

A társadalomkutatás módszerei I.

13. hét

Daróczi Gergely

Budapesti Corvinus Egyetem

2011. december 8.



- 1 A mintaválasztás célja
- 2 Alapfogalmak
- 3 Mintavételi eljárások
- 4 További fogalmak
- 5 Mintavételi hiba számítása
- 6 Mintanagyság számítása
- 7 Reprezentativitás
- 8 Egyéb kérdések

A mintaválasztás

A mintaválasztás célja

Miért vegyünk mintát?

- Nem áll rendelkezésre megfelelő információ, ismeret az érintett csoporton belül (pl. „*karácsonyi ajándékozás Póföldén*”).
- Teljes populáció megkérdezésének lehetetlensége, nehézsége (költséghatékonyság, korlátozott racionalitás).
- Az alapsokaság egyes tulajdonságainak, paramétereinek becslése annak egy kiválasztott része alapján.

Mi célból, hogyan vegyünk mintát?

- Felderítő kutatás
 - Leíró adatok gyűjtése
 - Hipotézisvizsgálat
 - Minőségbiztosítás
 - Döntéshozatal segítése
- } esetén

- Populáció
- Mintavételi keret
- Minta
- Vizsgálati populáció
- Elem
- Mintavételi egység

Mintavételi egység, válaszadó, megfigyelési egység, eset

Határozzuk meg a fenti fogalmakat:

- népszámlálás
 - telefonos piackutatás
 - árvízi mentés hatásvizsgálat
- } esetén

A mintaválasztás

Valószínűségi és nem-valószínűségi mintavétel

Valószínűségi mintavétel:

amikor a populáció (I. mintavételi keret) minden eleme (a mintavétel előtt már) ismert, nem nulla (és egyenlő) eséllyel kerül a mintába.

Feladatok:

- definíció,
- típusok,
- előnyök/hátrányok,
- mikor használjuk,
- példa.

Nem-valószínűségi mintavétel:

minden egyéb kiválasztás.

„A Bime Analytics most összegyűjtötte az alapadatokat a [Google+] 25 millió felhasználójáról. A minta nem reprezentatív, a kérdésekre önként válaszolt, akinek kedve volt hozzá, de az összesített adatokból jól látszanak a trendek.”

„a szajt felhasználóinak 70 százaléka férfi [. . .] kezdetben a programozók [. . .] voltak többségben [. . .] a diákok lettek a legtöbben, ők a felmérés szerint most 200 ezren vannak a fent a Google +-on [. . .] a Google+ felhasználóinak 83 százaléka inaktív.”

„[. . .] az eredmények közzététele után nem sokkal az elemző cég az aktivitásra vonatkozó adatokat visszavonta: mint kiderült, csak a publikus aktivitást monitorozta [. . .]”

Forrás: index.hu

Ki az aktív?

„A Facebook saját statisztikái szerint a 750 millió felhasználó 50 százaléka aktív: akik egy hónapon belül legalább egyszer belépnek a szájtra.”

„A Twitternél: az ottani koncepció szerint az alkalmi felhasználóból akkor válik aktív user, ha több mint 30 accountot követ, és ezeknek legalább a harmada vissza is követi őt. A mikroblogszájt 175 millió regisztrált felhasználója közül mindössze 21 millió [...], miközben 56 millió [...] egyetlen árva feedet sem követ, és 90 millió [...] akiket senki nem követ. A Twitterezőők 80 százaléka életében összesen nem posztolt többet 10 sornál, 40 százalékuk meg soha egyetlen betűt sem.”

Határozzuk meg:

- mintavételi hiba,
- megbízhatósági szint,
- reprezentativitás,
- torzítatlanság,
- kvótamátrix,
- súlymátrix,
- adathiány,
- valószínűségi mintavétel,

„Karácsonyi ajándékvásárlásra a magyarok mindössze négy százaléka venne fel hitelt, és közel ugyanennyien tartják elfogadhatónak az ilyen célú hiteligénylést – állapítja meg a Cofidis és az NRC reprezentatív kutatása.”

„A legtöbben öt-hat ember számára vásárolnak [...] a magyarok átlagosan 29 ezer forintot szánnak karácsonyi ajándéokra, egy megajándékozottra átlagosan 6 ezer forintot költenek [...] a nagyobb összegű beruházásra vállalkozók 28 százaléka tervezi hitelből finanszírozni kiadását [...] a felvenni kívánt hitelek átlagos értéke 99 ezer forint.”

„A kutatást az NRC novemberben a 18-69 éves korosztályon, ezer fő bevonásával, országos reprezentatív mintán végezte.”

„Mennyit szán (ezer Ft) karácsonyi ajándékokra idén?”

15, 40, 70, 30, 50, 10, 150, 100, 30

„Mennyit szán (ezer Ft) karácsonyi ajándékokra idén?”

