



2008. X.6.

# Mintavételi eljárások

Daróczi Gergely, PPKE BTK

# Óravázlat

- A mintavétel célja
- Alapfogalmak
  - Alapsokaság, mintavételi keret, megfigyelési egység, mintavételi egység ...
- Nem valószínűségi mintavételezési technikák
  - Könnyen elérhető alanyok megkérdezése
  - Szakértői mintavétel
  - Hólabda
  - Kvótás mintavétel
- Valószínűségi mintavételezési technikák
  - Egyszerű és szisztematikus mintavétel
  - Rétegzett mintavétel
  - Többlépcsős technikák
- A lebonyolítás módjai
  - Önkitöltős / kérdezőbiztos által felvett kérdőívek
  - Posta / telefon / Internet / SMS ... (csatorna)
  - Egyéb megfontolások

## A mintavétel célja

- Nem áll rendelkezésre megfelelő információ, ismeret az érintett csoporton belül (pl.: egyetemisták zenehallgatási szokásai)
- Teljes populáció megkérdezésének lehetetlensége, nehézsége (költséghatékonyság, avagy: a korlátozott racionalitás jótékony hatásai)
- Az alapsokaság egyes tulajdonságainak, paramétereinek becslése annak egy *kiválasztott* része alapján

Felderítő kutatás („Vajon hallgatnak-e zenét...”)

Leíró adatok gyűjtése („Milyen zenét hallgatnak...”)

Hipotézisvizsgálat (elméletek, modellek tesztelése)

(„Tényleg a trash-metal a legkedveltebb műfaj?”)

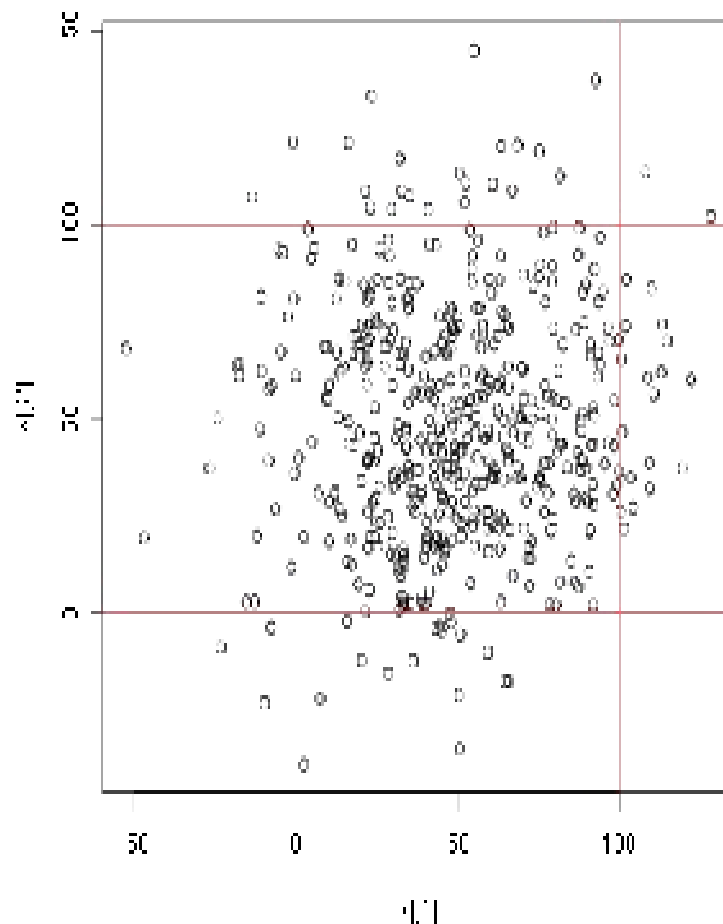
Minőségbiztosítás („A kortárs zene jótékony hatásai...”)

Döntéshozatal segítése („A büfében érdemes-e egy állandó DJ-t alkalmazni?”)

