BIOMETRIA

Az élővilág kutatásának matematikai, statisztikai eszköztára

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Témakörök

- Alapismeretek
- Variancia Analízis
- Korreláció- és Regresszió Analízis
- Esetszám- sorok és táblázatok elemzése

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

I. rész: ALAPISMERETEK

- Bevezetés
- Alapsokaság és minta
- Változók, adatok típusai, változatai
- Átlagok
- A minta jellemzői
- Az alapsokaság jellemzői
- Fontosabb sokasági megoszlások
- Paraméterbecslés, konfidencia intervallum
- Statisztikai következtetés: Hipotézis vizsgálat

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Bevezetés

- A ***kutatás***, amely statisztikai vizsgálatokra épül, az ***alapsokaság***(ok)ban fennálló összefüggést vizsgálja minta alapján.
- Az alapsokaságra vonatkozóan ***hipotéziseket*** állítunk fel és ezeket a mintára épülő statisztikai próbákkal ellenőrizzük.
- E szemléletben ne feledjük, hogy a minta ***esetleges***, a végkövetkeztetés függ attól, hogy az alapsokaság mely egyedei kerültek a mintába. Ebből adódóan a statisztikai következtetés nem abszolút érvényű, csak valószínűsíthető.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A biometriai vizsgálódás fázisai

- Kérdés felvetés, modellválasztás vagy modellalkotás
- Kísérlet-, ill. adatgyűjtés tervezése
- A kísérlet vagy adat felvételezés végrehajtása
- Adatelemzés
- Az eredmények értelmezése (interpretáció)

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ Az alapsokaság (populáció)

- a vizsgálat tárgyát képező egyedek, esetek összessége
- állhat véges sok egyedből, de általában végtelen sok egyedből áll

  Szűkebb értelemben az egyedek (esetek) valamely vagy egyszerre több *ismérv*ének összessége

  - Például: a magyar állampolgárok 2011. január elsején.
- Szűkítve (ismérvek): ezen emberek életkora, neme, egészségi állapota stb. a jelölt napon

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A minta

- ***Minta*** az alapsokaságból kiválasztott véges sok egyed, ***megfigyeléssel***, felméréssel vagy kísérletezéssel nyerjük. Szűkebb értelemben alapsokaság az egyedek valamely (vagy több) ***ismérvének*** összessége, a minta pedig a megfigyelési egységeken mért vagy megállapított adatok

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ Változók és adatok

- **Változó:**

  az alapsokaság egyedei ismérvének „értéke” mintavétel, megfigyelés előtt, jelölése a továbbiakban: X, Y, X_1, X_2, …

- **Adat:**

  a mintába felvett egyed(ek) szóbanforgó ismérvének „értéke” a mintavétel (megfigyelés, adatfelvétel) után

  - kis latin betűkkel jelöljük: x, y, x_1, x_2, …,

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

➢ Változók és adatok, példa

- **Valamely adott helyen**
  a holnapi csapadékmennyiség
- **ma még változó: X**
- **holnapután már adat, pl x = 8 mm**

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Ismérvek (változók) típusai, változatai

- megkülönböztetünk

- *kvalitatív* (minőségi, megállapítható) ismérveket
  Pl: „nem”, „szín”, „hivatali beosztás”

- *és kvantitatív* (mennyiségi, mérhető) ismérveket
  ennek két altípusa van:
  - diszkrét (pl: „iskolák száma adott településen”)
  - folytonos (pl: „hőmérséklet adott helyen és időben”)

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvalitatív ismérv változatai:

- **Osztályok, kategóriák (ezek is adatok!)**

Pl:	típus	változatok
	nem	férfi, nő
	szín	fehér, piros, stb.

- **Dichotom ismérv: két változata van**
- **Trichotom ismérv: három változata van**

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Diszkrét kvantitatív változó

- Lehetséges értékei (változatai)
- véges, sok

  pl: „fiúk száma egy 30 fős osztályban”

  lehet 0, 1, 2, ……..,30
- megszámlálhatóan végtelen sok

  (gyakorlatilag nincs felső határa)

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Folytonos kvantitatív változó

- **lehetséges értékei egy intervallum bármely értéke**

  pl: vércukorszint

  Ph érték

  életkor

  testsúly

  hőmérséklet

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Adat-transzformációk, átskálázás

Gyakran a mért (megfigyelt, megállapított) adatok helyett célszerűbb ezek

„ transzformált”–jaival dolgozni.

- Kvalitatív adatokat olykor kvantifikáljuk (pl: bonitálás)
- Kvantitatív adatok leggyakoribb transzformációja:

  log-transzformáció

  négzetgyök transzformáció

  reciprok- képzés

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Mérési skálák

- Az ismérveket megfelelő skálán mérjük.

a) ***Nominális skála*** tipikus kvalitatív skála. Értékei nem sorrendezhetők, csak két egyed azonos kategóriába, vagy különböző kategóriába tartozása állapítható meg (X=Y) illetve (XY).

b) ***Ordinális skála*** olyan kvalitatív skála, melyen a kategóriák sorrendje is megállapítható (X<Y), pl. bonitálási skála.

c) ***Intervallum skála***, amelyen két egyed távolsága (X-Y) mérhető. A skálának nincs valóságos nullpontja, X=0 nem jelenti az ismérv hiányát (pl. hőmérséklet).

d) ***Arány – (hányados) skála*** olyan kvantitatív skála, amelynek valódi nullpontja van. Ilyen skálán két érték aránya (Y/X) értelmes viszonyszám (pl. tömeg).

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ Kvantitatív adatok átlagai

- Jelölje $x_1, x_2, \ldots ..x_n$ az adatokat

Többféle átlagról beszélhetünk

- számtani (aritmetikai) átlag $\bar{x}$
- mértani (geometriai) átlag $\bar{x}_g$
- harmonikus átlag $\bar{x}_h$
- négyzetes (kvadratikus) átlag $\bar{x}_n$
- és általánosabban: f-átlag.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvantitatív adatok átlagai 1

a***) számtani átlag*** (jele: $\bar{x}$) a mintaelemek átlaga.