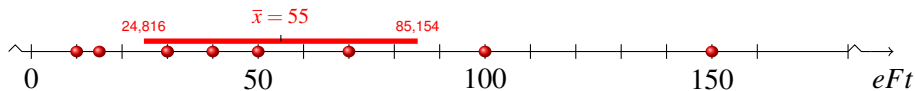
15, 40, 70, 30, 50, 10, 150, 100, 30

- **számtani átlag:** $\bar{x} = \frac{15+40+70+30+50+10+150+100+30}{9} = 55$
- **korrigált empirikus szórás:** $S^* = 45,277$
- **standard/mintavételi hiba:** $SE = \frac{45,277}{\sqrt{9}} = \frac{45,277}{3} = 15,092$
- **konfidencia-intervallum:** $45,277 \pm 2 \cdot 15.092 = [24,816; 85,154]$

„Mennyit szán (ezer Ft) karácsonyi ajándékokra idén?”

15, 40, 70, 30, 50, 10, 150, 100, 30

- **számtani átlag:** $\bar{x} = \frac{15+40+70+30+50+10+150+100+30}{9} = 55$
- **korrigált empirikus szórás:** $S^* = 45,277$
- **standard/mintavételi hiba:** $SE = \frac{45,277}{\sqrt{9}} = \frac{45,277}{3} = 15,092$
- **konfidencia-intervallum:** $45,277 \pm 2 \cdot 15,092 = [24,816; 85,154]$



Mintanagyság meghatározása

Példa

Mekkora mintára van szükségem ahhoz, hogy 5 ezer Ft-os pontossággal meg tudjam állapítani a karácsonyi ajándékokra szentelt megtakarítások nagyságát felnőtt magyar lakosság körében?

Mekkora mintára van szükségem ahhoz, hogy 5 ezer Ft-os pontossággal meg tudjam állapítani a karácsonyi ajándékokra szentelt megtakarítások nagyságát felnőtt magyar lakosság körében?

- 5 ezer Ft pontosság 95 %-os döntési szinten: $SE = 2,5$
- becsült szórás: 45
- $SE = \frac{S^*}{\sqrt{n}}$

Mintanagyság meghatározása

Példa

Mekkora mintára van szükségem ahhoz, hogy 5 ezer Ft-os pontossággal meg tudjam állapítani a karácsonyi ajándékokra szentelt megtakarítások nagyságát felnőtt magyar lakosság körében?

- 5 ezer Ft pontosság 95 %-os döntési szinten: $SE = 2,5$
- becsült szórás: 45
- $SE = \frac{S^*}{\sqrt{n}}$



- $2,5 = \frac{45}{\sqrt{n}}$



- $2,5 \cdot \sqrt{n} = 45$
- $\sqrt{n} = \frac{45}{2,5} = 18$
- $n = 18 \cdot 18 = 324$

Az NRC NetPanel

„Az NRC NetPanel több, mint 140.000 aktív tagot számlál, mérete pedig folyamatosan növekszik. Az új tagok közvetlen regisztráció útján vagy az NRC online és offline (nem panel) kutatásai során kerülnek a panelrendszerbe.”

„Az internet felhasználók közvetlen úton jelentkezhetnek a NetPanel rendszerbe, amennyiben regisztráltják magukat a NetPanel weboldalán (Hozzájárul az internetezők jobb megismeréséhez, véleményével, kritikai megjegyzéseivel és javaslataival segíti az Önért dolgozó cégek munkáját. A kutatásokban való részvétele emellett lehetőséget nyújt számos értékes ajándéktárgy megnyerésére is.)”

Az NRC NetPanel mintái

„A különböző célcsoportok esetében a megfelelő minta kiválasztásához a paneltagok rögzített szociodemográfiai, internet használattal és fogyasztási szokásokkal kapcsolatos adatai nyújtanak segítséget.”

„A felhasznált minta 10.000 fő, a TNS-NRC InterBus (offline) kutatásának adataival súlyozott minta, mely ily módon reprezentatív a 14-69 éves, hetente internetezőkre.”

Forrás: nrc.hu

Az NRC NetPanel lehetséges célcsoportjai

- Magyar internet populációra vonatkozó reprezentatív kutatás
- Speciális internet célcsoport kutatás
- Általános célcsoport vizsgálat abban az esetben, ha a vizsgált csoporton belül viszonylag magas az internet penetráció, és a kutatás eredményeit nem befolyásolja az internet (fogyasztói kutatás, koncepció-, csomagolás-, és reklám tesztek)

„Amennyiben a megkérdezni kívánt célcsoport esetében az internetezők aránya eléri az 50%-ot és az a tény, hogy a válaszadó internetezik, nem befolyásolja a kutatás témáját, érdemes az olcsóbb és gyorsabb online adatfelvételt választani.”

- Értékelje az állítást: „*Az internetes vizsgálatok kiküszöbölik a kérdezői hibát*” (forrás: nrc.hu)
- Mikor érdemes súlyozni? A súlyozás előnyei és hátrányai.
- Az adathiány típusai, az adathiány kezelésének módjai (előnyei, hátrányai)
- A kvótás és rétegzett mintavétel összevetése
- Milyen társadalmi csoportok elérése (nem) alkalmas a telefonos/internetes/postai kérdőív?
- Mikor alkalmaznánk szakértői mintavételt?
- Mi lehet az alapsokaság/mintavételi keret telefonos/internetes/postai kutatás esetén?
- A társadalomtudományi kutatás tervezésének lépései

Köszönöm a figyelmet!

Daróczi Gergely

daroczi.gergely@btk.ppke.hu