# Alapfogalmak

- Alapsokaság: a vizsgálati populáció (elméletileg meghatározott) összes eleme
- Mintavételi keret: a kiválasztásnál figyelembe vett (elérhető) elemek összessége
- Megfigyelési egység: az alapsokaság elemeinek tekintett egységek
- Mintavételi egység ( $\neq$ alapsokaság elemi egységei): a kiválasztásnál figyelembe vett, legegységibb egységek

**Pontos meghatározás!**

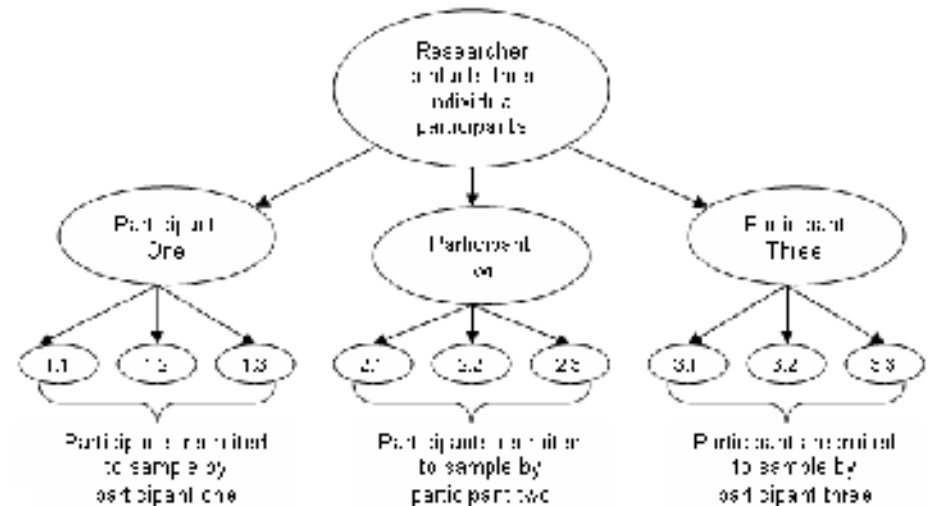


# Nem valószínűségi mintavétel

- A minta elemei nem valószínűségi alapon kerülnek kiválasztásra
- A minta torzítása (alapsokaságtól való eltérése, ill. annak nagysága) matematikailag nem számítható
- Bizonyos helyzetekben elkerülhetetlen (vagy nem célszerű) az alkalmazásuk
- Gyors, költséghatékony módszerek

## Típusai:

- egyszerűen elérhető alanyok megkeresése
- szakértői mintavétel
- hólabda módszer
- kvótás mintavétel (ld. Gallup)



# Kvótás mintavétel

- 1936: Gallup sikeresen „megjósolta” Rooseveltt győzelmét (vs Literary Digest – önkitöltős, postai kérdőívek milliós nagyságrendje)
- Az alapsokaságot leíró mátrix alapján kerülnek az egyedek kiválasztásra (nem konkrét célszemélyek!)

<b>Egyetemisták</b>	<b>Fekete</b>	<b>Barna</b>	<b>Szőke</b>	<b>Vörös</b>
<b>Fiú</b>	432	543	643	23
<b>Lány</b>	234	436	387	63

- A megkérdezett minta súlyozása → bizonyos jellemzőkre reprezentatív felmérés
- Hátrányok, buktatók: ld. 1948
  - Megfelelő kiindulási mátrix
  - Megfelelően kiválasztott tulajdonságok
  - „Marginális” elemek figyelmen kívül hagyása

# Valószínűségi mintavétel

- Minden egyed azonos (pontosabban: meghatározott) valószínűséggel kerül kiválasztása
- **Reprezentativitás általános meghatározása:** az alapsokaság minden eleme meghatározott, nem nulla valószínűséggel kerülhet kiválasztásra (pontosítást igényel!)
- A valószínűségek alapján lehetőség nyílik a mintavétel során elkövetett „hibát” (**mintavételi hiba**), a minta torzítását számolni, azaz meghatározni azt, hogy minta által felvázolt jellemzők milyen jól jellemzik az alapsokaságot