- Jellemzője, hogy a mintaelemek összege ugyanannyi, mint ha mindegyik elem helyébe $\bar{x}$ -ot teszünk
- $$\bar{x} = \frac{x_1 + x_2 + x_3 + ... x_n}{n} \qquad \bar{x} = \frac{\sum x_i}{n}$$

Fontos tulajdonsága még, hogy a $d_i = x_i - \bar{x}$ eltérések összege zéró.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvantitatív adatok átlagai 2

b) ***A mértani átlag*** (jele $\bar{x}_g$) pozitív mintaelemek esetén gyakran reálisabb a számtani átlagnál.

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot ..... x_n}$$, másként $$\overline{x_g} = \sqrt[n]{\prod x_i}$$

Ezt úgy jellemezhetjük, hogy

$x_1 \cdot x_2 \ldots = \underbrace{\bar{x}_g \cdot \bar{x}_g \cdot \ldots}_{n \text{ tényező}}$, a két szorzat azonos

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvantitatív adatok átlagai 3

c) ugyancsak pozitív mintaelemek esetén néha a ***harmonikus átlag*** a legjobb közép-jellemző

$$\bar{x}_h = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots \frac{1}{x_n}}{n}}$$

$$\bar{x}_h = \frac{n}{\sum \frac{1}{x_i}}$$

- Az adatok reciprokainak összege nem változik, ha mindegyik helyébe a harmonikus átlagot tesszük.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvantitatív adatok átlagai 4

d) ***négyzetes átlag*** (jele $\bar{x}_n$) az adatok négyzetösszegének a négyzetgyöke. Más szóval az adatok négyzetösszege nem változik, ha minden adat helyére $\bar{x}_n$ kerül.

$$\bar{x}_n = \sqrt{\frac{x_1^2 + x_2^2 + \dots x_n^2}{n}}$$, tömören: $$\bar{x}_n = \sqrt{\frac{\sum x_i^2}{n}}$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Általános átlag

- Az említetteken kívül egyéb átlagok is képezhetők. Mindezek úgy foghatók fel, hogy az eredeti x_i adatokat alkalmas módon ***transzformáljuk*** és a transzformált adatok átlagát visszatranszformáljuk.
- Például a geometriai középnél a $\log(x_i)$ transzformált adatok átlagát számítjuk, majd ezt az exp(.) „inverz transzformációval” alakítjuk $\bar{x}_g$ -vé.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Miért kell többféle átlag?

Hogy melyik átlag reális, azt az alapsokaság megoszlásának típusa dönti el (ld. később)

Számtani átlag reális szimmetrikus megoszlásnál.

Mértani átlag reális „ log normális” eloszlásnál,

pl. permetcseppek mérete

Harmonikus átlag reális „ exponenciális” eloszlásnál,

pl. túlélési idő inszekticidek alkalmazásánál

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ A minta jellemzői

1. Gyakorisági megoszlás

- ***Kvalitatív minta*** **gyakorisági megoszlása**
  az egyes kategóriákba,osztályokba eső esetszámok, gyakoriságok (f_1, f_2, f_3,…),
  vagy a relatív gyakoriságok (f_1/n, f_2/n, f_3/n,…) felsorolása.
  Szokásos szemléltetése: oszlop-diagram, torta-diagram
- ***Kvantitatív diszkrét ismérv*** **mintájának gyakorisági megoszlását megadhatjuk az egyes értékek gyakoriságainak vagy relatív gyakoriságainak felsorolásával. Grafikonja „gereblye fogak” (bot)**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Folytonos változó mintájának gyakorisági megoszlása

- Legyenek a minta elemei $x_1, x_2, \ldots, x_n$
- Soroljuk az adatokat a legkisebbtől a legnagyobbig
  c egyenlő közű osztályba
- az osztályok száma legyen
- $c \approx 1+3{,}3 lg n$, egészre kerekítve
- gyakoriságok: a k-adik osztályba eső minta elemek száma: f_k
  összegük: $\sum f_k = n$
- relatív gyakoriságok : $r_k = f_k / n$
  a relatív gyakoriságok összege 1=100 %

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Hisztogram

- **téglalapok sorozatával ábrázoljuk a gyakoriságokat vagy a relatív gyakoriságokat**
- **a k-adik téglalap alapja h (az osztályköz) magassága f_k vagy r_k**

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A minta jellemzői

2. Centrális jellemzők

- 1. **Kvantitatív minta mediánja**

- **Medián: nagyság szerint sorrendezett mintaelemek**
  - **középső tagja, ha n páratlan**
  - **a két középső tag átlaga, ha n páros**
- **Bonyolultabb a medián számítása, ha csak az osztály-gyakoriságokat ismerjük (itt nem részletezzük)**
- **Medián lényege: tőle balra is, jobbra is ugyanannyi adat van**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

2. Kvantitatív minta átlaga (mean)

- A minta átlagán a mért, vagy -szükség esetén- a transzformált adatok számtani átlagot értjük $\bar{x}$
- Főbb tulajdonságai
  1) a $\sum(x_i - a)^2$ négyzetösszeg akkora legkisebb, ha $a = \bar{x}$, a számtani átlag
- 2) az átlag mértékegysége azonos az adatok mértékegységével
- 3) az átlag skála-kezdőpont függő, azaz, ha minden adathoz egy *a* értéket adunk, az átlag is *a*-val változik
- 4) az átlag mértékegység-függő, azaz, ha minden adatot egy *c* értékkel szorzunk, az átlag is *c*-vel szorzódik

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Gyakoriságokkal súlyozott átlag-formula

Ha a mintában az x_i elem (lehetnek ezek transzformált adatok is) f_i-szer fordul elő, akkor az átlag (akár zseb-kalkurátorral is) gyorsabban számolható:

$$\bar{x} = \frac{\sum f_i x_i}{n}$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A minta jellemzői

3. Szóródás jellemzők

- ***Kvantitatív minta szóródásának mértékei***

A szóródásnak többféle mértékszáma van, ezek közül a legfontosabb a *szórás*

a *szórás* (s) „nagyjából” a $d_i = x_i - \bar{x}$ eltérések négyzetes átlaga, jele: s, olykor S.D. (Standard Deviation). Alapos okunk van arra, hogy n helyett n-1 –gyel osszunk

a *variancia* (Var vagy s^2) a szórás négyzete

Képletben:

$$Var = s^2 = \frac{1}{n-1}\sum d_i^{\,2} = \frac{1}{n-1}\sum \left(x_i - \bar{x}\right)^2 \qquad S.D. = s = \sqrt{s^2}$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Gyakoriságokkal súlyozott variancia-formula

Ha sok adatunk van és köztük az egyes értékek többszörösen, f gyakorisággal fordulnak elő (különösen diszkrét változó esetén) akkor nincs értelme minden adatot beütni a „gépbe”, a számítás egyszerűsíthető.

