkor – mindig gondoljunk rá!

Kiválasztási, Információs hiba

Kiválasztási hiba:
Különbség van a felmérésbe beválasztottak és nem beválasztottak vagy a beválasztottak csoportba sorolása között (egy kimenetelt befolyásoló paraméter tekintetében)

tipikus hibák: életkor, nem eltérő a csoportokban
alappopoláció eltérő
utánkövetés eltérő

Információs hiba:
Téves az alanyoktól kapott, alanyokról gyűjtött információ (, amely befolyásolja a kimenetelt)

tipikus hibák: visszaemlékezés rossz
elfogódottság

Ellenőrző kérdések

- Mit jelent a szisztémás és a véletlen hiba?
- Mit jelent, ha egy becslés torzítatlan?
- Mit jelent, ha egy becslés hatásos?
- Mit jelent, ha egy becslés konzisztens?
- Definiáld a konfidencia intervallumot!
- Mi a hipotézisvizsgálatok célja?
- Milyen a hipotézisvizsgálatok „jó kérdése”?
- Milyen a hipotézisvizsgálatok „jó hipotézise”:
- Mit jelent a szignifikancia szint?
- Mit jelent az, hogy egy hatás, különbség… releváns?
- Definiáld az elsőfajú hibát!
- Definiáld a másodfajú hibát!
- Fogalmazd meg a hipotézisvizsgálat p-értékének jelentését.
- Mit jelent a konfounding (összemosó) hiba?
- Mit jelent az információs és kiválasztási torzítás?