

STATISZTIKA

STATISZTIKAI ALAPFOGALMAK

Statisztikai alapfogalmak (1)

- Sokaság, populáció, véletlen kísérlet
- Statisztikai minta, minta realizáció
- Statisztikai mintavétel

Statisztikai sokaság, populáció

A vizsgálat tárgyát képező nagyszámú de véges elemszámú egyedek halmaza.

Cél:

- (1) a halmaz egészének **kevés paraméterrel** történő tömör jellemzése,
- (2) a populáció egyedeinek leírására bevezetett **paraméterek közötti kapcsolatok** feltárása.

Arra **általában** nincs lehetőség (erőforrás), hogy a populáció minden egyes eleméről adatokat szerezzünk be.

- Magyarország állampolgárai
- Egy egyetemi kar hallgatói
- Az érvényes forgalmival rendelkező autók halmaza
- Egy adott termék vásárlóinak halmaza
- Egy TV csatorna nézőinek halmaza

Statisztikai sokaság, populáció

Egy véletlen kísérlet megfigyelése

A statisztikai elemzés tárgya lehet egy **véletlen kísérlet** is, ami **időben változatlan körülmények** között elvileg akárhányszor lejátszódhat.

- A lottóhúzás
- Egy szerver működése
- Budapest januári átlaghőmérséklete
- Egy gyümölcsös termés hozama
- Egy új gyógyszer hatása
- Egy reklámkampány hatásossága
- Egy populáció egyedének véletlen kiválasztása

Statisztikai minta

A populáció egy kis elemszámú részhalmazára vonatkozó megfigyelések adatai.

A minta reprezentatív kell, hogy legyen.

- Egy felmérésbe bevont magyar állampolgárok halmaza
- Egy adott előadásra belátogatott hallgatók halmaza
- Adott biztosítóval szerződött autók halmaza
- Egy adott napon megkérdezett vásárlók halmaza
- Egy nézettségi felmérésbe bevont TV nézők halmaza
- Budapest januári középhőmérsékleteinek adatai

□ A statisztikai sokaság típusai:

- **Álló sokaság:** állapot, időpont jellegű; megfigyelése mindig adott időpontban végezhető el. Pl. a Mezőgazdasági Kar hallgatóinak létszáma 2015. szeptember 7.-én.
- **Mozgó sokaság:** időben változó sokaság, egy folyamatot érzékel, időintervallum alatt figyelhető meg. Pl. a Mezőgazdasági Kar 2015/2016. I. félévében vizsgát tett hallgatóinak a száma.
- **Diszkrét sokaság:** elemei jól elkülöníthetők egymástól, pl. hallgatók;
- **Folytonos sokaság:** elemei nem különülnek el, értékeit általában intervallumban adják meg, pl. a hallgatók magassága;
- **Véges sokaság:** véges sok elemből áll, pl. a népesség adott időben, térben;
- **Végtelen sokaság:** végtelen sok elemből áll, vagy olyan sok elemből, hogy végtelennek célszerű tekinteni;
- **Egynemű sokaság:** azonos elemekből áll, nem bontható részekre, pl. a homokbánya homokja;
- **Összetett sokaság:** az elemeknek nemcsak közös, hanem megkülönböztető tulajdonságai is vannak, pl. hallgatók;
- **Valós sokaság:** ténylegesen előforduló elemekből áll.
- **Elméleti sokaság:** az adott eseményre vonatkozó bekövetkezések összes lehetséges kimenetele alkotja.

□ **Statisztikai ismértv**

- **Statisztikai ismértv, vagy ismértvváltozó:** a statisztikai sokaság elemeit jellemző tulajdonság (pl. testsúly, testmagasság);
- **Ismértvváltozat:** az ismértvek által felvehető lehetséges értékek, pl.
 - statisztikai egyed: hallgató,
 - statisztikai ismértv: nem; osztályzat; testsúly, testmagasság;
 - statisztikai ismértvváltozat: férfi, vagy nő; 1,2,3,4,5; 65 kg; 172 cm;
- **Alternatív ismértv:** ha az ismértv csak két változattal rendelkezik;
- **Közös ismértv:** amely a statisztikai sokaságot meghatározza, mely alapján a sokaság egységei egyformák (pl. évfolyam);
- **Megkülönböztető ismértv:** amely szerint a sokaság egyedei különböznek egymástól (pl. nem, kor, lakcím, tanulmányi átlag);

- **Az ismérvek csoportosítása információtartalmuk szerint:**
 - ***Területi ismérvek:*** az egységek térbeli elhelyezésére szolgáló rendező elvek. Ismérvváltozatai általában földrajzi egységek (pl. a hallgatók születési helye);
 - ***Időbeli ismérvek:*** az egységek időbeli elhelyezésére szolgáló rendező elvek. Ismérvváltozatai időpontok és időszakok (pl. a hallgatók születési ideje). Kifejezhet állapotot és tartamot is.
- **Tárgyi ismérvek:**
 - ***Mennyiségi ismérvek:*** az egyedek számszerűen mérhető tulajdonságai.
 - ❖ folytonos – adott intervallumon belül bármilyen értéket felvehet (pl. testsúly, testmagasság);
 - ❖ diszkrét – csak egymástól jól elkülönülő értéket vehet fel (pl. a Kar gimnáziumban, szakközépiskolában, stb. végzett hallgatóinak a száma);
 - ***Minőségi ismérvek:*** az egységeket valamilyen tulajdonság alapján különböztetjük meg (pl. nem, foglalkozás, hajszín);

❑ Statisztikai adat I.

- **alapadat:** mérés, vagy számlálás útján nyerjük;
- **származtatott adat:** műveletek eredménye;

❑ Statisztikai adat II.

Számszerű adatok:

- **abszolút adat:** adatgyűjtésből származnak, számítással kapjuk meg (összegezés, különbségképzés, stb.);
- **relatív adat:** két statisztikai adat hányadosaként számítható ki;
- **egyéb adat:** szöveges adat, ezeket kódolni kell, hogy elemzésre tudjuk használni;

❏ **Adatgyűjtés**

Az elemzéshez szükséges adatok beszerzése adatgyűjtéssel történhet, amely lehet:

- **teljes körű:** a sokaság valamennyi egyedére vonatkozik;
- **részleges:** a sokaság egy részére vonatkozik;

■ ***A statisztikai adatok hibája***

Az összegyűjtött adatok általában korlátozott pontosságúak, azaz hibásak.

Keletkezése szerint a hiba lehet:

- adatgyűjtési hiba
- véletlen hiba.