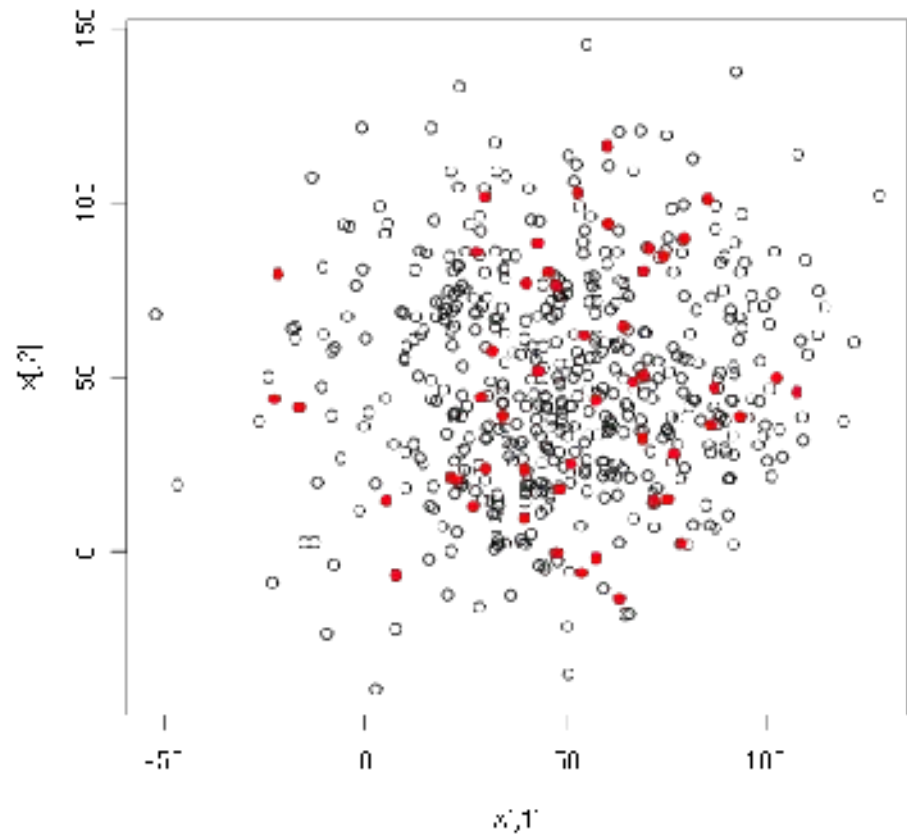
**Megbízhatósági szint:** a minta alapján számolt becslések milyen valószínűséggel lesznek igazak az alapsokaság tagjaira

**Mintavételi hiba:** a minta alapján becsült paraméter milyen mértékben ingadozik a valós érték körül (konfidencia intervallum)

Pl.: 95 százalékos megbízhatósági szinten elmondható, hogy a Blikk címlapján 3-5 szupersztár neve jelenik meg nap, mint nap

# Egyszerű véletlen mintavétel (SRS)

- A mintavételi keret minden tagja ugyanakkora valószínűséggel kerülhet kiválasztásra ( $P=n/N$ )
- Az alapsokaság minden tagjának ( $\rightarrow$  mintavételi keret) ismerni kell elérhetőségét, az elemeket listaiba kell tudnunk rendezni
- A listáról  $n$  (=mintaelemszám) darab, véletlenül kiválasztott egyedet választunk be a mintába





