

A társadalomkutatás módszerei I.

9. hét

Daróczi Gergely

Budapesti Corvinus Egyetem

2011. november 10.



1. Zh eredmények
2. Újra a hibatényezőkről
3. A mintavételi keret
4. Valószínűségi mintavételi eljárások
 - Ismétlés
 - Egyszerű véletlen mintavétel
 - A mintavételi hiba
 - Példák
 - Rétegzett mintavétel
 - Szisztematikus mintavétel
 - Csoportos mintavétel

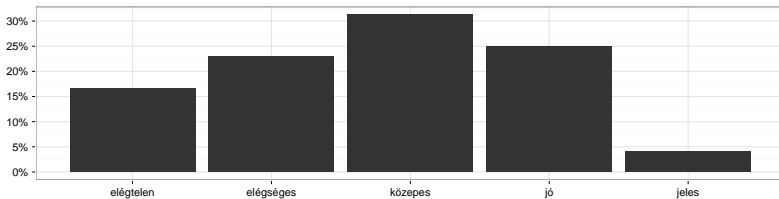
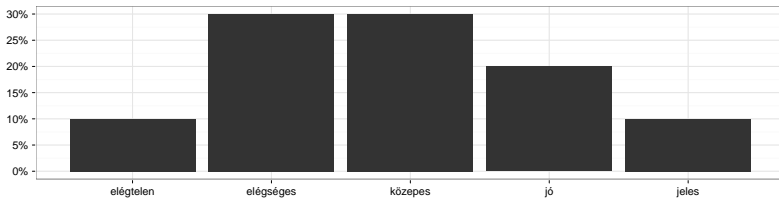
1. Zh

Egyéni eredmények

<u>Neptun</u>	Szerzett pontok													2. Zh min. pontszáma			
	1	2	3	4	5	6	7	8	9	10	11	12	Σ	5	4	3	2
<i>n pontszám</i>	1	1	2	2	3	3	2	2	1	2	2	4	25	20	14	8	1
BLKRJW	1	1	2	1	0	2	1,5	1	0	1	2	3,75	16,3	28,8	22,8	16,8	9,75
FNHLWF	0,5	1	2	2	3	1,5	1	0,5	1	1	1	4	18,5	26,5	20,5	14,5	7,5
AE0JFY	Nem írt, nincs lehetőség a pótlásra																
BREISZ	Pótlás																
GB7J7L	Pótlás																
YX9SRH	1	1	2	1	3	2,8	2	1,5	1	0	2	3,5	20,8	24,3	18,3	12,3	5,25
BD422G	0	1	2	2	3	1,5	2	2	0	1,5	2	2,75	19,8	25,3	19,3	13,3	6,25
I7PO5V	Pótlás																
Q01CHT	1	1	2	1,5	3	2,8	2	0,5	0,5	1	2	3	20,3	24,8	18,8	12,8	5,75
G0NTUR	0,5	1	2	1,5	3	1,5	1,5	2	0,5	0	2	2,5	18	27	21	15	8
H8BURC	0,8	1	2	2	3	3	2	2	1	1	1,9	2,75	22,4	22,6	16,6	10,6	3,6
NKB50P	0,5	1	1	0,5	2	1,5	2	0,5	0	1	2	2	14	31	25	19	12
D3AG6M	Nem írt, nincs lehetőség a pótlásra																
BYTBAT	0,1	1	1,5	1	3	3	1,5	1	0,8	1,5	1	2,75	18,2	26,9	20,9	14,9	7,85
J32XQA	0,5	0,5	1,5	0,5	0	0	2	1	0	1,5	1,5	2,6	11,6	33,4	27,4	21,4	14,4
I5G575	Nem írt, nincs lehetőség a pótlásra																
ROCX7A	0,5	1	1	1	0	2	2	1,5	0	1,5	2	2,5	15	30	24	18	11

1. Zh

Összessített eredmények



A *Time* az '50-es évek végén
a következő kutatási eredményt jelentette meg:

"the average Yaleman,
class of 1924,
makes \$ 25,111 a year"

Ez az összeg mai viszonylatban megközelítőleg 150,000 USD-t
(34,000,000 Ft) tesz ki!