Jellege szerint a hiba lehet:

- **Abszolút hiba (a):** a valóságos adat (A) és a mért adat (A') különbsége.

$$a = A - A'$$

- **Relatív hiba (α):** az abszolút hiba és a valóságos adat hányadosa, amelyet százalékban szoktak megadni

$$\alpha = a/A$$

Az adatok megadásánál beszélünk **szignifikáns számjegyről**, ez azt jelenti, hogy **nagyságrendben csak azokat a számjegyeket írják le (csak abban a nagyságrendben írják ki a hibát), amelyeket még megbízhatónak tartanak.** Az adatok jelentős része kerekített szám.

- A **becsült abszolút hiba** kiszámítható az alábbi képlettel:

$$\hat{a} \leq \frac{10^k}{2}$$

ahol \hat{a} : a becsült abszolút hiba

k: az utolsó szignifikáns számjegy helyi értékének hatványszáma
(azaz **a becsült abszolút hiba nagyságrendje**)

- A **becsült relatív hiba**:

$$\hat{\alpha} = \frac{\hat{a}}{A'}$$

- **Feladat**

Magyarország lakossága 2010. január 1-én 10.013 ezer fő.

$$\hat{a} \leq \frac{10^3}{2} = 500 \quad \hat{\alpha} = \frac{500}{10013000} = 0,0000499 = 0,005\%$$

Mintavételezési eljárások

A populáció minden egyes elemének ugyanakkora esélyt kell biztosítani a mintába kerüléshez.

A minta elemszámának elég nagynek kell lennie ahhoz, hogy a következtetéseink átvihetők lehessenek a populációra is.

Rétegzett mintavételezés: A populációt adott szempontok szerint csoportokba osztjuk, és a csoportok arányait a mintában is megtartjuk

Véletlen mintavételezés: A mintába kerülő egyedeket sorsolással választjuk ki.

Cenzus: népszámlálás

- **Valószínűségi mintavétel:** a populáció valamennyi egyedének lehetősége van (nullánál nagyobb az esélye) bekerülni a mintába. A valószínűségi mintavételhez az alábbi mintavételi eljárások tartoznak:

- egyszerű véletlen mintavétel;
- szisztematikus mintavétel;
- rétegzett mintavétel;
- többlépcsős csoportos mintavétel;

A különböző mintavételi eljárások két dologban megegyeznek:

- minden elemnek van egy nullánál nagyobb valószínűsége a bekerülésre;
- bizonyos tekintetben magukba foglalják a random kiválasztást;

- **Nem valószínűségi mintavételi eljárás:** olyan módszer, melynek során a populáció bizonyos elemeinek nincs lehetősége bekerülni a mintába, vagy a kiválasztás valószínűségét nem lehet pontosan meghatározni. A nem valószínűségi mintavételhez az alábbi eljárások tartoznak:

- önkényes mintavétel;
- koncentrált mintavétel;
- kvótás mintavétel;
- hólabda módszerű mintavétel;

Statisztikai alapfogalmak (2)

Eset

A minta egy eleme, az adatmátrix egy sora.

Mintaelemszám

Az adott minta elemeinek száma. Egy adatmátrix sorainak száma.

Adatmátrix

n db eset és p db változó adatainak mátrixba rendezett alakzata

Változó

A populáció egy mérhető jellemzője. Az adatmátrix egy oszlopa.

Példák változókra

- Magyarország állampolgárai: **fizetés, kor, nem, párt, stb.**
- Egy egyetemi kar hallgatói: **tanulmányi átlag, etr-kód, nem, szak, teljesített kreditek száma, stb.**
- Az autók halmaza: **gyorsulás; fogyasztás; lóerő; típus;...**
- Egy adott termék vásárlóinak halmaza: **vélemény az árról, minőségről, ...**
- Egy TV csatorna nézőinek halmaza: **kor, nem, tetszési index, iskolázottság, stb.**

Statisztika

A minta realizáció adataiból adott képlettel számolt adat a statisztika számított értéke.

átlag, standard szórás, medián, kvartilis, ferdeség, lapultság, módusz, gyakoriság, próbastatisztikák, stb.

Statisztikai alapfogalmak (3)

- **Statisztikai sokaság:** a statisztikai megfigyelés tárgyát képező egyedek összessége, halmaza.
- A **sokaság egységei:** a sokaságot alkotó egyedek.
- **Ismérvek:** Azok a kritériumok, amelyek szerint a sokaság egységeit jellemezzük.

Példák

Adathalmazok

■ Minta

- A Mezőgazdasági Kar hallgatóinak egy csoportja
- 20 egyetemi hallgató testsúlya
- ...

■ Sokaság

- A Mezőgazdasági Kar hallgatói
- A Mezőgazdasági Kar hallgatóinak a testsúlya
- Általában az egyetemi hallgatók testsúlya
- ...

Változók és adatok

■ Változó:

az alapsokaság egyedei **ismérvének „értéke”** mintavétel, illetve **megfigyelés előtt**, jelölése a továbbiakban: X, Y, X_1, X_2, \dots

■ Adat:

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

- kis latin betűkkel jelöljük: x, y, x_1, x_2, \dots ,

➤ Változók és adatok, példa

- Valamely adott helyen a holnapi csapadékmennyiség
 - ma még változó: X
 - holnapután már adat, pl. $x = 8 \text{ mm}$

Ismérvek (változók) típusai, változatai

- *kvalitatív* (minőségi, megállapítható) ismérvek
(pl. „nem”, „szín”, „hivatali beosztás”);
- *kvantitatív* (mennyiségi, mérhető) ismérvek
 - diszkrét (pl. „az iskolák száma adott településen”);
 - folytonos (pl. „a hőmérséklet adott helyen és időben”);

Kvalitatív ismerv változatai:

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

Pl. típusváltozatok

nem	férfi, nő
szín	fehér, piros, stb.

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

Diszkrét kvantitatív változó

- **lehetséges értékei (változatai)**
- **véges;**
pl. „a fiúk száma egy 30 fős osztályban”
lehet 0, 1, 2,,30
- **megszámlálhatóan végtelen sok;**
(gyakorlatilag nincs felső határa)

Folytonos kvantitatív változó

- lehetséges értékei egy intervallum bármely értéke

pl: vércukorszint

Ph érték

életkor

testsúly

hőmérséklet

Adat-transzformációk, átskálázás

Gyakran a mért (megfigyelt, megállapított) adatok helyett célszerűbb ezek „transzformált”-jaival dolgozni.

- A kvalitatív adatokat olykor kvantifikáljuk (pl. bonitálás);
- A kvantitatív adatok leggyakoribb transzformációja:
 - log-transzformáció
 - négyzetgyök transzformáció
 - reciprok-képzés

3. Sokaságok

Megfigyelési egység - Beszámolási egység

A megfigyelési egységek összességét statisztikai sokaságnak nevezzük.

