

7. A mintavétel alapvető kérdései

Bevezetés a kutatás módszereibe

Dr. Keszi Roland

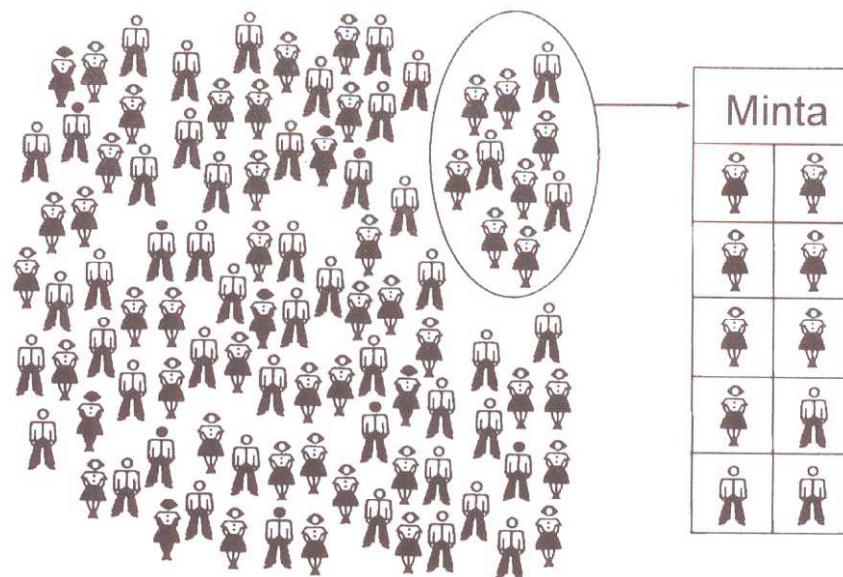
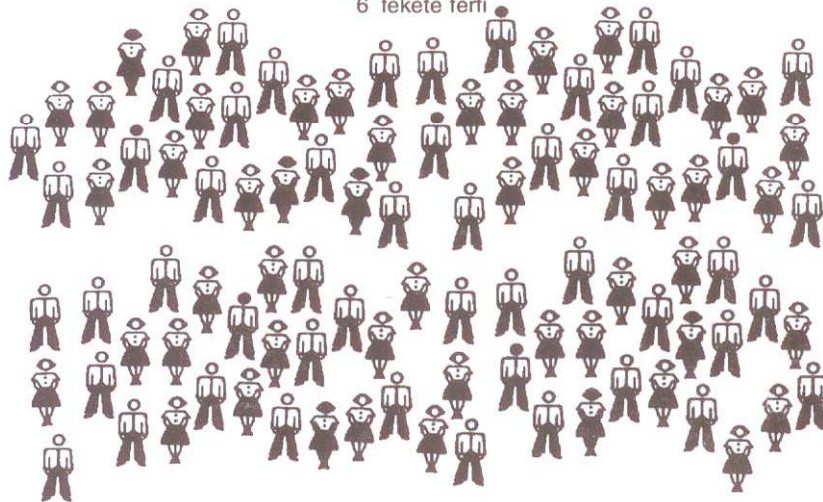
ELTE BGGyK

A mintavétel célja és alaptípusai

- Cél: a társadalomkutatás legtöbb típusában NINCS MÓDUNK megfigyelni az összes megfigyelési egységet => ki kell választanunk egy kezelhetőbb elemszámú halmazt
- Történeti gyökerei: USA
1920-as évek: Literaty Digest (önkitöltés, postai levelezőlapok)
1936-48: George Gallup (kvótás mintavétel=> 10 év siker, majd 48-ban bukás...)
- Alaptípusai:
 - A) Valószínűségi mintavétel
 - B) Nem valószínűségi mintavétel

A probléma: Miért és hogyan kell mintát vennünk az alapsokaságból?

44 fehér nő
44 fehér férfi
6 fekete nő
6 fekete férfi



A probléma oka

- Ha egy populáció minden eleme egyforma lenne, elég lenne egy elemet vizsgálni ahhoz, hogy levonjunk akármilyen következtetést...
- A társadalomtudományi kutatásban ez soha nincs így, **bármely populáció HETEROGÉN**, megfelelő vizsgálatukhoz, olyan mintát kell kiválasztanunk belőle, ami ugyanolyan (**nagyban HASONLÓ**), mint a vizsgált alapsokaság

Torzítás

- Vö. korábban a „mérleg” példa és az megbízhatóság-érvényesség esete a korábbi előadásban (*...de itt másról van szó!...*)
- **Általános definíció:** Az adott minta elemei nem jellemzőek arra az alapsokaságra, amelyből ki lettek válogatva. Lehet:

Tudatos torzítás: Csak bizonyos (*pl. mozgáskorlátozott, szőke, kövér, stb.*) embereket kérdezek meg arról, hogy X (*mert én is mozgáskorlátozott, szőke, kövér, stb... vagyok*).