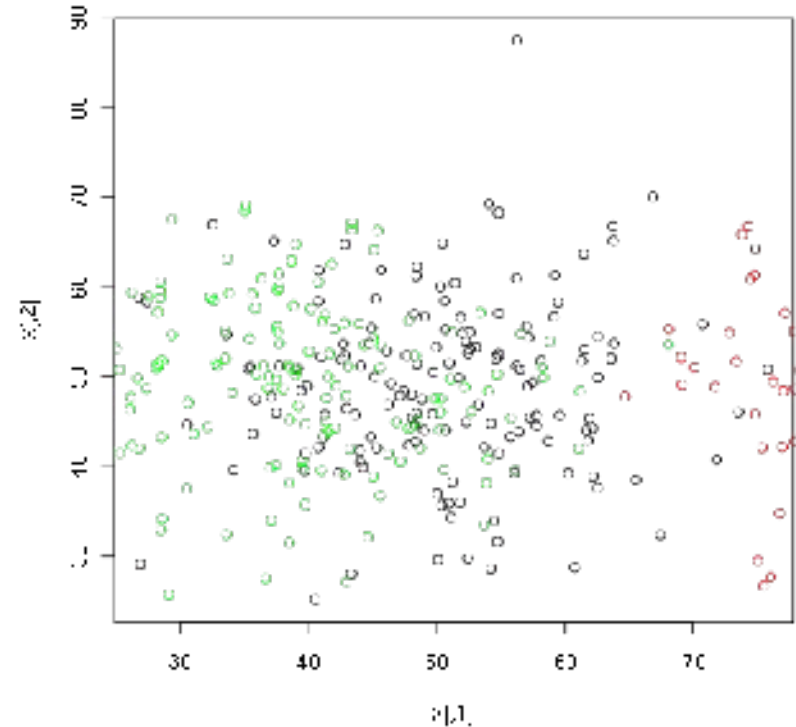
## Egyszerű véletlen mintavétel (SRS) II.

- A véletlen választás történhet:
  - véletlenszám-táblázat vagy számítógép segítségével
  - a lista véletlen szerinti elrendezése után szisztematikusan (minden  $k$ . elem kiválasztása, ahol  $k=n/N$ )
- A mintavételi hiba függ:
  - Minta homogenitásától (szórásától)
  - A minta nagyságától
  - Az alapsokaság elemszámától
- A szükséges mintanagyság számolható!
  - Hüvelykujj-szabály:  $n > 1.000$  fő elfogadható reprezentativitást garantál (nem minden esetben - ld. a mintavételi hibát meghatározó tényezők)

$$Se = \frac{S}{\sqrt{n}} * \sqrt{1 - \frac{n}{N}}$$

# Rétegzett mintavétel (stratification)

- Célja: a minta bizonyos változók szerinti teljes reprezentativitása
- Módszere: a mintavételi hiba csökkentése a homogenitás növelésével (a minta „rétegekre” bontásával)
  - Egy nagy, véletlen minta helyett számos kisebb, homogén részcsoport vizsgálata
  - A részcsoportok kiválasztása előzetes feltételezéseken alapul
  - Nem megfelelő réteg-képzés nagyobb mintavételi hibával jár együtt (mint akár SRS esetén!)
  - Legegyszerűbb módja: a mintavételi keret tagjainak listájának rétegek szerinti rendezése, majd szisztematikus mintavételezés



# Az egyszerű és a rétegzett mintavétel összehasonlítása

ALAPSOKASÁG		
	Hajszálok száma (100e)	
Fiú	1	3
Lány	7	9
Átlag	5	

Véletlen mintavétel		
Minta	Átlag	Valószínűség
1,7	4	1/6
1,9	5	1/6
3,7	5	1/6
3,9	6	1/6
1,3	2	1/6
7,9	8	1/6

Véletlen mintavétel		
Minta	Átlag	Valószínűség
1,7	4	1/4
1,9	5	1/4
3,7	5	1/4
3,9	6	1/4