Típus	Diszkrét	Folytonos	Aggregát
Álló	A hallgatók létszáma	Egy sütöde lisztkészlete	Egy bolt árukészlete
Mozgó	Egy fodrász vendégeinek száma	Egy üzem vízfogyasztása	Egy bútorgyár termelési értéke

4. Ismérvek

A megfigyelési egységek tulajdonságait ismérveknek nevezzük.

Ismérvváltozatok: ismérvek lehetséges kimenetelei

Az ismérvek típusai:

- a) közös - megkülönböztető
- b) mennyiségi (pl. testsúly, testmagasság)
- c) nem mennyiségi
 - ❑ területi (a hallgatók születési helye)
 - ❑ időbeli (a hallgatók születési éve)
 - ❑ minőségi (hajszín)

Ismérvek mérési skálái

- Az **adattípusok** fajtáit a **rendezhetőség** és a köztük értelmezhető **távolságfüggvény** alapján megfelelő skálán **mérjük**.
- a) **Névleges (nominális) skála:** tipikus kvalitatív skála. Értékei nem sorrendezhetők, csak két egyed azonos kategóriába, vagy különböző kategóriába tartozása állapítható meg ($X=Y$) illetve (XY) (pl. útlevélszám).
- b) **Sorrendi (ordinális) skála:** olyan kvalitatív skála, melyen a kategóriák sorrendje is megállapítható ($X<Y$) (pl. iskolai végzettség).
- c) **Intervallum skála:** olyan kvalitatív skála, amelyen két egyed távolsága ($X-Y$) mérhető. A skálának nincs valóságos nullpontja, $X=0$ nem jelenti az ismerv hiányát (pl. hőmérséklet).
- d) **Arány – (hányados) skála** olyan kvantitatív skála, amelynek valódi nullpontja van. Ilyen skálán két érték aránya (Y/X) értelmes viszonszám (pl. tömeg).

Névleges (nominális) skála (1)

- Legegyszerűbb mérési forma, számok kötetlen hozzárendelése dolgokhoz. Minden megfigyelt egyedet olyan adattal írunk le, melyek egymással nagyság szerint nem összehasonlíthatók. **Az ismérvértékek csak azonosságuk, vagy különbözőségük alapján mérhetőek. Sorrend nem állapítható meg.**
- Az objektumokhoz rendelt szimbólumok, számok csak az objektumok, vagy azok osztályainak azonosítására szolgálnak (egyéb jelentésük nincs!)
- Csak a megkülönböztethetőséget követeljük meg, így csak az egyenlőségi reláció értelmezhető.
- Pl. útlevélszám, repülőjáratok számozása, megszámok, a dolgozó neve, születési helye, neme...stb. akkor is nominális, ha számban kódolt: pl. a dolgozó törzsszáma. További példák: foglalkozás, családi állapot;
- **Számítható statisztikai mutató: osztályok azonosítása esetén a gyakoriság, módusz.**

Sorrendi (ordinális) skála (2)

- Az ismértértékeket nemcsak az azonos, vagy különböző kritériumok szerint rendezzük, hanem egy természetes sorrend szerint is.
- Az egyenlőségi reláció mellett a sorrendiségre vonatkozó reláció is érvényes.
- A sorrendi skálán mért egységek nincsenek egymástól egyenlő távolságra!
- Számtani átlag és szórás nem számítható!!!! Számítható a kvantilis, medián, rangkorrelációs együttható.
- Minden olyan transzformáció végezhető, amely a skála eredeti sorrendjét változatlanul hagyja.
- Pl. termékek minőségi osztályozása, kérdőíves felméréseknél 3, 5, 7 fokozatú skála, tűzveszélyességi osztály, iskolai osztályzatok és értékelési kategóriák (kiváló, jó, stb.), helyezési számok egy versenyen, stb.

Sorrendi (ordinális) skála (2)

- **Mérést jelent, ugyanis a skálaértékek különbségei is valós információt adnak a sokaság egységeiről. Bármely két adat összehasonlítható.**
- Példa: dolgozók iskolai végzettsége.
- Jellemző:
 - Nincs távolság értelmezve az adatok között. *(Pl. Nem lehet megmondani, hogy mennyivel értékesebb az érettségi a 8 általánosnál.)*
 - **Egyetlen művelet: adatok rendezése** – olyan rangstatisztika alkalmazható, amelyet csak **az adatok egymáshoz képesti rendezettségére** használnak. *(Pl. átlagnak nincs értelme, de mediánnak és módusznak igen – ezekről a későbbiekben lesz szó).*

Intervallum skála (1)

- Rendelkezik a **sorrendi skála** tulajdonságaival + a skála bármelyik két pontja közötti különbség, azaz a távolság is értelmezhető.
- Nincs rögzített nullpont, a skála nullpontját és mértékegységét szabadon választhatjuk meg.
- A közös és állandó mértékegység jellemzi és a számokat ennek alapján rendeljük a sorba rendezett dolgokhoz.
- A skála bármilyen lineáris transzformációja megengedett.
- A **mértani átlag és a relatív szórás kivételével valamennyi statisztikai jellemző és mutató számítható**.
- **Pl. hőmérséklet, naptári idő, tengerszint feletti magasság**

Intervallum skála (2)

Pl. a rangsorba rendezett tulajdonságokat **egyenlő** közök választják el. (Pl. IQ tesztek: 100-110, 110-120: egyenlő távolságok. **De: nem mondhatjuk, hogy a 150 IQ-jú személy 50%-kal intelligensebb, mint a 100-as IQ-jú személy.)**

- *Értelmetlen az egymáshoz viszonyított arányról beszélni: pl. hogy a 20°C kétszer olyan meleg, mint a 10°C .
(A hőmérséklet a Kelvin skálán nem intervallum típusú!)*

- A skálaértékek különbségei valós információt nyújtanak a sokaság egységeiről.

(Pl. két ember összehasonlításánál mondhatjuk, hogy **különböznek** egymástól (nominális), egyik **magasabb** a másiknál (ordinális) és meg tudjuk mondani, hogy **ennyivel** (intervallum).

Arányskála (1)

- A legmagasabb rendű, a legerősebb mérési formát jelenti.
- Rendelkezik a korábbi skálák tulajdonságaival és teljesülnek az additivitási követelmények is:
- A skálának valódi nullpontja van, és bármelyik két pontjának aránya független a mértékegységtől.
- Pl. termelés, forgalom, jövedelem, kereset, stb. mérése

Arányskála (2)

- Valós számokkal jellemezhető adatok.
- Minden olyan műveletet végrehajthatunk ezekkel, amelyeket a valós számokkal.