Akaratlan (nem tudatos) torzítás: Bármilyen rosszul megválasztott mérési módszer okozhatja
Pl. Egy mátészalkai szórakozóhely népszerűségi mutatóinak feltérképezése az Ecseri úti aluljáróban vett minta alapján

Reprezentativitás

Önmagában a véve a szónak NINCS ÉRTELME ! Kizárólag a köznyelvben használják, legtöbbször tévesen.

Általános munkadefiníció: egy minta akkor reprezentatív, valamely jellemző(k)re /változó(k)ra/, ha jellemzői jól közelítik a populáció /alapsokaság/ ugyanezen összesített jellemzőit.

*Általános reprezentativitás-képző ismérvek lehetnek:
kor, nem, iskolai végzettség, településtípus*

A kutatási **cél** és a rendelkezésre álló **adatok** dönti el, hogy egy minta mely jellemzők (**VÁLTOZÓK**) szerint legyen reprezentatív.

Minden mintavétel célja olyan mintát kiválasztani, mely jól jellemzi az alapsokaságot (reprezentatív). A valószínűségi mintavétel alapja a véletlen kiválasztás, mely során minden elemnek egyenlő esélye van, hogy a populációból a mintába kerüljön.

Equal Probability Selection Method;

(**EPSEM** – **EKV** Egyenlő Kiválasztási Valószínűségű módszerek) Ez csak akkor lehetséges, ha nincs tudatos vagy nem tudatos torzítás, részrehajlás. Éppen ezért számítógépekkel végzik általában a mintába való kiválasztást, de ilyen a kockával való dobálás is.

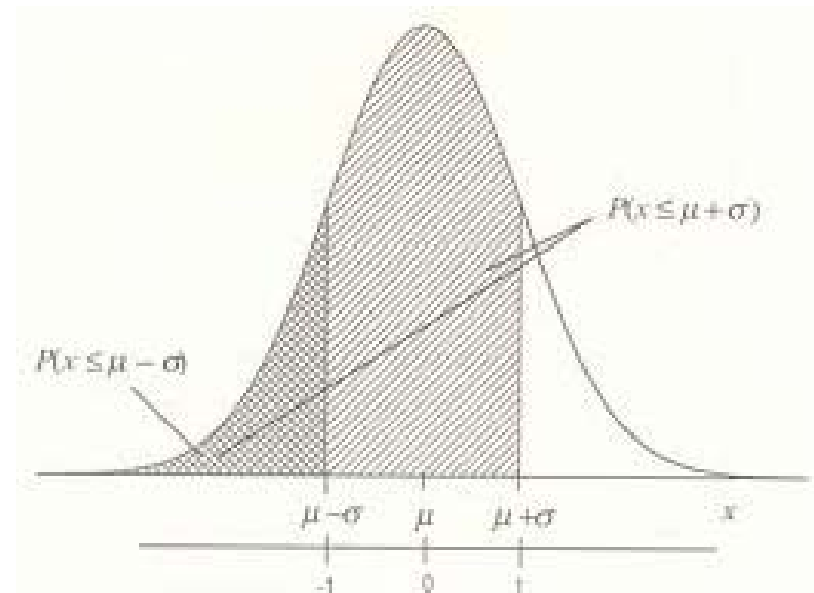
A mintavétel alapfogalmai – I.

- **Elem:** a vizsgálat azon alapegységei, amelyekről az információt gyűjtjük, az elemzés alapjai: emberek, családok, cégek, stb. – ld. korábban, az elemzési egységeknél tárgyaltak
- **Populáció:** a megfigyelt elemek elméleti összessége, amelyből a tényleges mintát vesszük. Pl.: Bármely fogyatékosági ághoz tartozó személyek összessége.
- **Vizsgálati populáció:** azon elemek összessége tartozik a vizsgálati populációba, melyek közül a mintát ténylegesen kiválasztjuk. Ez nem mindig fedi a tényleges populációt, hiszen sokan kimaradnak abból, hogy bekerülhessenek a mintába: pl. Telefonos elérhetőséggel rendelkező fogyatékossgal élő személyek.
- **Mintavételi egység:** azon elemek, illetve azon elemek csoportja, melyeket mintavétel során a kiválasztás(ok)nál figyelembe veszünk. Egylépcsős mintavételnél maga az elem, töblépcsősnél az egyes mintavételi lépcsők. Pl: Kétlépcsős mintavételnél: 1. lépcső: fogyatékossgai ágak, 2. lépcső: nemek
- **Mintavételi keret:** a mintavételi egységekről készült felsorolás. Egylépcsős mintavételnél a keret a (vizsgálati) populáció listája.
- **Megfigyelési egység:** olyan elemek, akiktől az adatokat konkrétan gyűjtjük. Lehet maga az elemzési egység, maga pl. az egyes személy, de nem feltétlenül, hiszen néha a nagykorú (nem feltétlenül értelmi) fogyatékos gyermek helyett az anya válaszol .
- **Változó:** az egymást kölcsönösen kizáró attribútumok halmaza. – vö. korábban...
- **Paraméter:** egy változó valamely összefoglaló jellemzője egy adott POPULÁCIÓBAN. Pl: Egy évfolyam átlagéletkora