Ha a mintában az x_i elem (lehetnek ezek transzformált adatok is) f_i-szer fordul elő, akkor a szórásnégyzet (variancia) (akár zseb-kalkurátorral is) gyorsabban számolható:

$$s^2 = \frac{\sum f_i \left(x_i - \bar{x}\right)^2}{n-1} = \frac{\sum f_i x_i^2 - n\left(\bar{x}\right)^2}{n-1}$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A szórás főbb tulajdonságai

- mértékegysége azonos az adatok mértékegységével
- a szórás kezdőpont-független, az x_i+a adatok szórása azonos az x_i adatok szórásával.
- A szórás mértékegység függő, pontosabban a cx_i adatok szórása $|c|$-szer akkora, mint az x_i adatok szórása

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Az átlag hibája (szórása)

- **a szórás (s) valójában egyetlen mintaelem „megbízhatatlanságát” méri.**
- **A minta-átlag annál pontosabb minél nagyobb a mintanagyság (n)**
- **$\bar{x}$ „ megbízhatatlanságát” méri az átlag hibája, $s_{\bar{x}}$ vagy S.E. (Standard Error)**
- **Számítása**

$$s_{\bar{x}} = S.E = s/\sqrt{n}$$

tehát pl., ha a mintaelemek számát meg-16-szorozzuk, az átlag pontossága meg-4-szereződik

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A relatív szórás (CV, variációs koefficiens)

$$CV\% = 100\ s / \bar{x}\ \%$$

- akkor értelmes, ha az adatok pozitívak
- s és $\bar{x}$ is mértékegységfüggő (azonos dimenziójúak) hányadosukból kiesik a mértékegység, ennélfogva szemléletesebben (%-ban) méri a szóródást
- értéke 0%-tól $100\sqrt{n}$ %-ig eshet

  (tehát lehet 100 %-nál nagyobb is)

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvalitatív minta jellemzői

- Kvalitatív mintánál átlagról nem beszélünk, (kivéve, ha változatait kvantifikáltuk)
- Variabilitását diverzitás mutatókkal mérhetjük.
- Legyenek az egyes kategóriákba eső egyedek gyakoriságai $f_1, f_2, ..., f_c$, összegük n
- Simpson-Yule féle diverzitási index

$$D_{S-Y}=1-\sum(f_k/n)^2$$, maximális értéke 1-1/c

- Shannon-Weaver féle diverzitási index

$$D_{S-W}=-\sum(f_k/n)\ln(f_k/n),$$

maximális értéke lnc, ahol c a kategóriák száma

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ Az alapsokaság jellemzői

- **Megoszlás**
  - **sűrűség függvény**
  - **eloszlás függvény**
- **sokasági átlag (várható érték)**
- **sokasági variancia és szórás**
  - **kvalitatív ismérvnél: diverzitás**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Megfeledkezve a matematikai szabatosságról, gondolatban tekintsük mintának a teljes sokaságot.

Ekkor a minta jellemzői „átnőnek” az alapsokaság jellemzőivé:

- **a relatív gyakoriságokból valószínűség (p) lesz**
- **a „gereblyéből” valószínűség eloszlás lesz**
- **a hisztogramból „sűrűség függvény„ lesz**
- **a minta átlagából várható érték (μ) lesz**
- **a minta szórásból sokasági szórás (σ) lesz**
- **kvalitatív minta diverzitása átmegy a sokaság diverzitásába**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Alapsokaság statisztikai megoszlása

Kvalitatív **ismérv statisztikai megoszlását az ismérv változatainak (kategóriáinak) a populációbeli relatív gyakoriságával adjuk meg (pl. 20% - 50 %- 30 %).**

Kvantitatív **ismérv statisztikai megoszlását a gyakorisági megoszlással (sűrűségfüggvény) vagy a kumulatív gyakorisági megoszlással (eloszlásfüggvény) jellemezzük**

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

- **A sűrűségfüggvény *diszkrét* esetben az ismérv x_1, x_2, …,x_k,..lehetséges értékeinek *valószínűségeiből* (sokasági relatív gyakoriságok) áll: $p(x_1)$, $p(x_2)$, …, $p(x_k)$,.., vagy tömörebben, p_1, p_2,…,p_k, összegük 1. A „valószínűség” megjelölés itt azt jelenti, hogy ha például az X ismérv értéke az alapsokaság 30%-ban x_1, akkor egy véletlenszerűen kiválasztott egyed X értéke 30% valószínűséggel x_1 lesz.**

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

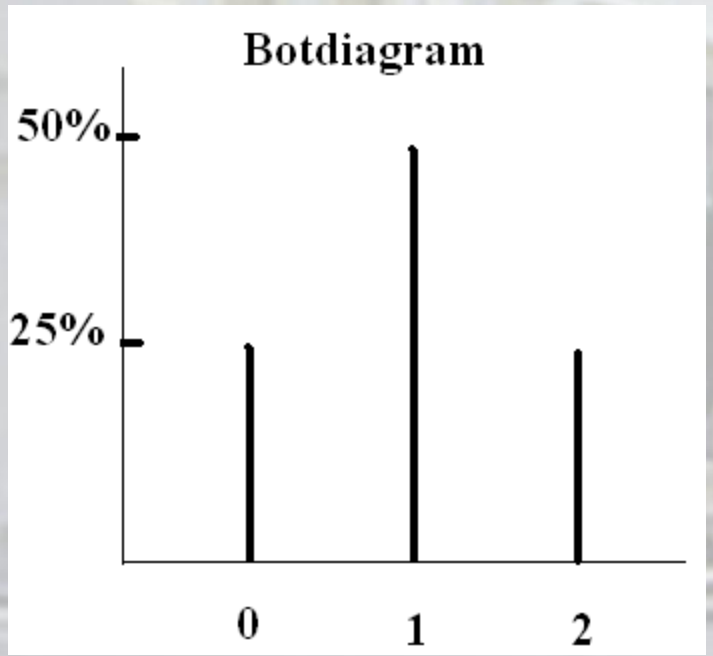
Példa diszkrét változó gyakorisági megoszlására