A kezdőpont egyértelműen adott és rögzített, s így a skálaértékek egymáshoz való **aránya** is meghatározható (pl. életkor, jövedelem, stb.)

Mérési skálák	Tulajdonság	Értelmezhető relációk	Sajátosságok	Jellemző példák
Nominális	Megkülönböztetés	$X_a = X_b$ vagy $X_a \neq X_b$ <u>Számítható:</u> Gyakoriság, módusz,	Nem számszerű	Név, születési hely, nem, útlevekszám, repülőjáratok számozása, mezszámok, nemzetiségi/vallási hovatartozás
Ordinális	Megkülönböztetés, sorrend	$X_a = X_b$ vagy $X_a \neq X_b$ és $X_a \geq X_b$ vagy $X_a < X_b$ <u>Számítható:</u> Kvantilis, medián, rangkorrelációs együttható	Nehezen mérhető, csak sorrendbe állítható	Sorrendek, (katonai) rangok, iskolai végzettség, termékek minőségi osztályozása, kérdőíves felméréseknél 3, 5, 7 fokozatú skála, tűzveszélyességi osztály
Intervallum	Megkülönböztetés, sorrend, különbség	$X_a = X_b$ vagy $X_a \neq X_b$ és $X_a \geq X_b$ vagy $X_a < X_b$ <u>Értelmezhető:</u> $X_a - X_b$ <u>Számítható:</u> Valamennyi stat. jellemző és mutató, kivéve: mértani átlag és relatív szórás	Pozitív és negatív értékek	Hőmérsékleti skála (kivéve Kelvin-skála), naptári idő, tengerszint feletti magasság, IQ-teszt egyenkénti kategóriái
Arány	Megkülönböztetés, sorrend, különbség, arány	$X_a = X_b$ vagy $X_a \neq X_b$ és $X_a \geq X_b$ vagy $X_a < X_b$ <u>Értelmezhető:</u> $X_a - X_b$ valamint X_a / X_b <u>Számítható:</u> Minden olyan művelet amelyek valós számokkal végrehajtható.	Van elméleti minimum, azonos előjelű	Darabszámmal, vagy intenzitással rendelkező mennyiségek, termelés, forgalom, népességszám, életkor, jövedelem, kereset

A különböző mérési skálájú adatok tulajdonságai

Példa nominális mérési szintű változóra:



Nő



Férfi

Példa ordinális mérési szintű változóra: vallásosság – „Mennyire fontos a vallás az Ön számára?”

Nem igazán fontos

Elég fontos

Nagyon fontos

A legfontosabb



Alacsony

Magas

Példa intervallum mérési szintű változóra: IQ



95



100



105



110



115

Példa arányskálára: jövedelem



0



10 000 \$



20 000 \$



30 000 \$



40 000 \$



50 000 \$

*Példa különböző ismérvekre, s
a hozzájuk rendelhető mérési
skálákra*

Nominális

Név

Nominális

Nem

Ordinális

Legmagasabb iskolai végzettség

Arány

Reggel és este mért vérnyomás

Intervallum

Kezdő és jelenlegi fizetés

Nominális

Tájegység

Földkategória

Ordinális

A
r
á
n
y

Saját terület (ha)

Arány

Erőgépek száma

Ismérvek és mérési skálák

Ismérv

Mérési skála

Területi

Nominális skála

Minőségi

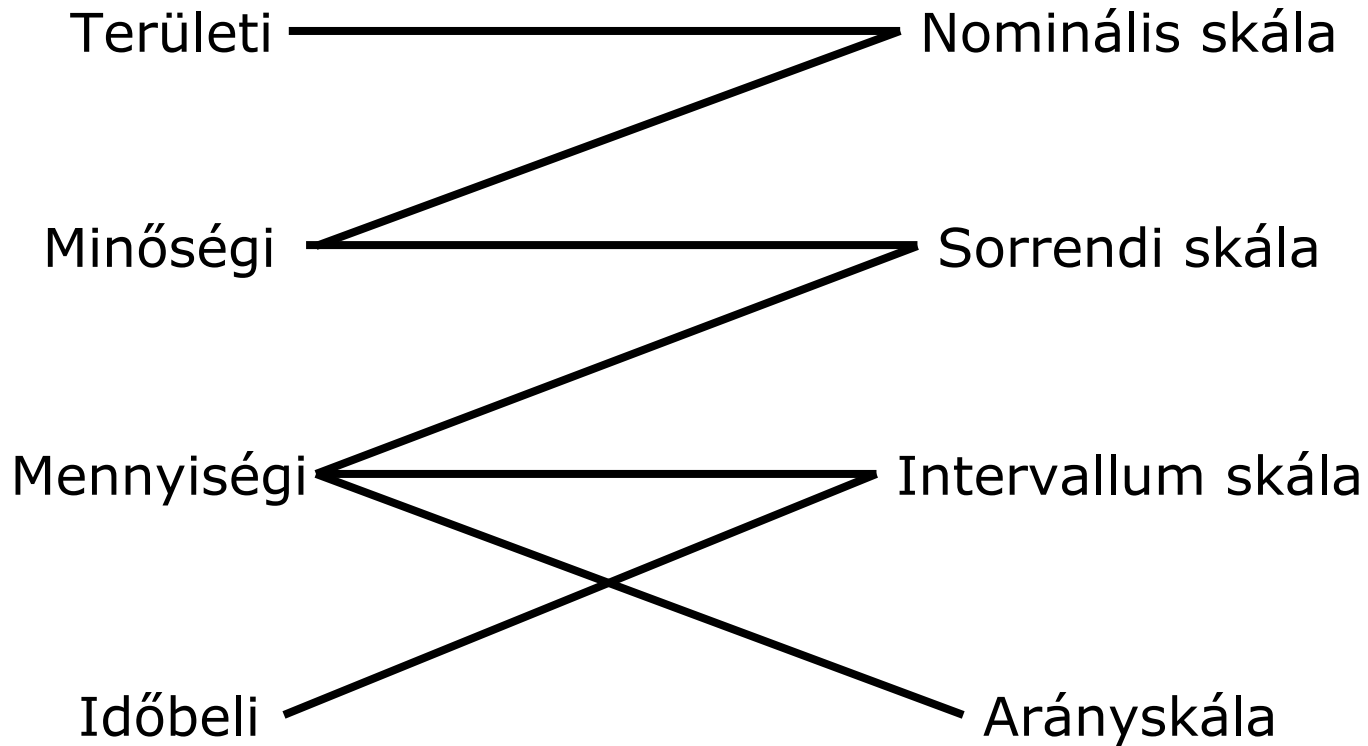
Sorrendi skála

Mennyiségi

Intervallum skála

Időbeli

Arányskála



Feladat/1.

