

# KIS HIÁNYBÓL NAGY HIBA

Internetes adatgyűjtések az empirikus társadalomkutatásban és ezek problémái a magyar lakosság jellemzői alapján

Fellner Zita, Szeitl Blanka

*Survey Methods Room Budapest, Statisztika Tanszék, ELTE Társadalomtudományi Kar*

Szociológiai tudás és közjó

MSZT 2022. évi vándorgyűlése, Október 14-15, 2022

**SURVEY  
METHODS  
ROOM**   
**BUDAPEST**



# Outline

- 1 Motiváció és háttér
- 2 Kutatási módszer
- 3 Eredmények
  - Online adatfelvételek elterjedtsége
  - Potenciális torzítás az online adatfelvételekben
- 4 Következtetések

# Outline

## 1 Motiváció és háttér

## 2 Kutatási módszer

## 3 Eredmények

- Online adatfelvételek elterjedtsége
- Potenciális torzítás az online adatfelvételekben

## 4 Következtetések

# Motiváció

- A klasszikus személyes adatfelvétel egyre nagyobb **nehézségekbe** ütközik: csökkenő válaszadási arányok, növekvő költségek, nehezen ellenőrizhető kérdezői hálózat.
- A kérdőíves adatfelvételek változó **technológiai környezete**: növekvő internetpenetráció, okostelefon-használat, kérdőívkészítő programok elérhetősége.
- Az online adatfelvételek **növekvő száma**, amelyek **minősége** megkérdőjelezhető.
- **Nemzetközi összehasonlításban** érdemi különbségek az országok között.

# Motiváció

- **Esettanulmány** a hazai adatfelvételekről:
- Hogyan mérhető az online adatfelvételek **elterjedtsége és minősége**?
- Hogyan mérhető az online adatfelvételek potenciális **torzítása**?
- Figyelembe véve a **populációs mintázatokat** (adminisztratív adatok és személyes survey adatok alapján).
- **Miért fontos mindez?** - érzékelés és bizalom

## Háttér

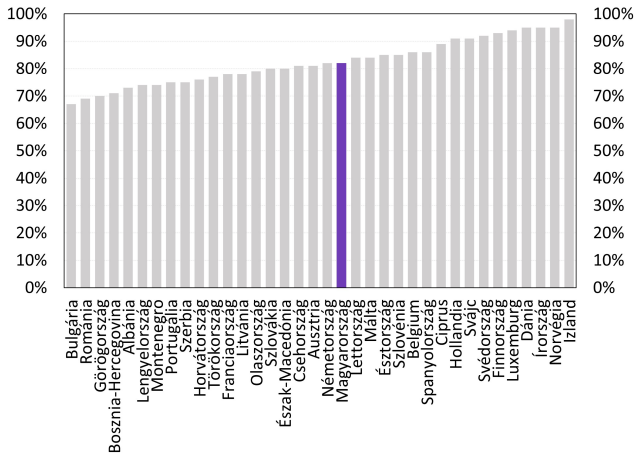


Figure: A mindennapos internethasználók aránya, 2021 (Forrás: Eurostat, %)

## Háttér

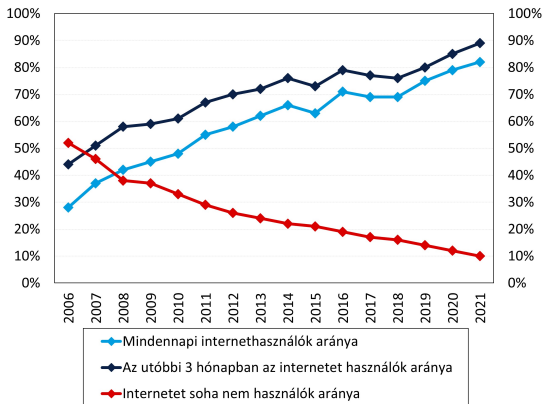


Figure: A hazai népesség az internethasználat gyakorisága szerint 2006-2021 (Forrás: Eurostat, %)

## Háttér

Ismérv	Demográfiai csoport	Arány a teljes népességen belül	Arány az internettel rendelkezőkön belül	Arány a mindennapos internethasználókon belül
Életkor	16-24	13%	15%	16%
	25-34	16%	18%	19%
	35-44	19%	21%	22%
	45-54	20%	21%	21%
	55-64	16%	14%	13%
	65-74	16%	11%	9%
Iskolai végzettség	Alapfokú	19%	16%	15%
	Középfokú	56%	56%	55%
	Felsőfokú	25%	27%	29%
Régió	Budapest	17%	19%	19%
	Pest	14%	14%	14%
	Közép-Dunántúl	11%	11%	11%
	Nyugat-Dunántúl	10%	10%	10%
	Dél-Dunántúl	9%	9%	9%
	Észak-Magyarország	11%	11%	10%
	Észak-Alföld	15%	14%	14%
	Dél-Alföld	13%	12%	12%