- Legyen az alapsokaság a kétgyermekes családok sokasága a földünkön a múltban, a jelenben, és a jövőben.
- Legyen X ismérv e családokban a fiúk száma, X lehetséges értékei 0, 1, 2,
- Kimutatható, hogy a kétgyermekes családok 25 %-ában nincs fiú, 50%-ában 1 fiú van, és 25%-ában mindkét gyermek fiú.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A példa folytatása

- **X valószínűség-eloszlása**



$P(X=0)=p_0=0,25=25\%$

$P(X=1)=p_1=0,50=50\%$

$P(X=2)=p_2=0,25=\underline{25\%}$

összesen 100%

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

- ***Folytonos*** **esetben az ismérv sűrűségfüggvénye egy nemnegatív p(x) vagy f(x) folytonos függvény, amely alatt a terület egységnyi. Ilyen például a jól ismert Gauss-féle haranggörbe.**
- **A sűrűségfüggvény lényege**

a sokaságnak az a részaránya, amely a és b érték közé esik, a sűrűségfüggvény alatti terület mérőszáma az (a, b) intervallum fölött,

képletben $$P(a \le x < b) = \int_a^b p(x)dx$$

Itt a „P” a probability (valószínűség) szóra utal.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Az *eloszlásfüggvény*, F(x)

- **az alapsokaság azon részaránya, amelybe tartozó egyedeken a szóban forgó X ismérv értéke x-nél kisebb. Más szóval, F(x) annak a valószínűsége, hogy egy véletlenszerűen választott egyeden X<x lesz, azaz**

$$F(x)=P(X<x).$$

Az alapsokaság (a,b) intervallumba tartozó egyedeinek részarányát a sűrűségfüggvénnyel és az eloszlásfüggvénnyel is kifejezhetjük:

$$P(a \le x < b) = \int_a^b p(x)dx = F(b) - F(a)$$

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Várható érték (sokasági átlag) és szórás

- **Az alapsokaság átlagát várható értéknek nevezzük, a továbbiakban μ-vel jelöljük, az alapsokaság szórásának jele σ. Ez az alapsokaság két legfontosabb *paramétere.* Képzésük a mintabeli megfelelőik értelemszerű kiterjesztésével történik:**

diszkrét esetben $\mu = \sum x_k \, p(x_k), \quad \sigma^2 = \sum (x_k - \mu)^2 \, p(x_k)$

folytonos esetben $\mu = \int_{-\infty}^{+\infty} x p(x) dx \quad \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \, p(x) dx$

- **μ a sűrűségfüggvény súlypontja.**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Kvalitatív változó jellemzői

- Kvalitatív sokasági átlagról nem beszélünk
- Variabilitását diverzitás mutatókkal mérhetjük.
- Legyenek az egyes kategóriákba sokasági relatív gyakoriságai p_1, p_2, ..., p_c, összegük 1 (100%)
- Simpson-Yule féle diverzitási index

$$D_{S-Y}=1-\sum p_k^2$$, maximális értéke 1-1/c

- Shannon-Weaver féle diverzitási index

$$D_{S-W}=-\sum p_k \ln(p_k),$$

maximális értéke ln(c), ahol c a kategóriák száma

(Mindkettő akkor maximális, ha $p_1= p_2= ...= p_c$)

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Ismeretek a várható értékről

A várható értéket **a továbbiakban μ szimbólum mellett E(.) –vel is jelöljük, tehát μ= E(X). Két alapvető tulajdonsága:**

$$E(a + c_1X_1 + c_2X_2 + \ldots) = a + c_1E(X_1) + c_2E(X_2) + \ldots$$

ahol $X_1, X_2, \ldots X_n$ tetszőleges véletlen változók és a, c_1, c_2 .. tetszőleges konstansok.

Speciálisan:

$$E(a)=a;\ E(cX)=cE(X);\ E(X+Y)=E(X)+E(Y);\ E(X-Y)=E(X)-E(Y)$$

A várható érték egy másik fontos tulajdonsága:

$$E(XY)=E(X)E(Y)$$ **, ha X és Y függetlenek**

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Ismeretek a sokasági varianciáról és szórásról

Sem a *szórás,* sem a *variancia* általában nem additívak

Ha viszont $X_1, X_2, \ldots X_n$ függetlenek, akkor

$$Var(a + c_1X_1 + c_2X_2 + \ldots) = c_1^2Var(X_1) + c_2^2Var(X_2) + \ldots$$

ahol a, c_1, c_2 .. tetszőleges konstansok.

Speciálisan:

$Var(a)=0$; $Var(cX)=c^2Var(X)$, és

ha X és Y függetlenek, akkor

$Var(X+Y)= Var(X)+Var(Y)$; $Var(X-Y)=Var(X)+Var(Y)$

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A sokasági átlag és variancia szabályainak néhány következménye a mintára vonatkozóan(1)

Felhasználva,hogy a minta elemei X_1, X_2, ...X_n független változók, igazolhatók az alábbiak

- A mintabeli relatív gyakoriság (f/n)
  - várható értéke azonos a sokasági relatív gyakorisággal (p)
  - varianciája pedig: Var (f/n) = p(1- p)/n

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A sokasági átlag és variancia szabályainak néhány következménye a mintára vonatkozóan(2)

- A minta átlagának ($\overline{X}$ a mintavétel előtt)
  - várható értéke azonos a sokasági átlaggal

    $E(\overline{X}) = \mu$
  - varianciája pedig:

    $Var(\overline{X}) = \sigma^2/n$
  - így az átlag szórása

    $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A sokasági átlag és variancia szabályainak néhány következménye a mintára vonatkozóan (3):

Két minta-átlag eltérésének várhatóértéke és szórása

- Tekintsünk két (idegen) sokaságot (1. és 2.), paramétereik μ_1 és σ_1 illetve μ_2 és σ_2.
- Vegyünk az 1. sokaságból n_1 elemű mintát, a 2.-ból n_2 eleműt, az átlagokat (a mintavétel előtt) jelölje rendre $\overline{X}$ ill. $\overline{Y}$.