Újra a hibatényezőkről

A hiba lehetséges okai

A *Time* becslése, mint kiderült, egy olyan kutatásra támaszkodik, amelyben a kérdőíveket postai úton juttatták el azon 1924-ben végzett volt hallgatókhoz, akiknek a címe ismert volt a Yale egyetemen.

Milyen módszertani problémák vethetők fel a mintavételezéssel kapcsolatban?

A *Time* becslése, mint kiderült, egy olyan kutatásra támaszkodik, amelyben a kérdőíveket postai úton juttatták el azon 1924-ben végzett volt hallgatókhoz, akiknek a címe ismert volt a Yale egyetemen.

Milyen módszertani problémák vethetők fel a mintavételezéssel kapcsolatban?

- 1 mintavételi hiba,
- 2 válaszmegtagadásból eredő hiba,
- 3 torzító válasz.

A mintavételi keret

Egy kevésbé szerencsésen megválaszott minta/mintavételi keret

Egy most induló piackutató cég tervei szerint publikus telefonkönyvből gyűjti majd az elérhetősegeket mintavételi keretének felépítéséhez. A kutatásai során a későbbiekben az szolgál majd alapul a várhatóan reprezentatív, egyszerű véletlen minták kiválasztásához.

Győzzük meg a cég vezetőit, hogy döntésük nem szerencsés módszertani szempontból!

A mintavételi keret

Egy kevésbé szerencsésen megválaszott minta/mintavételi keret

Egy most induló piackutató cég tervei szerint publikus telefonkönyvből gyűjti majd az elérhetősegeket mintavételi keretének felépítéséhez. A kutatásai során a későbbiekben az szolgál majd alapul a várhatóan reprezentatív, egyszerű véletlen minták kiválasztásához.

Győzzük meg a cég vezetőit, hogy döntésük nem szerencsés módszertani szempontból!

- 1 csak azok szerepelnek a listában, akik rendelkeznek vezetékes/mobil számmal,
- 2 csak a nyilvános számokat tartalmazza a lista,
- 3 bizonyos számok esetében a közvélemény-kutatás/marketing cél nem megengedett,
- 4 ugyanazt a számot többször nem kereshetjük rövid időszakon belül,
- 5 csak azokat érjük el végül, akik hajlandóak válaszolni kérdéseinkre.

A mintavételi keret

A lehetséges problémákról

Leslie Kish (1965): *Survey Sampling* könyvében négy alapvető problémát határozott meg az alapsokaságból képzett mintavételi keret lehatárolásával kapcsolatban:

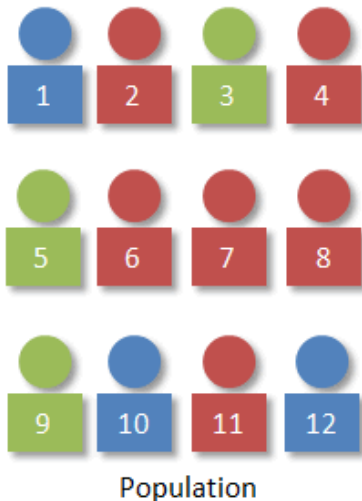
- 1 **Missing elements:** az alapsokaság egyes elemei kimaradnak a mintavételi keretből.
- 2 **Foreign elements:** olyan elemek is bekerülnek a mintavételi keretbe, akik nem részei az alapsokaságnak.
- 3 **Duplicate entries:** az alapsokaság egyes elemei többször kerülnek lekérdezésre.
- 4 **Groups or clusters:** a mintavételi keret esetek/egyedek csoportjait tartalmazza.