Véletlen mintavétel:

$$\text{átlag} = (4+5+5+6+2+8)/6 = 5$$

$$\text{szórás} = (1+0+0+1+9+9)/6 = \mathbf{3,33}$$

Rétegzett mintavétel:

$$\text{átlag} = (4+5+5+6)/4 = 5$$

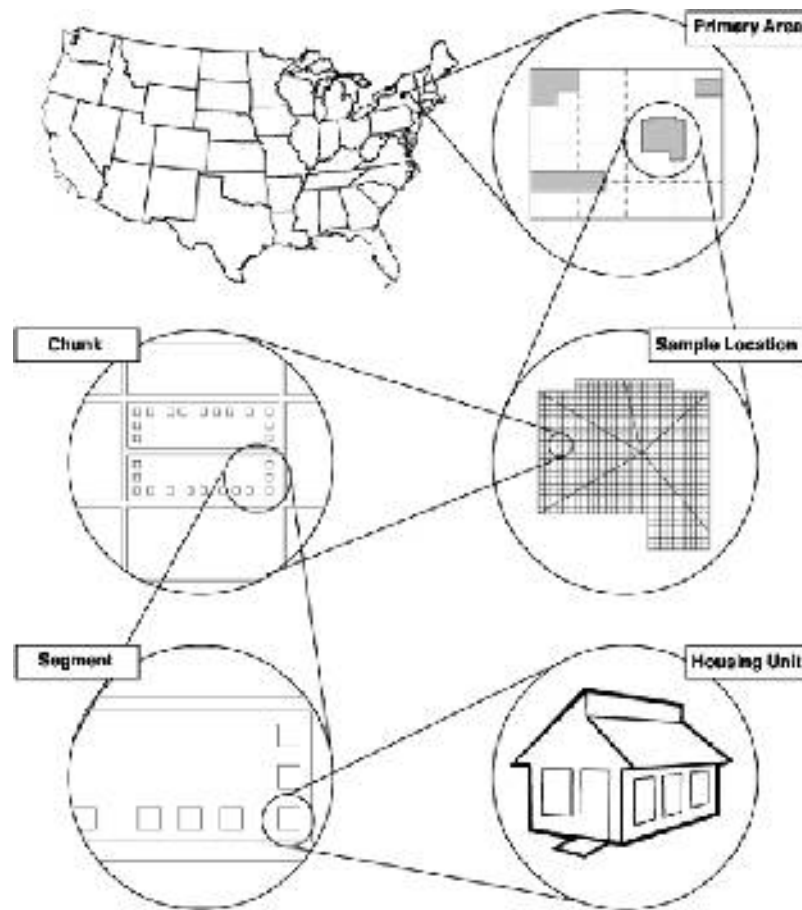
$$\text{szórás} = (1+0+0+1)/4 = \mathbf{0,5}$$

*Átlag: az egyes minták átlagainak átlaga (összegük / minták száma)*

*Szórás: a minták átlagainak eltérése az összesített átlagtól (egyedeltérések négyzetének összege / minták száma)*

„Rossz” rétegzés esetén: (1,7) (3,9) a szórás értéke **4,5** lesz!!!

# Többlépcsős csoportos mintavétel (clustering)



- Bizonyos esetekben nem áll rendelkezésre teljes lista a célcsoportról (vagy nem célszerű az alapján dolgozni)
- A mintavételi egységek több lépésen keresztül is elérhetőek valószínűségi alapon
- Viszont: minden egyes lépcső esetén újabb mintavételi hiba jelentkezik!

# A minta elérésének módjai

- Önkitöltős postai kérdőív
  - Olcsó (alacsony emberi erőforrás)
  - Kényes kérdések magasabb megváltási aránya
  - Összességében magasabb válasszmegtagadási arány
  - Alacsony visszaküldési arány (60% ?)



## – Kérdezőbiztosok által felvett adatok

- Megbízható (problémás pontok tisztázása a kérdezőbiztos által)
- Alacsonyabb válasszmegtagadási arány
- Kényes kérdések nehezen kezelhetőek
- Kérdezőbiztosok megbízhatósága, megfelelő felkészítése?
- Magas emberi erőforrás-szükséglet

# A kérdezést, adatfeldolgozást megkönnyítő eszközök

- Computer Aided Telephone Interview
- Computer Aided Personal Interview
- On-line kutatások
- SMS / Wap kérdőívek
- Emberi közreműködést nem igénylő telefonos kérdezések