Jelölje *D* a két átlag eltérését, ennek várható értéke és szórása jelentős szerepet kap a biometriai vizsgálatokban

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Két minta-átlag eltérésének… (folytatás)

Megmutatható, hogy

- az eltérés várható értéke $\mu_{\overline{D}} = E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$

- és a varianciája $\sigma^2_{\overline{D}} = Var(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

- Speciálisan

ha $\sigma_1 = \sigma_2 = \sigma$, akkor $\sigma^2_{\overline{D}} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2$

és ha emellett $n_1 = n_2 = n$, akkor $\sigma^2_{\overline{D}} = \left(\frac{2}{n}\right)\sigma^2$

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐Fontosabb sokasági megoszlások

- Diszkrét változók eloszlás-típusai
  - Binomiális eloszlás
  - Hipergeometrikus eloszlás
  - Poisson eloszlás
- Folytonos változók eloszlás-típusai
  - Egyenletes eloszlás
  - Exponenciális eloszlás
  - Normális eloszlás
  - „Normálisból származtatott” eloszlások

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Binomiális eloszlás

- Végezzünk n kísérletet, melyek mindegyikében p=P(A) eséllyel következik be a bennünket érdeklő „A” esemény és q=1-p eséllyel nem következik be (ilyen pl. a „visszatevéses mintavétel is véges sokaságnál)
- Legyen X az „A” bekövetkezésének száma az n kísérletből, X nyilván diszkrét véletlen változó, melynek lehetséges értékei 0,1,2,.., n. Az X változó eloszlását ***n, p paraméterű*** binomiális eloszlásnak nevezzük. Az X=k „esemény” valószínűségét p_k-val jelölve, kimutatható, hogy

$$p_k = P(X = k) = \binom{n}{k} p^k q^{n-k} \quad , \quad (k = 0, 1, 2, \ldots, n)$$

- X várható értéke és varianciája:

$$\mu = np \qquad \sigma^2 = npq$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Hipergeometrikus eloszlás

- Egy N elemű sokaságban legyen valamely „A” tulajdonságú egyedek száma S, ezek aránya p=S/N
- és „visszatevés nélkül” válasszunk ki n egyedet.
- Legyen X a kiválasztottak között az „A” tulajdonságúak száma. X diszkrét változó, melynek lehetséges értékei 0, 1, 2,.,(max)n. Az X véletlen változó eloszlását ***n,N,S paraméterű*** hipergeometrikus eloszlásnak nevezzük. Az X=k „esemény” valószínűségét p_k-val jelölve, kimutatható, hogy

$$p_k = P(X = k) = \frac{\binom{pN}{k}\binom{qN}{n-k}}{\binom{N}{n}}; \quad (k = 0,1,2,\ldots n) \qquad \mu = np, \qquad \sigma^2 = npq\left(1 - \frac{n-1}{N-1}\right)$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A ***Poisson eloszlás*** (ritka események eloszlása)

- a binomiális eloszlás határesete, ha n igen nagy és p pici. Ekkor az ***np*** =μ jelöléssel az X=k eset valószínűsége:

$$p_k = P(X = k) = e^{-\mu}\frac{\mu^k}{k!}, \qquad (k = 0,1,2,.....)$$

A Poisson eloszlású X valószínűségi változó várható értéke és szórásnégyzete egyaránt a μ paraméter.

Példa: ha egy területen bizonyos növény vagy rovaregyedek véletlenszerűen „szóródnak”, akkor az egységnyi területre eső X egyedszám Poisson eloszlású, μ az egységnyi területre eső átlagos egyedszámot jelenti

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Exponenciális eloszlás

- Alkatrészek élettartama, rovarok túlélési ideje a rovarirtó szer kipermetezésétől számítva (és általában véletlen időtartamok, távolságok) közelítően exponenciális eloszlásúak
- sűrűségfüggvénye $p(x) = \lambda \cdot e^{-\lambda x}$ *ha* $x \geq 0$ *különben* $p(x) = 0$
- eloszlásfüggvénye $F(x) = 1 - e^{-\lambda x}$ $(x>0)$
- várható értéke $1/\lambda$, szórása ugyanennyi
- Felezési időnek nevezzük azt a T értéket, amelyre

$$F(T) = ½, \text{ azaz } T = (\ln 2)/\lambda \approx 0{,}69/\lambda$$

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Normális eloszlás

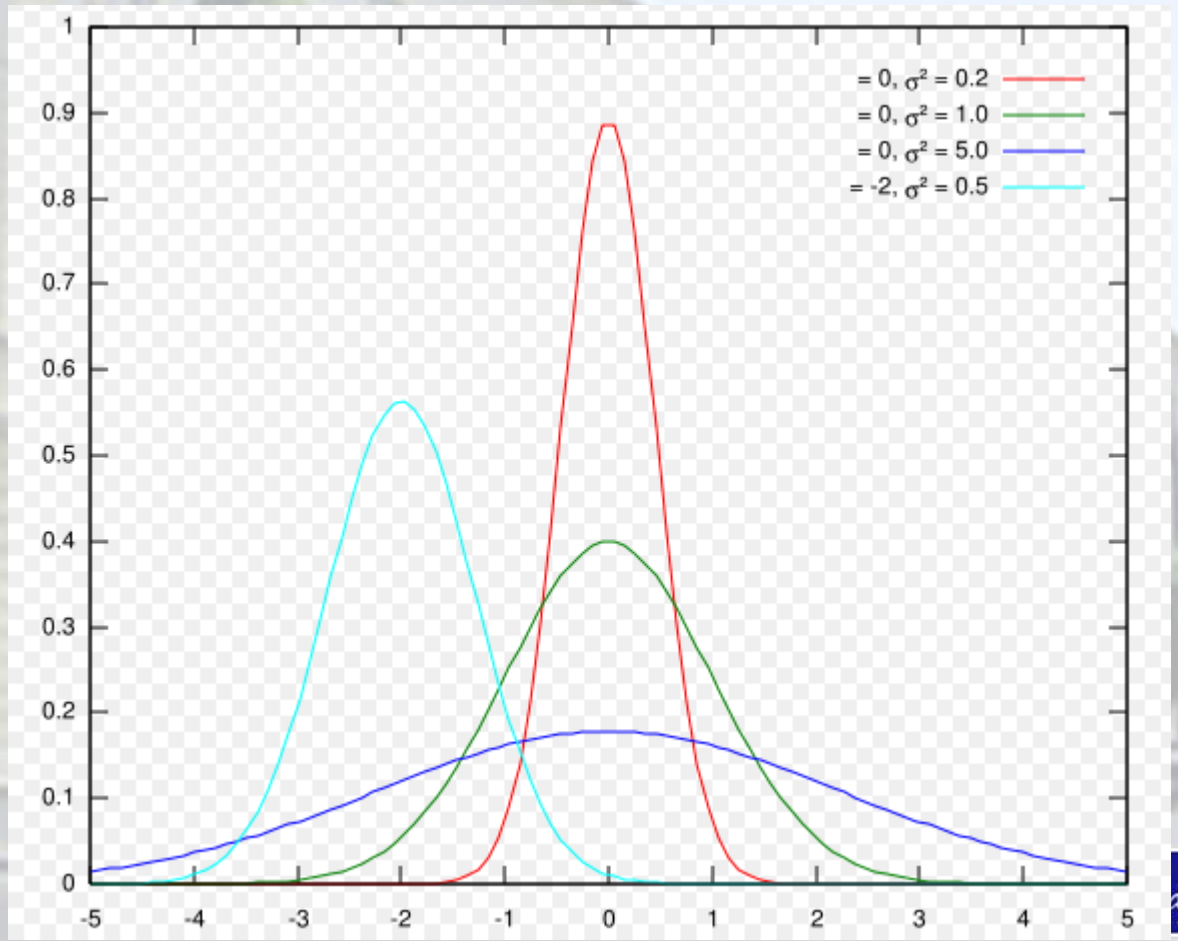
- A normális eloszlás a legfontosabb folytonos eloszlás
- sűrűségfüggvénye $p(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\cdot\frac{(x-\mu)^2}{\sigma^2}\right)$