***Véletlen* mintaválasztás történik-e, amikor:**

- Budapest egy véletlen módon kiválasztott buszmegállójában megkérdezek minden harmadik embert?
- a reprezentativitás elérése érdekében a mintatagokat az alapsokaság arányában választjuk ki: a kérdezőbiztosok 40 férfit és 60 nőt kérdeznek le a Campus véletlen bejárása során.
- felhívunk 500 számítógép által generált telefonszámot (l. 2+7 véletlen számjegy)?
- a kurzus látogatóinak neveivel ellátott papírfecniket tartalmazó kalapból csupa női nevet húzunk ki?
- matematika órán a tanár a napló „felcsapásával” választja ki a felelőt?

Egyszerű véletlen mintavétel

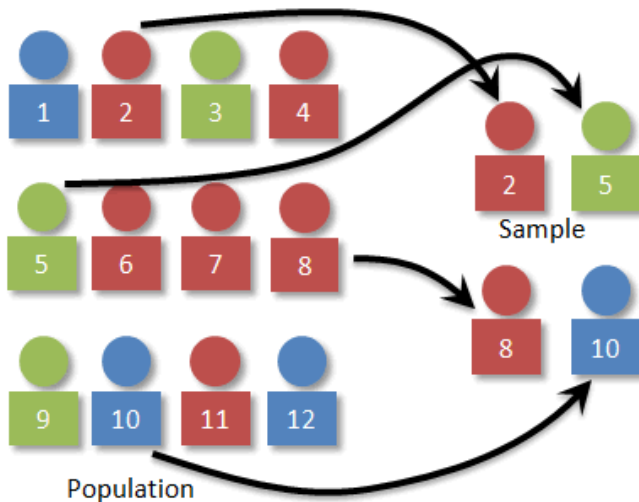
A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Egyszerű véletlen mintavétel

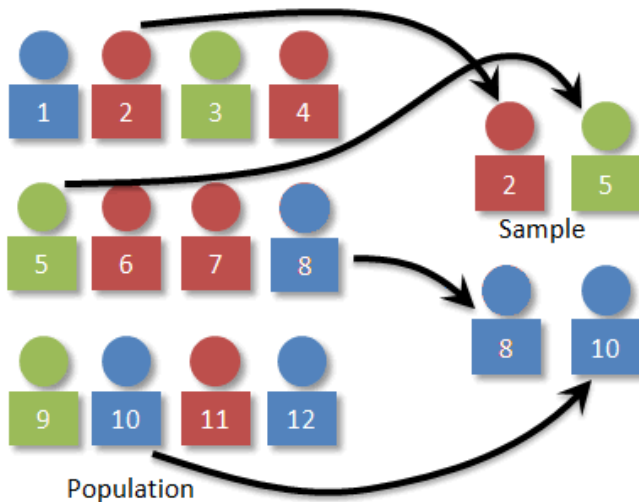
A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Egyszerű véletlen mintavétel

A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Ismétlés: a SRS esetén alkalmazható számítások

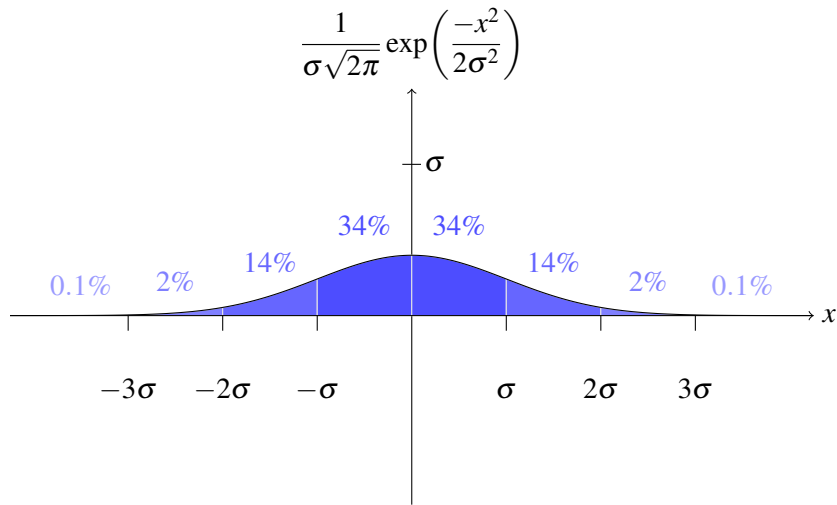
- **számtani átlag:** $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ (becslőfüggvény)
- **korrigált empirikus szórás:** $\sigma = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}$
- **standard/mintavételi hiba:** $SE = \frac{\sigma}{\sqrt{n}} \cdot FPC$
- **Finite Population Correction:** amennyiben a minta az alapsokaságnak nagyobb hányadát (>5%) teszi ki

$$FPC = \sqrt{1 - \frac{n}{N}}$$

$$SE = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$$

Egyszerű véletlen mintavétel

A mintavételi hiba | eloszlás



standard normális eloszlás: $\bar{x} = 0, \sigma = 1$

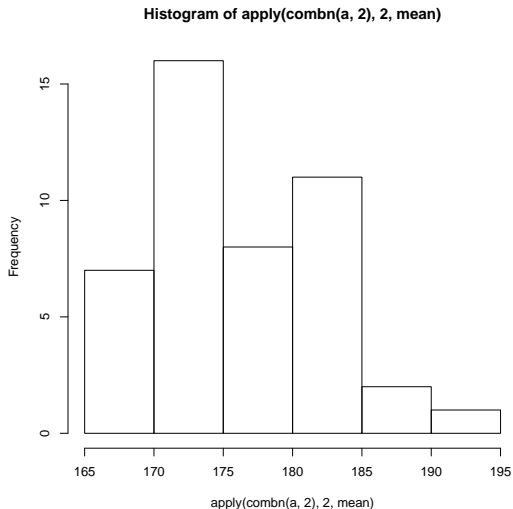
Egyszerű véletlen mintavétel

A mintavételi hiba | eloszlás

```
> a
[1] 170 174 175 182 193 191 167 168 171 172
> combn(a, 2)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]  170  170  170  170  170  170  170  170  170  170
[2,]  174  175  182  193  191  167  168  171  172  172
      [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25] [,26]
[1,]   175   175   175   175   175   175   175   175   182
[2,]   182   193   191   167   168   171   172   193
      [,33] [,34] [,35] [,36] [,37] [,38] [,39] [,40] [,41]
[1,]   193   193   193   191   191   191   191   167
[2,]   168   171   172   167   168   171   172   168
```

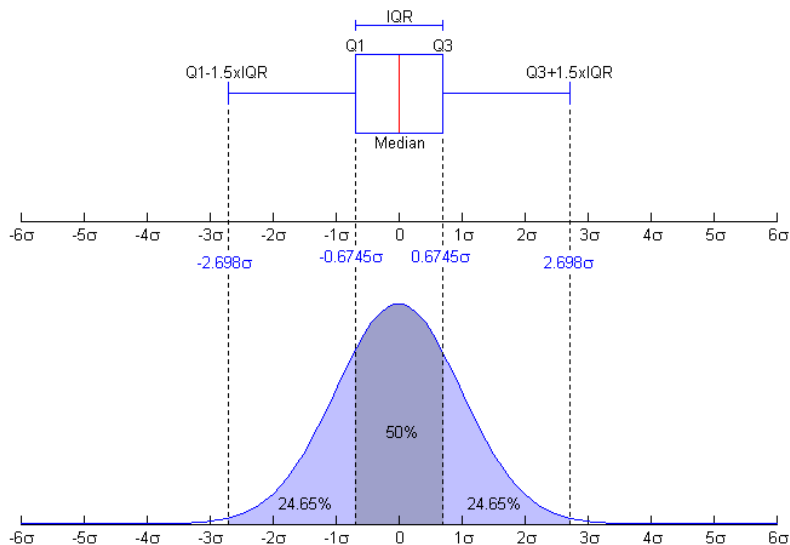
Egyszerű véletlen mintavétel

A mintavételi hiba | eloszlás

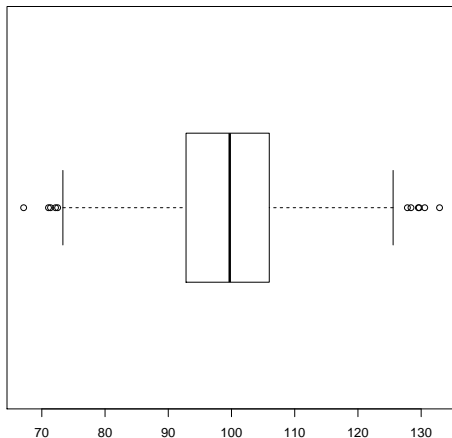


Leíró statisztikák

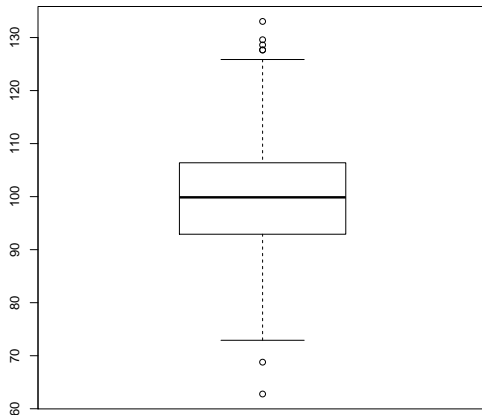
Interkvartilis terjedelem



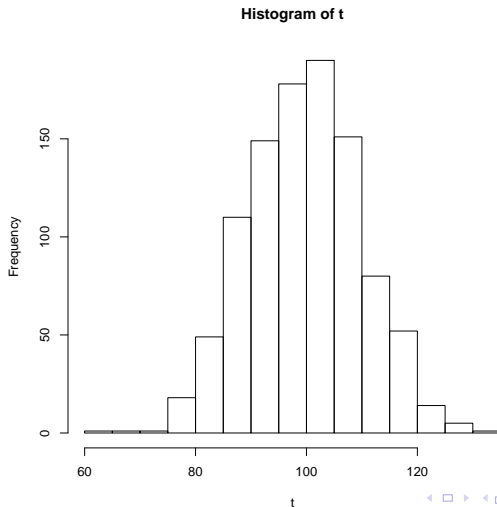
A tófalusi fesztiválon végzett intelligencia-kutatás eredménye:



A tófalusi fesztiválon végzett intelligencia-kutatás eredménye:



A tófalusi fesztiválon végzett intelligencia-kutatás eredménye:



A mintavételi hiba

Egy egyszerű példa

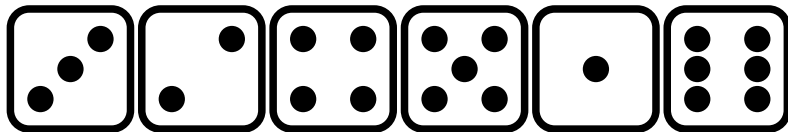
Játékszabály

Egy hatoldalú dobókockával gurítunk:

Ha a dobás páratlan, a játékos a dobott összeget megnyeri.

Amennyiben a dobás páros, úgy ő fizet két dollárt a banknak.

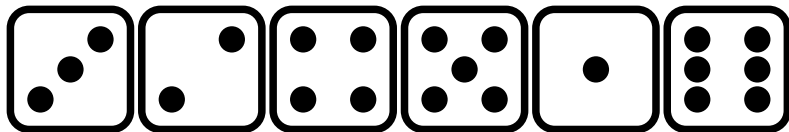
Adatgyűjtés: előzetes számítások helyett belevetjük magunkat a játékba, és a következő értékeket dobjuk.



Folytassuk a játékot?

A mintavételi hiba

Egy egyszerű példa



$$X = \{-2, 2, 4, -2, -2, 6\}$$

$$\bar{x} = \frac{-2+2+4+2+2+6}{6} = \frac{6}{6} = \frac{1}{1} = 1$$

$$\sigma = \sqrt{\frac{(-2-1)^2 + (2-1)^2 + (4-1)^2 + (-2-1)^2 + (-2-1)^2 + (6-1)^2}{5}} =$$
$$= \sqrt{\frac{9+1+9+9+9+25}{5}} = \sqrt{\frac{62}{5}} = \sqrt{12.4} = 3.521363$$

$$SE = \frac{3.521363}{\sqrt{6}} = \frac{3.521363}{2.44949} = 1.437591$$

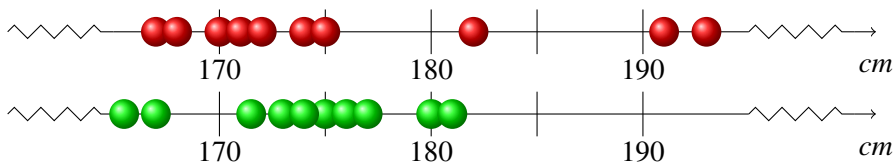
A várható érték valahol -1.87 és 3.87 között (CI)
található 95%-os döntési szinten.

Hogyan döntenénk?

A mintavételi hiba

Egy egyszerű példa

10-10 diák magasságát mértük 2 osztályteremben:

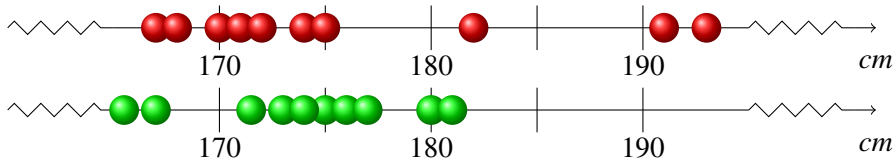


Melyik osztály diákjai a magassabbak a leíró statisztikák alapján adható becslések alapján?

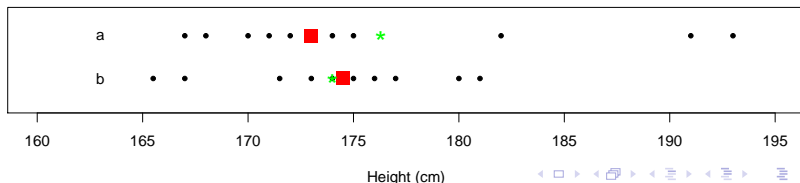
A mintavételi hiba

Egy egyszerű példa

10-10 diák magasságát mértük 2 osztályteremben:



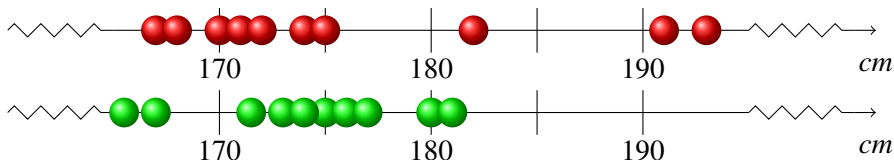
Melyik osztály diákjai a magassabbak a leíró statisztikák alapján adható becslések alapján?



A mintavételi hiba

Egy egyszerű példa

10-10 diák magasságát mértük 2 osztályteremben:



$$\bar{A} = 176.3, S_A^* = 9.2862, n = 10 \Rightarrow SE = 2.9366, CI_{66\%} = (173.36, 179.24)$$

$$\bar{B} = 174, S_A^* = 5.0277, n = 10 \Rightarrow SE = 1.5899, CI_{66\%} = (172.41, 175.59)$$

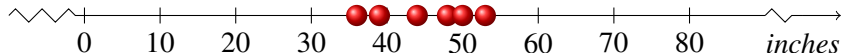
A mintavételi hiba

Egy egyszerű példa

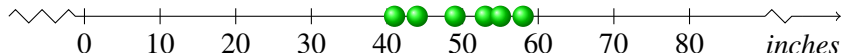
Két időpontban mértük (*inch*) a nevelt gesztenyefák magasságát egy faiskolában.

Határozzuk meg a minták átlagát, szórását és a mérés (mintavételi) hibáját! Ezek alapján mondhatjuk-e, hogy az eltelt időszakban a fák kimutatható módon nőttek?

1 **2011. március 22.:** 36 48 50 44 53 39



2 **2011. április 1.:** 41 53 55 49 58 44



A mintavételi hiba

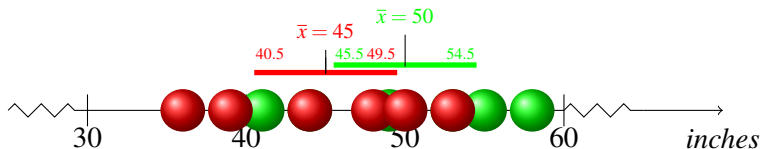
Egy egyszerű példa

Két időpontban mértük (*inch*) a nevelt gesztenyefák magasságát egy faiskolában.

Határozzuk meg a minták átlagát, szórását és a mérés (mintavételi) hibáját! Ezek alapján mondhatjuk-e, hogy az eltelt időszakban a fák kimutatható módon nőttek?

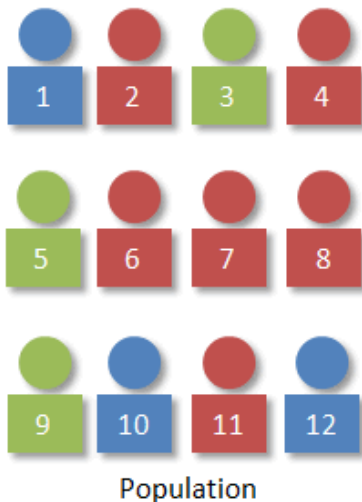
1 **2011. március 22.:** 36 48 50 44 53 39

2 **2011. április 1.:** 41 53 55 49 58 44



Rétegzett mintavétel

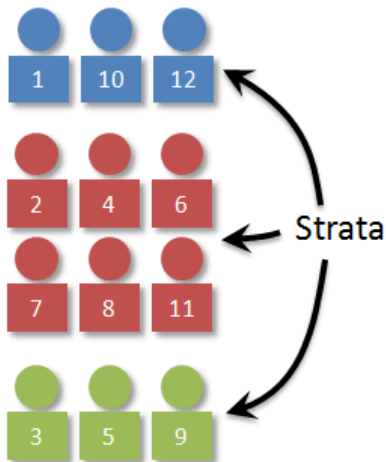
A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Rétegzett mintavétel

A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Rétegzett mintavétel

A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Rétegzett mintavétel

A mintavételi hiba

Négy hallgatót kérdeztünk meg arról, hogy hány macskát tart otthon:

	Budapest	vidék
Lányok	9	7
Fiúk	3	1

Mit gondolunk, hogyan alakulna a mintavételi hiba ha egyszerű véletlen mintát, és hogyan, ha rétegzett mintát vennénk?

Négy hallgatót kérdeztünk meg arról, hogy hány macskát tart otthon:

	Budapest	vidék
Lányok	9	7
Fiúk	3	1

Mit gondolunk, hogyan alakulna a mintavételi hiba ha egyszerű véletlen mintát, és hogyan, ha rétegzett mintát vennénk?

Két fős mintákat választva:

- 1 SRS: 6 lehetséges minta: (1,7) (1,9) (3,7) (3,9) (1,3) (7,9)

$$\bar{x} = \frac{4+5+5+6+2+8}{6} = 5,5^* = \frac{1+0+0+1+9+9}{6} = 3.33$$

- 2 Rétegzett: 4 lehetséges minta: (1,7) (1,9) (3,7) (3,9)

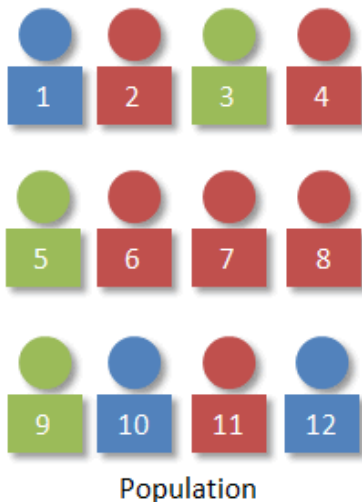
$$\bar{x} = \frac{4+5+5+6}{4} = 5,5^* = \frac{1+0+0+1}{4} = 0.5$$

- 3 Rétegzett: 4 lehetséges minta: (1,3) (1,9) (3,1) (3,7)

$$\bar{x} = \frac{2+5+2+5}{4} = 3.5, S^* = \frac{1.5+1.5+1.5+1.5}{4} = 1.5$$

Szisztematikus mintavétel

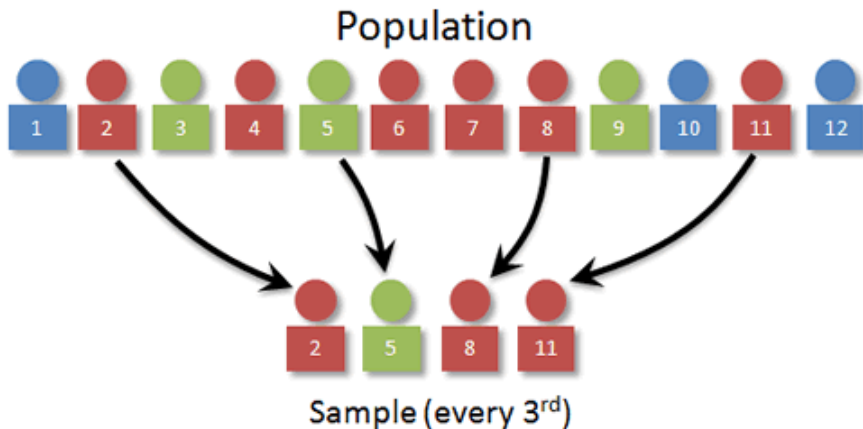
A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Szisztematikus mintavétel

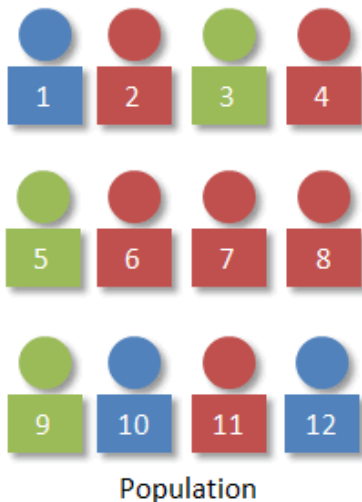
A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Csoportos mintavétel

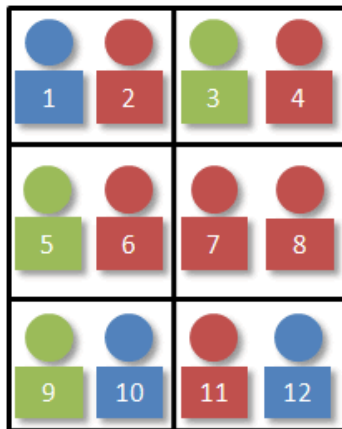
A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Csoportos mintavétel

A kiválasztás menete

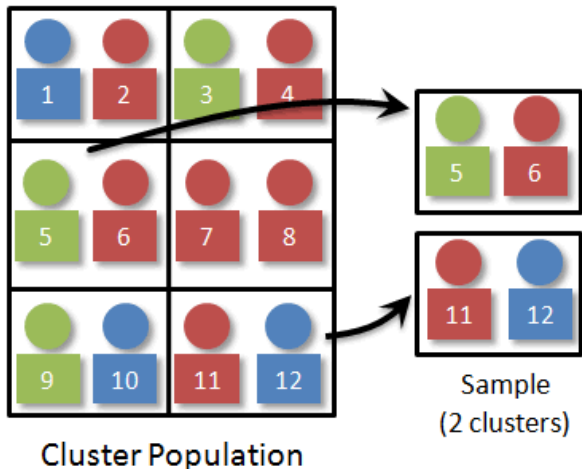


Cluster Population

Forrás: Dan Kerlner, Elgin Community College

Csoportos mintavétel

A kiválasztás menete



Forrás: Dan Kerlner, Elgin Community College

Köszönöm a figyelmet!

Daróczi Gergely

daroczi.gergely@btk.ppke.hu