A mintavétel alapfogalmai – II.

- **Statisztika:** egy változó valamely jellemzője a MINTÁBAN(a populációs paraméterek becslésére használjuk)
- **Mintavételi hiba:** Matematikai statisztikai eljárások alkalmazásával történő becslési eljárással nyert mutató
- **Megbízhatósági szint:** Annak valószínűsége, hogy a mintából kapott statisztikánk a paraméterrel egyezik (pl. 95%)
- **Konfidencia intervallum:** A mintából kapott statisztikánk szóródása(pl.+/- 5%)

A fenti két fogalom használata: A mintastatisztika helytállóságát lehet kifejezni vele: Mennyire lehetünk biztosak (**Megbízhatósági szint**) abban, hogy a statisztika a mintaparaméter körüli X intervallumban (**konfidencia intervallum**) helyezkedik el.



Konfidencia intervallumot NÖVELJÜK =>
Megbízhatóság: NÖVEKSZIK

.....

A valószínűségi mintavétel matematikai-statisztikai elméletének alapjai

- **Standard hiba:** megmutatja, milyen mértékben szóródnak a becslések a paraméter körül. (tulajdonképpen a mintaeloszlás szórása). A standard hiba nagysága függ a mintanagyságtól is, fordítottan arányosak egymással. Nagyobb a minta, kisebb lesz a standard hiba is. (a képlet miatt a standard hiba akkor feleződik, ha a mintanagyságot négyszeresére növeljük – a négyzetgyök miatt.)

$$s = \sqrt{\frac{P \times Q}{n}}$$

- *s – standard hiba*
- *P – az egyik attribútum valószínűsége*
- *Q – a másik attribútum valószínűsége*
- *n – a minta elemszáma*

Populációk és mintavételi keretek összefüggése:

- 1) egy mintáról megállapított állításokat csak arra az alapsokaságra tekinthetjük reprezentatívnak, amelyekből a mintát vettük (mintavételi keret).
- 2) a mintavételi keretek többségükben nem tartalmazzák az összes mérni kívánt elemet.
- 3) Figyelni kell, hogy a keretben található sokaság minden eleme csak egyszer (egyformán) legyen felsorolva, így egyenlő eséllyel kerülhetnek mintába. Ha valamelyik elem többször is bekerülhet, azokat a minta *túlreprezentálja*.

A mintavétel típusai – valószínűségi mintavétel

- **Az egyszerű véletlen mintavétel (EVM):**

Viszonylag ritkán alkalmazzák, hiszen többnyire kivitelezhetetlen, és vannak gyakran jobban és eredményesebben alkalmazható mintavételi eljárások is. Szükség van hozzá az összes elemet tartalmazó listára.

A lényege az, hogy a mintavételi keretet alkotó listán megjelöljük az összes elemet úgy, hogy minden elemet számításba vesszünk. Ezután véletlenszerűen (számítógép, véletlenszám táblázat) segítségével választjuk ki a minta elemeit.

- **Szisztematikus mintavétel:**

Az egyszerű véletlen mintavételnél szükség van az elemek listájára. Ha ez megvan, akkor gyakran inkább a szisztematikus módszert használják mintavételeknél. Ennek lényege, hogy a felsorolás pl. minden 5-ödik elemét veszik ki az alapsokaságból a mintába. (ez a mintavételi intervallum: az elemek közötti távolság) Azt, hogy ne lehessen bizonyos torzítás a mintában, az első számot véletlen módon választjuk ki jelen esetben az első 5 számból, utána pedig minden 5-ödik kerül a mintába. A kiválasztási arány a mintába kerülő elemek aránya az alapsokasághoz képest, itt: 1/5.

A módszer veszélye: ha a mintavételi keret valamilyen sorrend szerint csoportokból áll, periodikus. Pl.: Kosárcsapatokat kérdeznek meg, és a csapatokon belül posztok szerint vannak felsorolva a játékosok, és mi minden ötödiket választjuk ki, akkor mind egy posztról kerül majd a mintába, pl. mind center lesz.