Sokaság	Egy konkrét egység	Ismérv	Ismérv-változat	Ismérvfajta/ Mérési skála
A magyar népesség 2007. január elsején	Kiss Réka	Születési idő	1976	Időbeli/ intervallum
		Lakóhely	Budapest	Területi/ nominális
		Nem	Nő	Minőségi/ nominális
		Életkor	29	Mennyiségi/ arány

Feladat/2.

Adottak az alábbi sokaságok:

- Magyarország népessége 2006. jan.1-jén 10 076 581 fő.
- A budapesti férfiak sörfogyasztása a 2006-os VB idején.
- BCE oktatói 2006. szept. 4-én.
- Jótékonysági koncertek 2006-ban a Zeneakadémián.

Feladat:

- Állapítsa meg a sokaságok típusát és egységeit!

Feladat/3.

Döntse el az alábbi ismérvekről, hogy mennyiségi vagy minőségi ismérvek-e!

- ☐ Nem (férfi, nő)
- ☐ Életkor
- ☐ Magasság
- ☐ Testsúly
- ☐ Családi állapot
- ☐ Iskolai végzettség
- ☐ Foglalkozás
- ☐ Bruttó havi fizetés

Statisztikai adat és mutatószám

■ Statisztikai adat:

Az egyedeokról
szerezhető információ.
(szám, vagy
számszerű jellemző)
fogalmi jegy
időbeli azonosító
térbeli azonosító
számérték
mértékegység
(mérés vagy számlálás)

Statisztikai mutatószám:

Valamilyen statisztikai
módszerrel a
rendelkezésre álló
adatokból számított
származtatott statisztikai
mérőszám.

Például:

(Havi) Átlagbér Magyarországon 2008-ban bruttó **194.000** Ft/fő/hó

Statisztikai sorok

A sokaság egy ismerv szerinti tömör jellemzése.

Sorkészítés célja szerint:

- Csoportosító sor
- Összehasonlító sor
- Leíró sor

} Valódi statisztikai sorok
(azonos fajtájú adatokból)

→ Nem valódi statisztikai sor
(különböző fajtájú adatokból)

Ismervfajtáknak megfelelően:

Időbeli (tartam-állapot), területi, minőségi, mennyiségi
+ leíró sorok

Sorok készítése: ismervváltozatok | számszerű értékek

■ **Statisztikai sorok**

A statisztikai sor **a statisztikai adatok** meghatározott összefüggésben történő felsorolása, **egyetlen ismerv szerinti csoportosítása** eredményeként jön létre. A csoportosítás lehetőséget nyújt a sokaság szerkezetének, struktúrájának tanulmányozására.

A statisztikai sorok felosztása:

■ **Az adatok fajtája szerint:**

- **valódi sorok:** egy sokaságra vonatkoznak, azonos fajtájú adatokból keletkeznek
- **nem valódi sorok:** valamilyen jelenség leírását adják többféle szempontból, több sokaságra vonatkoznak:
 - ❑ **leíró sor:** A megfigyelés tárgyának különböző tulajdonságait írja le, az adatok között logikai kapcsolat van. Az intenzitási viszonzyszámok forrásai.

- ***A keletkezés módja szerint a valódi sorokat tovább bonthatjuk:***
 - **Összehasonlító sorokra:** az adatokat felsorolásszerűen tartalmazza:
 - ❖ állapot idősor
 - ❖ leíró sor
 - **Csoportosító sorokra:** az ismerv változat szerinti részsokaságokból tevődik össze:
 - ❖ minőségi sor,
 - ❖ mennyiségi sor
 - ❖ területi sor
 - ❖ tartam idősor
- ***Az ismerv fajtája szerint a valódi sorok az alábbiak lehetnek:***
 - ❖ idősor,
 - ❖ területi sor,
 - ❖ minőségi sor,
 - ❖ mennyiségi sor.

■ **Idősorok**

Bizonyos időpontban meglévő, vagy bizonyos időtartam alatt bekövetkező jelenségek adatait sorolják fel.

- ❖ **Állapot idősor:** az álló sokaságok időbeli változását mutatja be. Az adatok összegzésének nincs tárgyi értelme. **Pl. a személygépkocsi állomány éves bontásában.**
- ❖ **Tartam idősor:** egy időintervallum alatti mozgó sokaság időbeli alakulását mutatja. Dinamikus helyzetet tükröz. Az adatok összegzésének van tárgyi értelme, az összegzéssel egy hosszabb időtartamra jellemző adatot képezhetünk. **Pl. turisták száma évente.**

■ **Területi sorok**

- ❖ Ha a csoportosítás a rendező elv, akkor egy nagyobb területi egység adatait az összetartozó kisebb területi egységeknek megfelelően osztályozzuk. Azonos időpontban rögzített statisztikai sokaság térbelileg bontott részsokaságait tartalmazza. Az ismérvváltozatok sorrendje kötetlen. A gyakoriságok összeadásának van tárgyi értelme. **Pl. a népesség megyék szerinti megoszlása.**
- ❖ Ha azonban az összehasonlítás a rendező elv, akkor az adatok összegzésének nincs tárgyi értelme.

■ ***Minőségi sorok***

A sokaságelemek minőségi tulajdonságainak formailag előre rögzített csoportosítása révén jön létre. Betekintést nyújt a sokaság összetételébe, szerkezetébe. Az ismérvváltozatok sorrendje kötetlen, a gyakoriságok (az előfordulások száma) összeadásának van tárgyi értelme. **Pl. A külföldiek megoszlása az utazás jellege szerint 2015-ben: csoportos turista, egyéni turista, átutazó.**

■ ***Mennyiségi sorok***

Mennyiségi ismérvek szerinti csoportosítással jön létre. Ha az ismérvértékek konkrét pontossággal megadott számok, akkor *diszkrét mennyiségi sorról* beszélünk. Általában számlálással kapjuk az ilyen sorokat. Ha az ismérvértékek adott intervallumban bármilyen értéket felvehetnek, akkor *folytonos mennyiségi sorról* beszélünk.

■ A mennyiségi sorok fajtái:

Gyakorisági sor: az ismerv előfordulásának gyakoriságát tüntetjük fel. A gyakoriság (f_i) megmutatja, hogy az egyes ismervváltozatok hányszor fordulnak elő a megfigyelt sokaságban. Ha az egyes gyakoriságokat azok összegéhez viszonyítjuk, akkor az adott ismervérték relatív gyakoriságát (g_i) kapjuk meg:

$$g_i = \frac{f_i}{\Sigma f_i}$$

ahol: g_i : az i -edik ismervérték relatív gyakorisága

f_i : az i -edik ismervérték gyakorisága

$\Sigma f_i = n$: a sokaság elemeinek száma

- Ha az ismérvváltozatok száma nagy, akkor az adatokat rangsoroljuk, és ez megkönnyíti a változó osztályozását. Az osztályozás sűríti az információt. A legnagyobb és legkisebb ismérvek által adott intervallumot úgy osztjuk osztályokba, hogy az egyes osztályközökön belül a gyakoriságok közel egyenlő eloszlásúak legyenek, így az osztályközép alkalmas lesz az osztály jellemzésére.