- ahol μ és σ a normális eloszlású ismérv várható értéke ill. a szórása, képe a Gauss-féle haranggörbe

- A normális eloszlás-család tehát két-paraméterű, jelöljük N(μ, σ)-val.
- E családban a $\mu=0$ és $\sigma=1$ paraméterű esetet ***standard normális*** eloszlásnak nevezik. A sűrűségfüggvényét p(x) helyett konvencionálisan φ(u) -val jelölik, eloszlásfüggvénye pedig F(x) helyett Φ(u).

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Normális eloszlás sűrűségfüggvénye



A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Átszámítás ☞(u)-ból F(x) –re (Normális eloszlás folyt.)

- A Φ(u) és a ☞(u) függvény táblázatba foglalva megtalálható minden statisztika témájú könyvben (Excelből is kikereshető)
- Tetszőleges N(μ, σ) eloszlás eloszlásfüggvény értéke – F(x) – kiszámítható a standard normális eloszlásfüggvényből. Az „átszámítás”:

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- Eszerint egy N(μ, σ) eloszlású alapsokaságnak az (a,b) közbeeső egyedeinek részaránya:

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Átszámítás ☞(u)-ból F(x) –re (Normális eloszlás folyt.)

Az átszámítási formula szerint egy N(μ, σ) eloszlású alapsokaságnak az (a,b) közbeeső egyedeinek részaránya:

$$P(a \le x < b) = F(b) - F(a) = \Phi(u_b) - \Phi(u_a)$$

ahol

$$u_b = \frac{b - \mu}{\sigma} \quad \text{és} \quad u_a = \frac{a - \mu}{\sigma}$$

Megjegyezzük, hogy tetszőleges eloszlású X változó ***standardizáltjának*** nevezzük az

$$\overset{\frown}{X} = \frac{(X - \mu)}{\sigma}$$

változót. Ennek várható értéke mindig 0 és szórása 1

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Normális eloszlás(ok)ból képzett statisztikai eloszlások (1)

Véletlen változók függvényei is véletlen változók.

1) Lognormális eloszlásúnak nevezzük X változót, ha logX normális eloszlású.

2) n „független” standard normális eloszlású véletlen változó négyzetösszege ***n szabadságfokú chi² eloszlású*** valószínűségi változó, tehát:

$$\chi^2 = X_1^2 + X_2^2 + ... + X_n^2$$

ahol az X_i valószínűségi változók „független”, N(0,1) eloszlásúak. A függetlenség durván azt jelenti, hogy nincsenek kapcsolatban egymással (de erre még kitérünk).

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Normális eloszlás(ok)ból képzett statisztikai eloszlások (2)

- 3) A t-eloszlás

- Legyen X standard normális eloszlású és $\chi^2_{[n]}$ chi^2 eloszlású változó, legyenek függetlenek. Ekkor a

$$t_{[n]} = \frac{X}{\sqrt{\chi^2 / n}}$$

véletlen változó eloszlását ***n-szabadságfokú t-eloszlásnak*** hívjuk (Student-eloszlás)

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Normális eloszlás(ok)ból képzett statisztikai eloszlások (3)

- 3) Az F-eloszlás
- Két független chi2 –eloszlású valószínűségi változó legyen $\chi^2_{[m]}$ *és* $\chi^2_{[n]}$
- Ekkor az

$$F_{[m,n]} = \frac{\chi^2_{[m]} / m}{\chi^2_{[n]} / n}$$

- hányados F-eloszlású, m,n szabadságfokokkal.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ Paraméter becslés és konfidencia intervallum

- ***Paraméterbecslés***(1)
- Az alapsokaság valamely θ paraméterét (lehet ez μ, σ, ρ, regressziós állandók, stb.) minta alapján becsüljük. A becsült érték, a mintaelemek valamely T(X1, X2, …,Xn) függvénye. E függvényt igyekezni kell úgy választani, hogy várható értéke θ legyen ***(torzítatlanság***) és szórása a lehető legkisebb legyen.

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

- ***Paraméterbecslés***(2)
- **Ha pl. θ az alapsokaságban egy „A” tulajdonság relatív gyakorisága, θ = p =P(A), akkor a mintabeli relatív gyakoriság (f/n) torzítatlan becslése p -nek, hiszen E(f/n) = p.**
- **Ugyanígy, a mintaátlag az alapsokaság μ átlagának torzítatlan becslése, hiszen** $E\left(\overline{X}\right) = \mu$
- **Továbbá s^2 torzítatlan becslése σ^2 -nek**
- **Megmutatható, hogy mindhárom minimális szórású a lehetséges becslések között.**

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

- ***Paraméterbecslés***(3)
- A ***becslési elvek (kritériumok)***

közül a két leggyakrabban alkalmazottat említjük: a legkisebb négyzetek elvét (LN) és a legnagyobb valószínűség elvét (ML, maximim likelihood).