- **Rétegzett mintavétel:**

A rétegzett mintavétel az egyszerű véletlen és a szisztematikus mintavétel módosítása, mellyel tovább növelhető a minta reprezentativitása. A lényege az, hogy megfelelő számban kerülhessenek elemek az alapsokaság bizonyos vizsgálni kívánt egység részcsoporthaiból. Ezt úgy érjük el, hogy a sokaságot homogén részcsoporthaokra osztjuk valamilyen kritérium, tulajdonság alapján, és ezen csoportokból megfelelő számú elemet választunk ki a mintába. A rétegzés attól függ, milyen információink vannak a listáról. Pl. évfolyam, nem. Bizonyos vizsgálatoknál ez fontos lehet, pl. ha kézilabdázókat nézünk, hogy meccsenként hány gólt dobznak, de előtte posztok szerint rétegzést készítünk, hiszen a gólok száma függ-(het) a poszttól.

- **Többlépcsős csoportos mintavétel:**

Vannak olyan sokaságok, melyekről nem tudunk listát készíteni, beszerezni. Ekkor használjuk azt az eljárást, hogy az elemek csoportjaiból vesszünk mintát, majd második lépésben az így kapott csoportokon belül vesszünk újabb mintát. A többlépcsős csoportos mintavétel jelentése, hogy egymás után többször készítünk listát és alkalmazzuk a kiválasztást.

Pl.: populáció: Magyarország könyvtárainak taglistája, ebből akarunk valamit megtudni az olvasókról. *Első lépés*: mintát vesszünk a könyvtárak közül (pl. szisztematikus minta, bekerül minden 10-dik könyvtár neve). *Második lépés*: a kiválasztott könyvtárak tagjairól listát kérünk, és azokból mind vesszünk mintát. Itt is be lehet vezetni a rétegzést, pl. a taglistánál gyermek és felnőtt olvasókat szétválasztjuk.

- **Elemzámmal arányos valószínűségű mintavétel (EAV)**

Olyan esetekben érdemes használni ezt az eljárást, amikor a vizsgálni kívánt csoportok mérete egymástól jelentősen eltér. Ilyenkor is olyan mintát kapunk, amelybe minden elem egyforma eséllyel kerülhet be. Ehhez a csoport méretével azonos valószínűséget kell adnunk a csoport elemeinek, hogy bekerüljenek a mintába. Tehát pl. kétszer akkora eséllyel kap a mintába kerülésre az a Budapesti kerület, mely 100.000 lakosú, mint az 50.000 lakosú. Ennek módja: SÚLYOZÁS

A mintavétel típusai – nem valószínűségi mintavétel

Alkalmazása: Ha nincs lehetőség a valószínűségi mintavételre.

Pl. : „A piros lámpán átkelő gyalogosok közlekedési kultúrája” c. kutatásban – nem áll rendelkezésre lista a populációról.

Előnye: Olcsóbb, egyszerűbb

Hátránya: Kevésbé megbízható

Típusai:

- **Szakértői mintavétel:** olyan módszer, mely során az általunk tudott információk, megítélések alapján választunk mintát az alapsokaságról. Ez alap információk gyűjtésére alkalmas lehet, de nem tekinthető reprezentatívnak semmilyen alapsokaságra. Pl. egy jelzőlámpánál lévő újságárust kérdezek meg., aki jól ismeri a helyieket és évek óta figyeli az átkelőt..
- **Kvótás mintavétel:** a kvótás mintavételi eljárás alapja az ún. kvótamátrix, mely a célpopuláció összes jellemzője szerint sorolja táblázatba az elemeket. Ehhez minden szükséges tulajdonságról szükségünk van a pontos adatokra, amilyen kategóriákat a táblázatba felveszünk. (pl. nem, életkor, vallás) Itt a hibalehetőség, hiszen az adatoknak helyesnek kell lenniük! Ezután a táblázat kockáiba beírjuk a hozzá rendelhető tulajdonságú elemek gyakoriságát. Ezután adatokat gyűjtünk az egyes cellákban lévő tulajdonságokkal rendelkező személyektől, majd súlyozzuk őket az összpopulációban képviselt arányuknak megfelelően. Így egy reprezentatív mintát kapunk.
- **Egyszerűen elérhető alanyokra vonatkozó mintavétel:** lényege: a kutató csak olyan elemeket (embereket) vizsgál, akik könnyen elérhetőek, mondjuk épp ott sétál. Általánosítható adatokat ritkán érünk el vele, csak akkor jó, ha ténylegesen azok az emberek érdekelnek, akiket vizsgálók. Pl.: egy meccs előtt sorban állva megkérdezett 100 szurkoló véleménye, akik épp szembe jöttek a pénztárnál, még nem reprezentatív az összes nézőre nézve.