- Az osztályok olyan adatcsoportok, ahol az egyes osztályok közötti mennyiségi változás minőségi változást takar.
- Az osztályközök száma az a legkisebb k , amelyre: $2^k > N$
- Az osztályköz hossza:

$$h = \frac{x_{\max} - x_{\min}}{k}$$

- Ha a csoportosító ismerv folytonos jellegű, a mennyiségi sort osztályközös gyakorisági sornak nevezzük.

■ Példa

A sárgarépa nitrogén-tartalmának meghatározására 24 elemű mintát vettek és megmérték a nitrogéntartalmat. A mérések eredménye (mg/100g):

104,2	100,2	93,4	98,6	117,4	100,8	88,2	100,1
101,9	116,8	112,9	97,8	99,3	111,5	109,3	100,7
118,0	96,4	89,4	98,9	105,3	112,4	99,5	102,7

Feladat:

Készítsen az adatokból rangsort!

Mennyi az osztályközök optimális száma?

Megoldás:

88,2	89,4	93,4	96,4	97,8	98,6	98,9	99,3
99,5	100,1	100,2	100,7	100,8	101,9	102,7	104,2
105,3	109,3	111,5	112,4	112,9	116,8	117,4	118,0

$n=24$ $2^5=32 > 24$, ezért $k=5 \rightarrow h=(118,0-88,2)/5=5,96$.

\Rightarrow Az osztályköz hossza kerekítve 6 lesz

<i>Nitrogéntartalom (mg/100g)</i>	<i>Sárgarépa száma (db)</i>	<i>Osztályközép</i>	<i>Kumulált gyakoriság</i>
88,1-94,0	3	91	3
94,1-100,0	6	97	9
100,1-106,0	8	103	17
106,1-112,0	2	109	19
112,1-118,0	5	115	24
Összesen:	24	-	-

Az Excelben a következő módon számolhatunk: Írjuk be az osztályközök felső határát a B2-B6 cellákba, majd jelöljük ki a C2-C5 cellákat. A '*Beszúr*' menü '*Függvény*' almenüjéből válasszuk ki a Statisztikai függvények közül a '*Gyakoriság*' függvényt. Az 'adattömbnek' adjuk meg az A2-A21 tömböt az egér kijelölésével. A 'Csoport_tömb' a B2-B6 lesz. Ezután kattintsunk a kész gombra. A szerkesztőlécben megjelenik a: =GYAKORISÁG(A2:A21;B2:B6). Az egérkurzorra a szerkesztőlécre állva a SHIFT, a CTRL és az ENTER billentyűk együttes lenyomása után a C2-C6 cellákban megjelenik a gyakoriság.

- **Értékösszeg-sor:** Ha a gyakorisági soroknál az ismérvek változatainak a szorzatát tüntetjük fel, akkor értékösszeg-sort kapunk.

Az értékösszeg jele: s_i .

$$s_i = f_i \cdot x_i$$

Példa

1. táblázat

Egy településen a családok megoszlása a családban lévő gyermekek száma szerint

Gyermekek száma	Családok száma	Összes gyermekek száma (s_i)
0	992	0
1	954	954
2	761	1522
3	148	444
4	40	160
5	15	75
6	5	30
Összesen	2915	3185

- **Kumulált gyakorisági sor:** A gyakorisági soroknál sajátos információkat nyerhetünk a gyakoriságuk kumulált képzésével. A kumulálás halmozott összeadást jelent, s arról kapunk képet, hogy egy adott értékhatárnál kisebb (alulról kumulált), vagy nagyobb (felülről kumulált) érték összesen hányszor fordul elő. A kumulált gyakoriság jele: f_i' .

Példa

Az alábbi táblázat a gyermekvédelmi gondozásban részesülők számát mutatja be. Feladat a kumulált gyakoriság és a relatív gyakoriság kiszámítása.

2. táblázat

A gyermekvédelmi gondozásban részesülők száma

Korcsoport (év)	gyermekek száma (fő)	f_i'	g_i
0-3	1848	1848	0,1078
4-5	1221	3069	0,0712
6-9	3075	6144	0,1794
10-11	2164	8308	0,1262
12-14	3827	12135	0,2232
15-17	5010	17145	0,2922
Összesen	17145	-	1,0000

Forrás: Magyar statisztikai zsebkönyv, 2006.

Pl.: $f_3' = 3069 + 3075 = 6144$; $g_3 = 3075 / 17145 = 0,1794$

A kumulált gyakoriságot és a relatív gyakoriságot az Excel segítségével könnyen ki lehet számolni. A kumulált gyakoriság esetében az első helyre beírjuk, hogy egyenlő önmagával ($=B2$). A következő cellába (C3) pedig az összeadandókat. Ezután már csak az egérrel kell a cella jobb alsó sarkát lehúzni a többi cellára, és a megkapjuk a kumulált gyakoriságokat. A relatív gyakoriságot úgy tudjuk kiszámolni, hogy a megfelelő cellába az egyenlőségjel után beírjuk a művelet képletét. A „B2” jelölés a relatív hivatkozás, a „B\$8” pedig a sorra nézve az abszolút hivatkozás Így már a többi cellába nem kell képletet gépelni, hanem a cella jobb alsó sarkát az egérrel lefelé húzni a többi cellára és megkapjuk a helyes értékeket (1. ábra).

	A	B	C	D
1	Korcsoport (év)	gyermek száma (fő)	f_i'	g_i
2	0-3	1848	=B2	=B2/B\$8
3	4-5	1221	=C2+B3	
4	6-9	3075		
5	10-11	2164		
6	12-14	3827		
7	15-17	5010		

Az Excel munkalapjának részlete

■ **Statisztikai tábla**

A statisztikai sorok összefüggő rendszere. Több tulajdonság szerinti rendezéskor kapjuk.

❖ **Rendeltetés szerint lehet:**

alap-, vagy gyűjtő tábla,
feldolgozási-, vagy munkatábla,
közlési-, vagy eredménytábla.

❖ **Az összefoglalt sorok típusa szerinti felosztás alapján lehet:**

✓ **egyszerű tábla:** csoportosítást nem tartalmazó adatsorok összefüggő rendszere. Általában leíró és összehasonlító sorokat tartalmaz.

Megnevezés	2007	2008	2009
Vállalkozások száma			
Saját tőke			
Idegen tőke			

- ✓ **Csoportosító tábla:** egy ismerv szerinti csoportosítást tartalmazó statisztikai sorok összefüggő rendszere.

Életkor	2007	2008	2009
0-6			
7-14			
15-22			
.			
.			
Összesen			

- ✓ **Kombinációs tábla:** A sokaság több ismerv szerinti kombinatív osztályozásának eredményeként kapott adatokat tartalmazza.

Megnevezés	Budapest	Város	Község	Összesen
Komfortos				
Félkomfortos				
Komfort nélküli				
Összesen				

■ ***Statisztikai táblák készítése***

Formai és tartalmi követelményeknek kell érvényesülni, hogy megfeleljen a vizsgálat céljára.

• ***Formai követelmény:***

- táblázat címe,
- megnevezések a fej- és oldalrovatban,
- mértékegység feltüntetése,
- forrás megjelölése.

Példa

3. táblázat

A gyermekvédelmi gondozásban részesülők száma 2005-ben

Korcsoport (év)	Gyermekek száma (fő)
0-3	1848
4-5	1221
6-9	3075
10-11	2164
12-14	3827
15-17	5010
Összesen	17145

Forrás: Magyar statisztikai zsebkönyv, 2006.

Tartalmi követelmény.

A tábla minden celláját információközlésre kell felhasználni, ezért a következő szokásos jelöléseket alkalmazzák:

- konkrét adat: számszerű információ,
- kihúzott rovat (-): nincs információ
- ...: létezik adat, de nem áll rendelkezésre
- +: az adat előzetes becslés eredménye
- *: a táblázat alján lévő megjegyzésre utal
- 0,0: van adat, de a többi adathoz képest túl kicsi.

Statisztikai sorok

■ Csoportosító statisztikai sor:

A sokaság belső összefüggéseit fejezi ki, csoportosítás céljából készül, adatai összegeezhetők.

(időbeli, területi, minőségi, mennyiségi)

Ismérv-változatok	Egységek száma
C1	f1
C2	f2
.	.
.	.
Ci	fi
.	.
Ck	fk
Összesen:	N

Statisztikai sorok

■ Például:

A teremben ülő hallgatók
hajszín szerint.

minőségi csoportosító
statisztikai sor

Hajszín	Hallgatók száma/fő
barna	23
szőke	12
fekete	4
vörös	2
ősz	2
egyéb	1
Összesen:	44

Statisztikai sorok

- Összehasonlító statisztikai sor:
Összehasonlító adatok statisztikai sorba rendezve, összehasonlítási céllal, adataik nem összegeezhetők.
(idősor, területi)

Ismérv-változat	Számérték/ mértékegység
C1	adat
C2	adat
.	.
.	.
Ci	adat
.	.
Ck	adat

Statisztikai sorok

- Például: egy felsőoktatási intézmény nappali tagozatos hallgatóinak átlagos havi ösztöndíja 2004 és 2010 között

Összehasonlító időbeli sor

Év	Havi átlagos ösztöndíj (Ft/hallgató)
2004	12.600 Ft
2005	13.200 Ft
2006	13.800 Ft
2007	14.100 Ft
2008	14.000 Ft
2009	14.200 Ft
2010	15.000 Ft

Statisztikai sorok

Statisztika sorok kellékei:

- Cím (sokaság pontos megnevezése, a közös ismérvek felsorolása)
- Tulajdonságok, ismérvváltozatok felsorolása
- Ismérvváltozatoknak megfelelő gyakoriságok felsorolása
- Összesen rovat (csak a csoportosító sor estében)
- A forrás megnevezése

Statisztikai táblák

Statisztikai sorok összefüggő rendszere.

- Egyszerű tábla (összehasonlító és/vagy leíró sorok)
Nincs csoportosító sora, egy adata, egy statisztikai sor tagja.
- Csoportosító tábla (csoportosító és/vagy összehasonlító vagy leíró sorok)
Egyirányú csoportosítást tartalmaz, egy adata egy statisztikai sor tagja.
- Kombinációs tábla (csoportosító sorok)
Csak csoportosító sorokat tartalmaz, egy adata egyidejűleg több statisztikai sor tagja.

Statisztikai táblák

■ Egyszerű statisztikai tábla

Egy városban az orvosellátottság alakulása:

Év	Orvosok száma (fő)	Lakosok száma (fő)	Egy orvosra jutó lakosok száma
1990	240	80 000	333,3
1999	360	100 000	277,8

Statisztikai táblák

■ Csoportosító statisztikai tábla

Búzatermelés adatai 1991-ben:

Körzet	Termés (ezer tonna)	Termésátlag (t/ha)
Dunántúl	2000	5,2
Alföld	3000	5,31
Észak	705	4,71
Összesen	5705	...

Statisztikai táblák

■ Kombinációs statisztikai tábla

Egy felsőfokú intézmény nappali tagozatos hallgatóinak jegyei statisztikából
1991/1992 II. félév:

Osztályzat	A	B	C	Összesen
	kar hallgatóinak megoszlása			
5	19	23	19	61
4	32	49	40	121
3	24	36	56	116
2	20	36	82	138
1	1	2	18	21
Összesen	96	146	215	457

Statisztikai táblák

A statisztikai táblák részei:

- Oszlop (a táblázat egy oszlopa)
- Sor (a táblázat egy sora)
- Rovat (sor és oszlop találkozása)
- Fejrovat (a táblázat első sora, mely szövegesen tartalmazza az egyik ismértv változatait)
- Oszloprovat (a táblázat első oszlopa, mely szövegesen tartalmazza a másik szerinti ismértv változatokat)
- Összegrovat (a sorok és oszlopok összességét tartalmazza)

