

# Bevezető: Változók, minták, kísérleti elrendezések

BME Számítógépes és Kognitív  
Idegtudomány, Statisztika és  
Kísérlettervezés

# Tankönyvek, hasznos weboldalak

- Goss-Sampson, *Statistical Analysis in JASP: A Guide for Students*
- Navarro et al, *Learning Statistics with JASP*
- Szokolszky Ágnes, *Kutatómunka a pszichológiában.*  
(feladatlap is van hozzá)
- <http://www.mathsisfun.com/> (data)
- <http://students.brown.edu/seeing-theory/>
- <http://www.socscistatistics.com/>
- <http://www.graphpad.com/guides/prism/6/statistics/>
- <http://distributome.org/V3/calc/index.html>

# Amire szükség lesz

- Laptop (lehet hozni az előadásra is)
- Excel vagy Google Sheets
- JASP (Jeffrey's Amazing Statistical Package)
- Előadásdiák, stb:

[http://www.cogsci.bme.hu/~ktkuser/KURZUSOK/BMETE47MC38/2019\\_2020\\_1/Ea/](http://www.cogsci.bme.hu/~ktkuser/KURZUSOK/BMETE47MC38/2019_2020_1/Ea/)

# Kurzuskövetelmények

- 2 zh
  - adatelemzés és eredmény jelentése
  - laptopon
- Labormunka

Aki mindkét zh-t elégtelennél jobban teljesíti, és meg van elégedve a jegyével, annak nem kell vizsgáznia.

- Egyéb esetben vizsga a vizsgaidőszakban

# Tervezett témák

Hét	Téma
1	Hipotézistesztelés, populáció és minta, változók
2	Disztribúció, átlag & szórás, kiugró értékek, grafikus ábrázolás
3	Normál eloszlás, z-értékek, konfidencia intervallumok, hatásméret, statisztika erő
4 (okt 2.)	Dékáni szünet
5	Korreláció
6	Regresszió
7 (okt 23)	szünet
8 (okt 30)	1. zh
9	t próbák
10	ANOVA és GLM 1
11	ANOVA and GLM 2
12	ANOVA and GLM 3
13	Nem-paraméteres tesztek
14 (dec 11.)	2. zh
Pót hét (dec 18.)	pótzh

# Mai témák

- Deduktív és induktív következtetés
- Minták és populációk
- Leíró és inferenciális statisztika
- Sampling error
- Függő és független változók
- Független mintás és ismételt méréses elrendezés
- Hipotézistesztelés
- Miért nem lehet a számítógépre bízni a statisztikát?
- Mérési skálák

# Indukció és dedukció

- Dedukció: logikán alapuló következtetés:  
Minden ember halandó. Szókratész ember.  
Tehát:
- Indukció: megfigyelésen alapuló következtetés:  
amikor sok fagyalt fogy, megnövekszik a  
fulladásos halálok száma (????)

# A statisztika

- A megfigyelésen alapuló következtetés segédeszköze: általánosítható-e a megfigyelés tapasztalata
- Numerikus adatok elemzésére alkalmas
- A megfigyelésnek numerikus adatokat kell létrehoznia



# Egy probléma

- Hatással van-e a stressz a kognitív teljesítményre
  - Természetes megfigyelés
  - Kvázi-kísérlet (természetes csoportok tesztelése)
  - Kísérlet (a kísérleti csoportban a stressz mesterséges előidézése)
- Mindegyik esetben **mintákat** mérünk
  - Mert mindenkit nem tudunk lemérni
  - De nem a minta, hanem a **populáció** érdekel (valamennyi olyan entitás akiről/amiről szeretnénk megtudni valamit)

# A statisztika két ága

## **Leíróstatisztika**

A minta leírására szolgál

- Átlag
- Min-max
- Százalék

## **Inferenciális statisztika**

- A minta leíróstatisztikáiból indul ki
- És matematikai módszerekkel megpróbálja megbecsülni a populáció jellemzőit (paramétereit)

# Miért nem olyan egyszerű a becslés

- Egy minta sem tudja tökéletesen reprezentálni a populációt (sampling error)
  - A magyar „mintavételi hiba” kifejezést egy kicsit másképp használják
  - Valójában nem “hiba” – abban az értelemben, hogy nem lehet kiküszöbölni
  - Inkább természetes változatosság
- Amit tehetünk a mintavételi hiba csökkentése érdekében: **random mintavétel**

# Hogyan lesz a megfigyelésből adat?

- A mintánk tagjait valamilyen módon lemérjük. A mérés eredménye a **függő változó** (pontszám, kimenet).
  - Pl. a kognitív teljesítmény mérésére csináltatunk velük egy tesztet, ahol egy sor szót fordított sorrendben kell visszamondani. 4 szóval kezdjük, ha az sikerült, jön 5, stb. A leghosszabb sikeresen visszamondott szó sor hossza a pontszám.
- Függő változónak hívják, mert többnyire azt feltételezzük, hogy az elért pontszám a körülményektől függ. A körülmény, amit a kutató manipulál vagy határoz meg, a **független változó**.
  - Pl. Ha az a feltételezés, hogy a stressz befolyásolja a teljesítményt, a stressz szintje a független változó.

A független változót két módon manipulálhatjuk:

**1. Ismételt méréses elrendezés:**

Ugyanazokat a személyeket a független változó több értéke mellett is lemérjük: pl. egy órával egy nagyon fontos vizsga előtt és a nyári szünetben (**kondíciók**)

**2. Független mintás elrendezés:**

A független változó különböző értékei mellett más-más személyeket mérünk: egy csoportot egy órával a vizsgájuk előtt, egy másik csoportot a nyári szünetben (**csoportok**)

Mindkét esetben a független változó különböző szintjei esetén elért pontszámokat fogjuk összehasonlítani, de a statisztikai eljárás más lesz.

# Hipotézistesztezés

- Elmélet (pl. *a stressz befolyásolja a kognitív teljesítményt*)
- **Hipotézis:** Az elméletből következő feltételezés, miszerint két (vagy több) változó között valamilyen kapcsolat áll fenn (pl. *a stressz szintje összefügg a memóriateszten elért eredménnyel*).
  - A hipotézis lehet **egyszélű**: amikor az összefüggés irányát is megjósoljuk (*a stressz javítja a memóriateljesítményt*)
  - vagy **kétszélű**: amikor az összefüggés irányát nem jósoljuk meg (*a stressz valamilyen módon befolyásolja a memóriateljesítményt*)

- Cél: megtudni, hogy a hipotézis beigazolódik-e, vagyis támogatják-e az eredmények az elméletet.
- A minta lemérése és statisztikai elemzés után kapunk egy számot: **p érték** (p mint probability, valószínűség)
  - a p közel-0 és 1 között lehet.
- A p érték annak a valószínűségét mutatja, hogy NEM áll fenn a megjósolt összefüggés a változók között -> ez a **null hipotézis**

# Döntés

- A számítógép kidobja a  $p$  értéket
- De az embernek kell eldöntenie, hogy elég kicsi-e ez a  $p$  érték ahhoz, hogy elvesse a null hipotézist (vagyis támogatottnak tekintse az elméletet)
- Mikor “elég kicsi” a  $p$  érték?
- Hagyományosan 0,05 vagy alatta számít elég kicsinek: **statisztikailag szignifikáns** az eredmény (pl. *a különbség a stressz alatt álló és a kontrollcsoport memóriateljesítménye között*)
  - Ha a  $p = 0,05$ , az annyit jelent, hogy 5% annak az esélye, hogy nincs összefüggés a változók (*a stressz és a memória*) között.
  - A döntés múlik azon, hogy milyen következményei lehetnek a hibázásnak
  - A legszigorúbb küszöb jellemzően: 0,001, vagyis egy ezrelék.
- A hipotézist bebizonyítani nem lehet, mert mindig van egy kicsi esély arra, hogy mégis a null hipotézis a helyes.



# Lehetséges kimenetek

	Null hipotézis elvetése	Null hipotézis megtartása
Null hipotézis igaz	1. Típusú hiba (Téves riasztás)	Helyes döntés (Helyes elutasítás)
Null hipotézis hamis	Helyes döntés (Találat)	2. Típusú hiba (Kihagyás)

# MÉRÉSI SKÁLÁK

# A függő és független változók típusai

Két matematikai szempontból fontos dimenzió mentén különböznek egymástól:

1. név vagy szám
2. van-e természetes sorrendjük az értékeknek  
vagy sem

# Név-skálák

- Nominális (kategorikus)
  - kategorizál, diszkrét csoportokra oszt
  - pl. születési hely, családi állapot, kedvenc sütemény
- Ordinális
  - diszkrét csoportokra oszt, amiknek van egy természetes sorrendje
  - pl. versenyen elért helyezés, rang egy rangsorban, osztályzat (?)
  - Likert skála: *soha - néha - gyakran - állandóan*

# Szám-skálák

- Intervallum
  - egy jól definiált mértékegység mentén mér, tehát két egymást követő érték között mindig ugyanannyi a különbség
  - pl. hőmérséklet, személyiség teszten elért pontszám (?)
  - Likert skála: 0-től 100-ig skálán mennyire szereti a fagyaltot (???)
- Arányskála
  - mint az intervallum + a 0 értéknek van értelme, tehát két érték arányának is van értelme
  - pl. emlékezeti teszt pontszám, reakcióidő

# Praktikusan

- Intervallum skálának fogjuk tekinteni
  - az intervallum skálát
  - az arányskálát
  - a sok numerikus értékkel rendelkező Likert skálát
  - a JASP ezeket **Continuous** skálának nevezi
- Nominális skálának tekintjük a nominális skálát, de két típusa lesz:
  - amit szó jelöl (Nominal text)
  - és amit szám jelöl (Nominal) (1: férfi, 2: nő)
  - néha nominális skálaként kezeljük a kevés értékkel rendelkező ordinális skálát is (Ordinal)

# A mérési skála típusa sokszor döntés kérdése

- A memóriateszten elért eredmény: helyesen visszamondott szavak száma - Intervallum skála
- De a stressz szintje lehet:
  - vizsga előtt kontra nyári szünetben: nominális
  - nagyon stresszelt > kicsit stresszelt > nem stresszelt: ordinális-nominális
  - kérdőívvel 1-től 100-ig skálán: ordinális-intervallum
  - hormonszint méréssel: intervallum

# Házi

Google Sheet:

<https://docs.google.com/spreadsheets/d/11eZgHWgEL4iG2QNyv-8asCtLOnnY6oiZJTEWYWK4ePM/edit?usp=sharing>

Neptun: saját Neptun kód, ellenőrzés végett

ksz: egy szám, ami majd azonosítja az embereket. **MÁS AZONOSÍTÁS ETIKAI OKOKBÓL TILOS!**

nem: férfi - nő - egyéb

magasság: cm

cipő: EU méret

hajszín:

szemszín:

Facebook: Facebook ismerősök száma

alvásKedd: hány órát aludt a kérészt megelőző keddről szerdára virradó éjszaka

alvásSzombat: hány órát aludt a kérészt megelőző szombatról vasárnapra virradó éjszaka

buliSzombat: elment-e valahova a kérészt megelőző szombat este: igen - nem

alkohol: hány egység alkoholt fogyasztott a kérészt megelőző hét napon összesen

1 egység alkohol: 1 korsó sör, 1 pohár bor, fél deci vodka, stb.