1. ***A legkisebb négyzetek*** **elvét használjuk többek között regressziós paraméterek meghatározásánál. Ha az alapsokaságban pl. lineáris összefüggést feltételezünk két ismérv, X és Y között,** $Y = \alpha + \beta X$ **akkor a paramétereket az yi- (a+bxi) eltérések négyzetösszegének minimálásával becsüljük, itt xi, yi az i-dik mintaelemnél kapott két ismérvérték,**

$$\hat{\alpha} = a; \quad \hat{\beta} = b$$

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

- 2. A ***maximum likelihood***

becslési elv lényege: θ becsléseként azt a értéket fogadjuk el, amely mellett a kapott (realizált) minta esélye a lehető legnagyobb.

- **Például a sokasági relatív gyakoriság (valószínűség) ML-becslése a mintabeli relatív gyakoriság**:

$\hat{p}$ =f/n, a mintabeli relatív gyakoriság.

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Konfidencia-intervallum
(megbízhatósági határok)

Egy sokasági paraméter becsült értéke még hibával terhelt, amit a szórása jelez. A becsült értékből az alapsokaság tényleges paraméterértéke csak hibahatáron belül állapítható meg.

Ezt a célt szolgálja a ***konfidencia-intervallum*** (alsó határa L (lower), felső határa U (upper)).
A θ paraméter pl. 95%-os konfidencia-intervalluma (L,U) egy olyan számköz, amely 95%-os valószínűséggel lefedi a valódi θ paramétert:

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Első példaként

képezzünk 95%-os konfidencia intervallumot egy N(μ, σ) eloszlású X sokaság μ várható értékére, legyen σ ismert.

Ekkor Δ=1,96 $\frac{\sigma}{\sqrt{n}}$ **, un. *hibahatár* jelöléssel a sokasági**

átlag (μ) 95% biztonsággal
L = mintaátlag – Δ és U = mintaátlag + Δ
közé esik .

Ha a szórás nem ismert, azt a mintából becsült szórással (s-sel) helyettesítjük és 1,96 helyett megfelelő „t értéket” írunk (ld. később).

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Második példaként

- **az alapsokaságbeli ismeretlen relatív gyakoriságra (p) keressünk konfidencia intervallumot.**
- **Legyen r=f/n a mintabeli relatív gyakoriság**
- **Ha a mintanagyság (n) legalább 10, p-nek a 95% -os konfidencia határai (L,U) – jó közelítéssel - az alábbi, p -ben másodfokú egyenlet két gyöke**

$$n(r - p)^2 = 3{,}84p(1 - p)$$

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

☐ Statisztikai következtetés: Hipotézis vizsgálat, statisztikai próbák

A hipotézisvizsgálat elve (1)

- **A statisztikai hipotézisvizsgálat arra irányul, hogy az alapsokaság(ok)ra vonatkozóan megfogalmazott feltevéseket minta alapján ellenőrizzük, elfogadjuk, vagy elvessük.**
- **A kísérlet (megfigyelés) *előtt* kérdéseket fogalmazunk meg az alapsokaságra vonatkozóan,**
- **majd ezeket formálisan hipotézisekbe öntjük:**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A hipotézisvizsgálat elve (2)

- **Bármi is az igazolni kívánt hipotézis, először meg kell fogalmaznunk a H_0, u.n. *null-hipotézist***
- **A null-hipotézist mindig tagadó értelemben fogalmazzuk:**
  a kezelésnek nincs hatása, két alapsokaság átlaga nem különbözik, két ismérv nem korrelál, stb.
- **A H_0 munkahipotézishez u.n. *ellenhipotézist* csatolunk, H_1, ez általában a H_0 egyszerű tagadása, néha viszont az ellenhipotézis valamely irányú egyenlőtlenséget fejez ki, pl. $\mu_2>\mu_1$ (a 2.sokaság átlaga nagyobb az 1.sokaság átlagánál) /ld. később: *egy-* illetve *kétoldali próba/***

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A hipotézisvizsgálat elve (3)

- **A hipotézisvizsgálathoz mintát veszünk, adatokat kapunk**
- **Az ellenhipotézist is figyelembe véve, kiszámítjuk a kapott- és annál szélsőségesebb minták együttes esélyét, ha a null-hipotézis igaz**
- **Ha ez az esély (*P*) túl kicsi, *elutasítjuk* a H_0 hipotézist és elfogadjuk a H_1 hipotézist**
- **Ha *P* „elég nagy”, akkor *elfogadjuk* a H_0 hipotézist**
- **Azt, hogy mely *P* értéket tekintjük elég kicsinek, a kutató dönti el a vizsgált kérdéstől függően. Konvenciális értékei α = 5%(=0,05) vagy 1%(=0,01) vagy 0,1%(=0,001). α neve: *szinifikancia szint***
- ***Szignifikanciá*ról beszélünk, ha elutasítjuk a H_0-t, de hozzá kell tennünk, hogy mely α „hibaszinten”**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A hipotézisvizsgálat elve (4): döntési hibák

- ***Mivel a minta estetleges, a statisztikai döntés mem abszolút érvényű, hibás lehet, erre utal a „szignifikáns” jelző***
- ***a statisztikai tévedés két fajtája: az első- és a másodfajú hiba.***
- ***Tévedhetünk úgy, hogy az alapsokaságban H0 igaz, mégis elutasítjuk, ennek esélye α (első fajta hiba),***

  és úgy is, hogy a hamis nullhipotézist elfogadjuk (második fajta hiba), ennek esélye β, értéke függ attól, hogy H0 helyett pontosan mi igaz

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A hipotézisvizsgálat elve (5): Modell-példa

- **➢ Vizsgáljuk egy kistelepülésen az újszülöttek között a fiú:leány arányt.**
- **a H_0 null-hipotézis: a fiú:lány arány 50:50%**
- **A minta: a település szülőotthonában adott hónapban 1 leány és 7 fiú születik (n=8)**
- **A *P* esély itt egyszerű valószínűségszámítási meggondolással közvetlenül számítható**