Statisztikai táblák

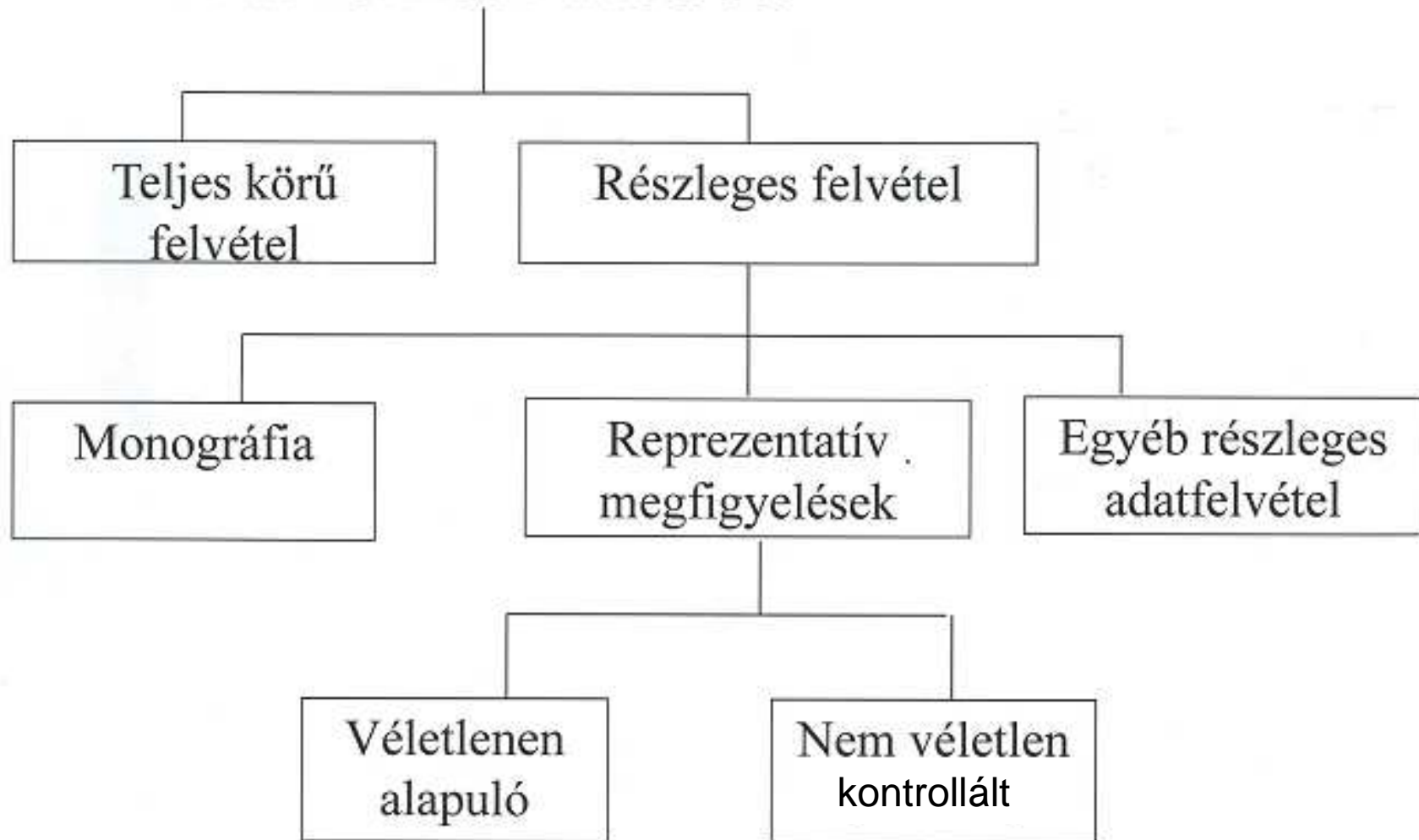
Dimenziószám:

Azt mutatja, hogy a tábla egy statisztikai adata egyidejűleg hány statisztikai sor tagja.

Táblakészítés szabályai:

- Cím (azonosítókkal!, idő, hely, stb.)
- Oldalrovatok (fejrovat és oszloprovat) megnevezése
- Egy rovat sem üres (--, ●(●); ... , 0,0)
- Megjegyzés (ha valamely rovatában lévő adat eltérő mértékegységű)
- Forrásmegjelölés (!)

Adatszerzési módok



Kérdőívszerkesztés

- Alapos szakmai hozzáértés
- Tömör, egyértelmű, könnyen megválaszolható kérdések
- Főleg feleletválasztós (karikázós, x-elős és kevés kifejtendő választ igénylő)
- Ne legyen túl hosszú
- Ajánlott az anonim adatfelvétel
- Kompromisszum: csak a legfontosabb dolgokat kérdezzük
- Véglegesítés előtt: próbalekérdezés
- Ha nyereményhez kötjük, növelhető a válaszadási arány

Kérdőív

Nem: férfi 1 nő 2

Születési dátum négyjegyű év-hó-nap:
Születési hely (irányítószámmal):
Jelenlegi/utolsó foglalkozás:
Korábbi foglalkozás:

Iskola (aláhúzással jelölve):
Kevesebb, mint 8 általános
8 általános
Középiskola (szakiskola vagy egy része)
Főiskola, vagy egy része
Egyetem, vagy egy része

Apa betegségei (max. 5 db betegség):

Anya betegségei (max. 5 db betegség):

Testvérek betegségei (minden egyes testvér esetében külön-külön, max. 5 db betegség):

Ha gyermekei vannak, a gyermek(ek) betegségei (minden egyes gyermek esetében külön-külön, max. 5 db betegség):

Saját betegségei (max. 5 db betegség):

Anyatejjel táplálták: igen – nem (aláhúzással jelölve):
Hány hónapig?:

Nem allergiás betegségei (aláhúzással jelölve):
Magas vérnyomás: igen – nem
Érbetegség: igen – nem
Szívbetegség: igen – nem
Tüdőbetegség: igen – nem
Cukorbetegség: igen – nem
Elhízás: igen – nem
Daganatos betegség: igen – nem

Allergiás megbetegedései (aláhúzással jelölve) saját megítélése szerint. (Ha az alábbiaktól eltérő(ek), akkor írja ide szövegesen a betegség(ek)et!)

Porallergia
Virágpor allergia
Ételallergia
Bőrállergia
Háziállat allergia
Gyógyszerallergia
Rovarcsípés allergia

Dohányzás (aláhúzással jelölve):
Nem dohányzik
Leszokott
Dohányzik, éve,db cigaretta/nap

Alkoholfogyasztás (aláhúzással jelölve):
Alkoholt nem fogyaszt
Hetente fogyasztmiből, mennyit (szöveges)
Naponta fogyasztmiből, mennyit (szöveges)

Lakókörülmények (aláhúzással jelölve):
Városi lakásban lakik: igen – nem
Ezen belül lakótelepi lakásban lakik: igen – nem
Családi/falusi házban lakik: igen – nem

Lakás paraméterei (aláhúzással jelölve):
Ház anyaga: Beton – Téglá – Vályog
Falak: szárazak – nyirkosak
Padló: parketta (laminált padló) – szőnyegpadló – kő – egyéb
Ágynemű: toll – nem toll
Háziállatok: kutyá – macska – baromfi – sertés – tehén

Vannak-e olyan tünetei, amelyek allergiára utalnak? Igen – nem (aláhúzás)

Van-e orvos által diagnosztizált allergiája? Igen – nem (aláhúzás)

Milyen gyógyszereket szed rendszeresen? (felsorolás)



Nézzük mindig a dolgok napos
oldalát!

Mára befejeztük, viszontlátásra!