Figure: A teljes népesség, az internettel rendelkezők és a mindennapos internethasználók életkor, iskolai végzettség és régió szerint 2021-ben (Forrás: KSH, %)



# Háttér

- Az internetpenetráció emelkedik, így egyre nagyobb csoportok válhatnak **online adatfelvételek célcsoportjává**.
- A különböző időpontokból származó online adatfelvételek eredményei nehezen összehasonlíthatóak, mert **az online elérhető populáció folyamatosan változik**.
- Jelentős különbségek vannak az online elérhető populáció és a teljes populáció között **a főbb demográfiai változók mentén**.

# Outline

- 1 Motiváció és háttér
- 2 **Kutatási módszer**
- 3 Eredmények
  - Online adatfelvételek elterjedtsége
  - Potenciális torzítás az online adatfelvételekben
- 4 Következtetések

# Kutatási módszer

## Az online adatfelvételek elterjedtsége - tartalomelemzés:

- egyszerű véletlen minta online híroldalakon megjelent cikkekben/közlésekből (n=130);
- minta tudományos publikációkból (n=110);
- ezek értékelése előre definiált módszertani kritériumok alapján: mintavétel módja, adatfelvétel módja, rétegzés/súlyozás, interpretáció.

# Kutatási módszer

## Potenciális torzítás az online adatfelvételekben - szimuláció:

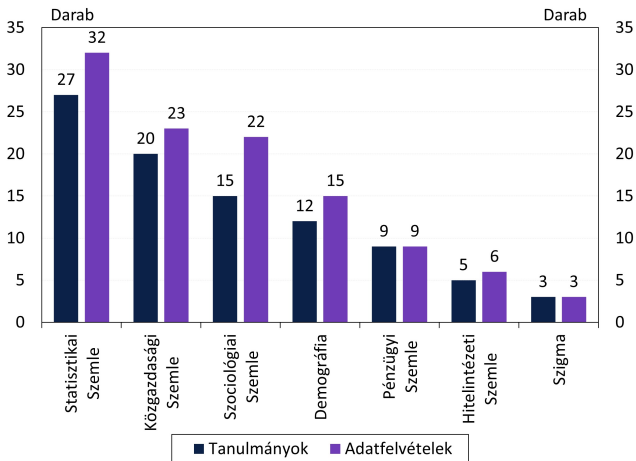
- három különböző populáció: teljes népesség; internet-hozzáféréssel rendelkezők; mindennapos internethasználók;
- populációs paraméterek szimulációja mindhárom populációban;
- becslés készítése ezen populációs paraméterek mellett.

**Kiegészítés:** a populációk összehasonlítás a European Social Survey (ESS) adatain.

# Outline

- 1 Motiváció és háttér
- 2 Kutatási módszer
- 3 Eredmények
  - Online adatfelvételek elterjedtsége
  - Potenciális torzítás az online adatfelvételekben
- 4 Következtetések

# Online adatfelvételek elterjedtsége



**Figure:** A tartalomelemzésbe bevont tudományos publikációk és adatfelvételek száma folyóiratanként

# Online adatfelvételek elterjedtsége

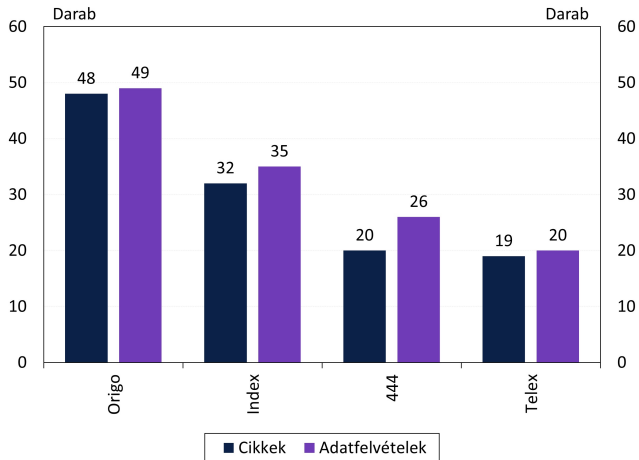


Figure: A tartalomelemzésbe bevont online cikkek és adatfelvételek száma híroldalanként