  A) egyoldali próba
- **Ha az ellenhipotézis (H_1) az , hogy a településen több fiú születik mint lány (egyoldali ellenhipotézis), akkor a mintánál szélsőségesebb csak az az eset, hogy mind a 8 újszülött fiú, azaz *P* = P(0 vagy 1 leány)**

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

- A modell-példa folytatása

- **A leányok száma a mintában Binomiális eloszlású n=8 és p=0,5 paraméterekkel, eszerint**

  P **= P(0 vagy 1 leány) =** $0{,}5^8 + 8\times 0{,}5^8 = 0{,}035 = 3{,}5\%$

  - mivel 3,5% < 5%, a H_0 **hipotézist** $\alpha = 5\%$**-os szignifikancia szinten elutasítjuk és a** H_1 **hipotézist fogaduk el: a településen szignifikánsan több fiú születik, mint leány**

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

A modell-példa folytatása

B) kétoldali próba

- **Ha az ellenhipotézis (H_1) az , hogy a településen nem 50%:50% az újszülöttek fiú:leány aránya (kétoldali ellenhipotézis),**

- akkor figyelembe kell venni a „legfeljebb egy fiú” esetet is, így

***P* = P(0 vagy 1 leány) + P(0 vagy 1 fiú)**

= 2× 0,035 = 0,07 = 7%

Mivel P>5%, elfogadjuk a fele fiú, fele leány hipotézist

Új Magyarország Fejlesztési Terv

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Hipotézisvizsgálat (6)

- a *P* hiba-esély kiszámítása ritkán megy közvetlenül
- általában a mintaelemekből először képezünk egy alkalmas függvényt (próba függvény, statisztika, ST(.))
- e statisztika (mint véletlen változó) eloszlása H_0 fennállásának feltételezésével meghatározható
- kiszámoljuk az ST statisztikát a kapott mintára, majd - az ellenhipotézist is figyelembe véve - megállapítjuk annak esélyét, hogy H_0 fennállása estén ST legalább olyan szélsőséges érték, mint amit mintából számoltunk, ez *P*
- az eljárásokra szoftverek állnak rendelkezésre

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Hipotézisvizsgálat (7): példa

- Illusztrálásként vizsgáljuk egy bizonyos „kezelés” hatását n mintaegyeden. Az i-edik egyeden a jelzőérték legyen a kezelés előtt x_{0i}, utána x_{1i}, a növekmény $x_i = x_{1i} - x_{0i}$
- Tegyük fel, hogy $\{x_i\}$ az $N(0,\sigma)$ eloszlású alapsokaság egy reprezentációja
- Az ismeretlen σ szórást az x_i =adatokból becsüljük, s
- A kezelés hatástalan volta esetén az $X=X_1-X_0$ v.változó várható értéke μ=0, ez a H0. H0 fennállása esetén a

$$t = \frac{\overline{X} - 0}{s/\sqrt{n}}$$

statisztika n-1 szabadságfokú t-eloszlású v. változó

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Hipotézisvizsgálat (8): a példa folytatása

- Kiszámítva a t-értéket a mintából és a számított értéket összehasonlítva a *t*- táblázatbeli α-szintű kritikus értékkel, megítélhetjük a kezelés-hatás szignifikanciáját
- Legyen például n=20, és t=1,9. Mivel a táblázatbeli érték kétoldali próba esetén (azaz H_1: $\mu \neq 0$) α= 5%-os szinten 2,09, és ennél 1,9 kisebb, elfogadjuk a H_0 hipotézist (az eltérés nem szignifikáns!),
- Ha viszont az ellenhipotézis H_1: $\mu > 0$ (azaz jó okunk van arra, hogy pozitív kezeléshatást feltételezzünk), akkor egyoldali próbát alkalmazunk, elfogadjuk a H_1:hipotézist, mert t táblázatbeli értéke α= 5%-nál 1,78, ennél 1,9 nagyobb, a kezelés hatása tehát szignifikáns

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Ellenőrző gyakorlatok

- Vegyünk fel legalább n=10 elemű pozitív mintát (adatot). Számoljuk ki e minta jellemzőit (átlagok, átlagos eltérés, szórás, az átlag hibája, relatív szórás). Rakjuk nagysági sorrendbe az átlagokat. Ellenőrizzük a $\sum (x_i - \bar{x})^2 = \sum x_i^2 - n(\bar{x})^2$ egyenlőséget.
- Legyen x1>0 és x2=x3=….=xn=0. Mutassuk meg, hogy ez esetben $s^2=n(\bar{x})^2$.
- Egy n=25 elemű mintában f=10 egyed rendelkezik egy „A” tulajdonsággal. Adjuk meg a mintabeli relatív gyakoriságot és ennek hibáját (szórását).
- Egy diszkrét kvantitatív ismérv lehetséges értékei 0, 1, 2. Ezek részaránya az alapsokaságban p0=0,20=20%, p1=0,30=30%, p2=0,5=50%. Számoljuk ki az ismérv várható értékét és szórását.

Új Magyarország Fejlesztési Terv
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg

Ellenőrző gyakorlatok (folytatás)

- Írjuk fel az n=2, p=0,6 paraméterű binomiális eloszlás p0, p1, p2 tagjait. Mennyi μ és σ?
- Egy N(12;2) eloszlású alapsokaság egyedeinek hány %-a esik a (8;12) intervallumba? ($\Phi(1) = 0{,}841$)
- Legyenek X1, X2, …,Xn azonos eloszlású független v.-változók μ és σ paraméterekkel, továbbá c1, c2, …, cn konstansok, melyek összege 1. Igazoljuk, hogy az $Y= c_1X_1+ c_2X_2+\ldots+c_nX_n$ súlyozott átlag várható értéke μ, varianciája $\blacklozenge^2\blacklozenge c_i^2$.
- Számoljuk ki a normális eloszlású v. változó μ várható értékének 95%-os megbízhatósági intervallumát, ha n = 10 elemű mintából $\bar{x} = 5$ és s=2.
- Adjunk az alapsokaság valamely p arányára 95%-os konfidencia intervallumot, ha n=10 mintegyedből a relatív gyakoriság r=0,4.

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg