

# Elemi stat.

## 1. Bevezetés a statisztikába

### Áttekintés:

Adatok - összegyűjtött megfigyelések (mérések, kérdőíves válaszok, felmérések)

Statisztika - adatokon alapuló kísérlettervezési, gyűjtési, rendezési, összesítési, ábrázolási, analízisi, értelmezési és következtetési módszerek összessége

Populáció (alapsokaság) - a tanulmányozandó elemek összessége, teljessége (pl. eredmények, mérések, stb.). A gyűjtemény teljes abban az értelemben, hogy tartalmaz minden tanulmányozandó tárgyat.

Cenzus - adatok gyűjteménye a populáció minden eleméről

Minta - a populációból kiválasztott elemek rész halmaza

Kulcsfogalmak

\_ A mintát megfelelő módon kell gyűjteni, mint amilyen a véletlen kiválasztás.

\_ Ha az adatok nem így lettek gyűjtve, akkor általában statisztikai módszerekkel sem lehet ezt kijavítani, az adatokat nem lehet használni.

### Adatok típusai:

Paraméter - a populációt jellemző numerikus érték

Statisztika - a mintát jellemző numerikus érték.

Kvantitatív adatok méréseket vagy leszámításokat jellemző számok. Pl.: az emberek súlya A kvantitatív adatokat tovább bonthatjuk diszkrét és folytonos típusokra

Kvalitatív (kategória vagy tulajdonság) adatok - kategóriákra bonthatók, melyeket valamilyen nemnumerikus jellemzők alapján különböztethetők meg Példa: profi atléták nemei (férfi/nő).

Diszkrét - amikor a lehetséges adatok száma véges vagy legalábbis megszámlálható. 0, 1, 2, 3, . . . Példa: Tyúkok által tojt tojások száma.

Folytonos - (numerikus) adat ami végtelen sok lehetséges értéket vehet fel valamilyen folytonos skálán, és nincsenek benne lyukak, szakadások. Pl.: A tehén által naponta adott tej mennyisége (8.86965517 liter) .

A mérések szintjei

\_ Nominális – csak kategóriák

\_ Ordinális – kategóriák és rendezhetőség

\_ Intervallum – különbségeknek van értelme, de nincs természetes 0 pont

\_ Arány – a különbségeknek és arányoknak van értelme és létezik természetes kezdőpont

### Kritikus gondolkodás:

Buktatók

\_ Hibás minták

\_ Kicsi minták

\_ Félrevezető ábrák

\_ Becsapós ábrák

\_ Játék a százalékokkal

\_ Beugrató kérdések

\_ A kérdések sorrendje

\_ Válasz megtagadás

\_ Korreláció és kauzalitás

\_ Önérdelkség a vizsgálatban

\_ Precíz számok

\_ Részleges képek

\_ Készakart hamisítás

### A kísérletek megtervezése:

Megfigyeléses vizsgálat (Observational study) - bizonyos jellemző tulajdonságok megfigyelése és mérése anélkül, hogy megváltoztatnánk a vizsgálat tárgyát/alanyát pl.: közvéleménykutatás, csillagászati/asztrofizikai megfigyelése

Kísérlet (Experiment) - valamilyen kezelést végzünk és azután megfigyeljük a hatásait a kísérlet tárgyán/alanyán Pl.: klinikai gyógyszervizsgálat, részecske ütközések a CERN gyorsítójában

Keresztmetszeti vizsgálat (Cross Sectional Study) - Az adatokat egy időpontban mérjük, figyeljük meg és gyűjtjük be.

Utólagos vizsgálat (Retrospective Study) - Múltbéli adatokat használunk. (pl.: az autóbalesetben meghaltak és nem abban meghaltak összehasonlítása)

Előre tervezett (Prospective Study) - Az adatokat a jövőben gyűjtjük, olyan csoportokból, melyek valamilyen közös faktorban megegyeznek. (pl.: a mobil telefont használó és nem használó vezetők csoportjainak összehasonlítása)

Zavar (bezavarás) - akkor lép fel egy kísérletben, ha a kísérletet végző nem tudja megkülönböztetni az egyes faktorokat Pl.: Mindenkitől levonunk 1 pontot, ha nem jelenik meg az előadáson, javul-e a részvételi arány? Tfh. hogy javul. De lehet, hogy idén jobb volt az időjárás. A két faktor nem különböztethető meg.

#### A változók hatásának kontroll alatt tartása

Vak vizsgálat (Blinding), duplán vak vizsgálat - a vizsgálat alanya nem tudja, hogy kezelést kape vagy placebót, duplán vak, ha a kísérletező sem tudja (pl.: a gyermekbénulás Salk vakcina kipróbálása az USA-ban 1954-ben)

Blokkosítás – felosztjuk a populációt olyan alcsoportokra amelyekben a kísérlet szempontjából fontos tulajdonságai megegyeznek . Mindegyik blokkban véletlenszerűen választjuk ki a kezelteket

Teljesen randomizált (véletlenszerűsített) kísérleti elrendezés - véletlen kiválasztással választjuk ki azokat, akik kezelést kapnak

Szigorúan kontrollált elrendezés - nagyon körültekintően kiválasztott egyedek pl.: ha pl. vérnyomáscsökkentőt tesztlünk, akkor ha az egyik blokkban van egy 30 éves túlsúlyos cigarettázó férfi, aki szereti a sós és zsíros ételeket, akkor a másik blokkba is teszünk ilyen

#### Ismétlés és a minta mérete

Ismétlés - a kísérlet megismétlése, amikor van elegendő alany ahhoz, hogy észrevehessük a különböző kezelések közti eltéréseket

Minta mérete - akkora mintát kell használni, ami elég nagy ahhoz, hogy kimutathassuk benne az effektust

Véletlen mintavétel - a populáció minden tagjának ugyanakkora esélye van arra, hogy a mintába bekerüljön

Egyszerű véletlen mintavétel (n hosszúságú) - a minta tagjait úgy választjuk ki, hogy bármelyik n hosszúságú mintának ugyanakkora a kiválasztási esélye

Véletlen számok generálása

Szisztematikus mintavétel - Valamilyen kezdőponttól indulva kiválasztjuk minden K adik elemet a populációból

Kényelmes mintavétel - használjuk azt a mintát, amit a legkönnyebb beszerezni

Rétegzett mintavétel - oszd fel a populációt kettő vagy több csoportra (rétegre), melyeken belül bizonyos (a kísérlet szempontjából fontos) tulajdonságok azonosak vagy hasonlóak, majd vegyünk mintát mindegyik rétegből

Klaszter mintavétel - oszd a populációt valamilyen természetes módon klaszterekre; véletlenül válassz kö

Mintavételi hiba (Sampling error) - a minta és a populáció eredménye közti eltérés, ami a minták fluktuációjából származik

Nem mintavételi hiba (Non-sampling error) - olyan eltérés, ami az inkorrekt adatgyűjtésből, adat felvitelből vagy analízisből ered

## **2. Az adatok leírása, megismerése és összehasonlítása**

### **Áttekintés:**

Leíró statisztika - a legfontosabb tulajdonságok összegzése vagy leírása a populáció egy ismert részhalmazán

Következtet statisztika - a minta felhasználásával következtetések (vagy általánosítások) a populációról

#### Az adatok legfontosabb tulajdonságai

1. Középpont (center): Egy reprezentáns vagy átlag érték, ami megmutatja, hogy hol van a közepe az adathalmaznak
2. Szóródás: Annak a mértéke, hogy mennyire szóródnak az adatok
3. Eloszlás (distribution): Az adatok eloszlásának természetete, alakja (pl.: harang görbe, egyenletes, ferde)
4. Kiugró adatok (outliers): Olyan mintaértékek, melyek a mintaértékek döntő többségétől messze helyezkednek el
5. Idő: Az adatok idő ben változó tulajdonságai

### **Gyakorisági eloszlások:**

Gyakoriság eloszlás (frequency distribution) - az adat értékek listája (vagy egyenként vagy mint intervallum csoportok), együtt a gyakoriságukkal vagy számukkal

Alsó osztálykorlát - a legkisebb szám, ami még az adott osztályba kerülhet

Felső osztálykorlát - a legnagyobb szám, ami egy osztályba bekerülhet

Osztályhatárok - azok a számok, amelyek elválasztják az osztályokat egymástól, de az osztálykorlátok által kreált hézagok nélkül

Osztályhatár - az osztályokat elválasztó számok

Osztály felező pont - az osztályok felező pontjai az osztály felező pontot a felső és alsó osztálykorlátok összege elosztva kettővel adja meg

Osztályszélesség - két egymás utáni alsó osztálykorlát vagy osztályhatár különbsége