# Online adatfelvételek elterjedtsége

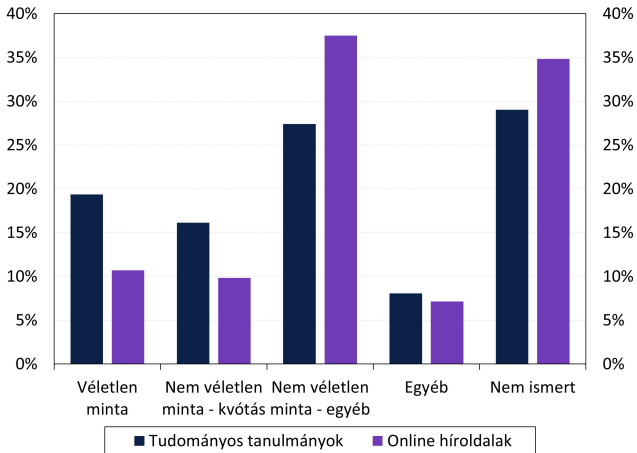


Figure: A vizsgált tanulmányok és online cikkek a mintavétel módja szerint



# Online adatfelvételek elterjedtsége

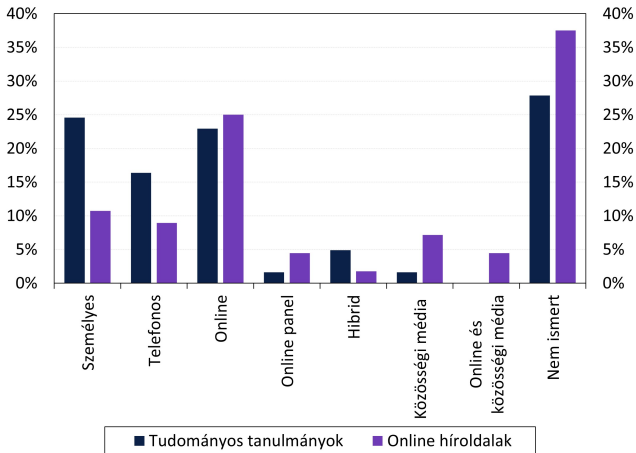


Figure: A vizsgált tanulmányok és online cikkek az adatfelvétel módja szerint

# Online adatfelvételek elterjedtsége

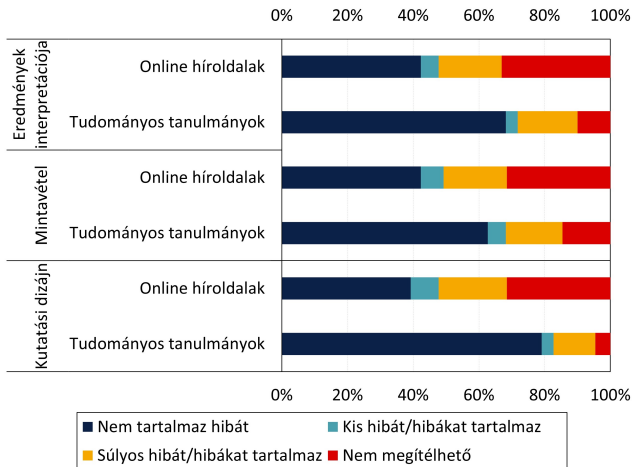


Figure: A vizsgált tanulmányok minőségi kategorizálása

# Potenciális torzítás az online adatfelvételekben

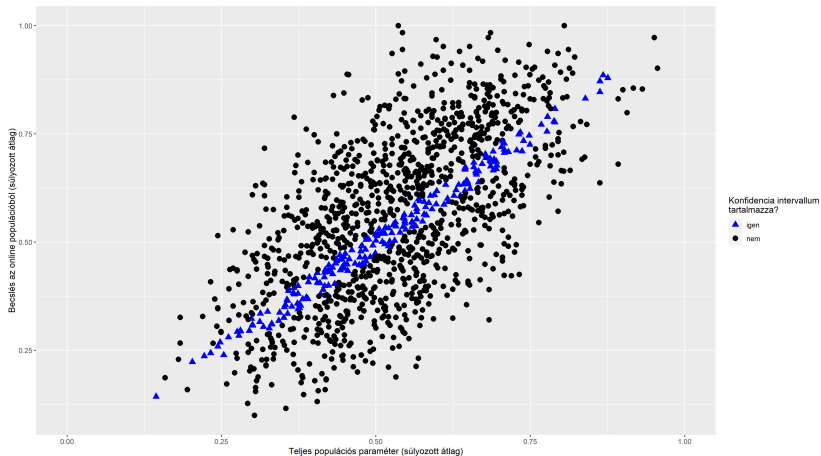
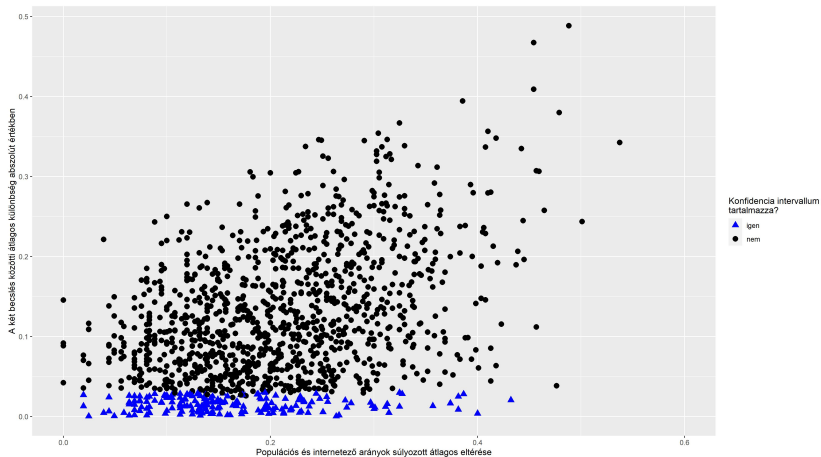


Figure: Az online minta alapján becsült és a populációs megoszlás különbsége

# Potenciális torzítás az online adatfelvételekben



**Figure:** Az online minta alapján becsült és a populációs megoszlás különbsége a teljes populációs paraméter értéke alapján

# Potenciális torzítás az online adatfelvételekben

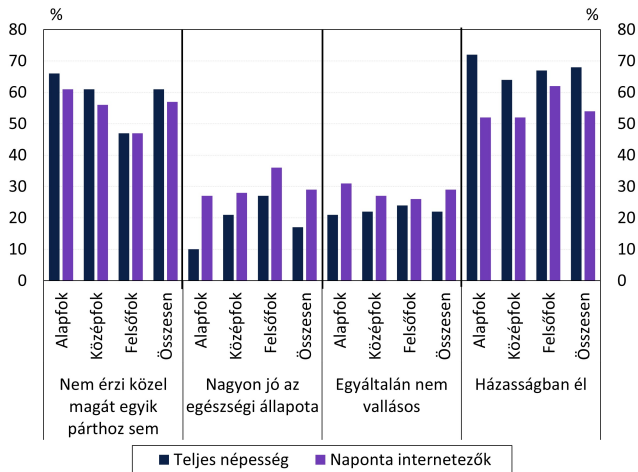


Figure: Becslés valós adatokon (European Social Survey)

# Outline

- 1 Motiváció és háttér
- 2 Kutatási módszer
- 3 Eredmények
  - Online adatfelvételek elterjedtsége
  - Potenciális torzítás az online adatfelvételekben
- 4 Következtetések

# Következtetések

- Az online adatgyűjtések **egyre népszerűbbek**.
- An online minta nem tekinthető a teljes populációból vett véletlen mintának, még magas arányú internethasználat mellett sem.
- Az online adatfelvételekből származó becslések **torzítottak lehetnek**, még abban az esetben is, ha néhány demográfiai ismérv mentén reprezentálják a teljes népséget.
- Az online adatfelvételek **számottevő arányban** fordulnak elő tudományos publikációkban és online cikkekben.
- **Az online adatfelvételekből származtatott eredmények gyakran - helytelenül - véletlen mintaként jelennek meg ezekben a közlésekben.**
- Az eredményeink nem általánosíthatóak más országokra, **a lokális kutatások szükségességére** szeretnénk felhívni a figyelmet.

# Köszönjük a figyelmet!

## Szeitl Blanka

Egyetemi tanársegéd  
Eötvös Loránd Tudományegyetem  
Társadalomtudományi Kar  
Statisztika Tanszék

e-mail: [szeitl.blanka@tatk.elte.hu](mailto:szeitl.blanka@tatk.elte.hu)

## Fellner Zita

Egyetemi tanársegéd  
Eötvös Loránd Tudományegyetem  
Társadalomtudományi Kar  
Statisztika Tanszék

e-mail: [fellner.zita@tatk.elte.hu](mailto:fellner.zita@tatk.elte.hu)