

stathelp.hu

Készítette: Soltész-Várhelyi Klára

1. Populáció és minta

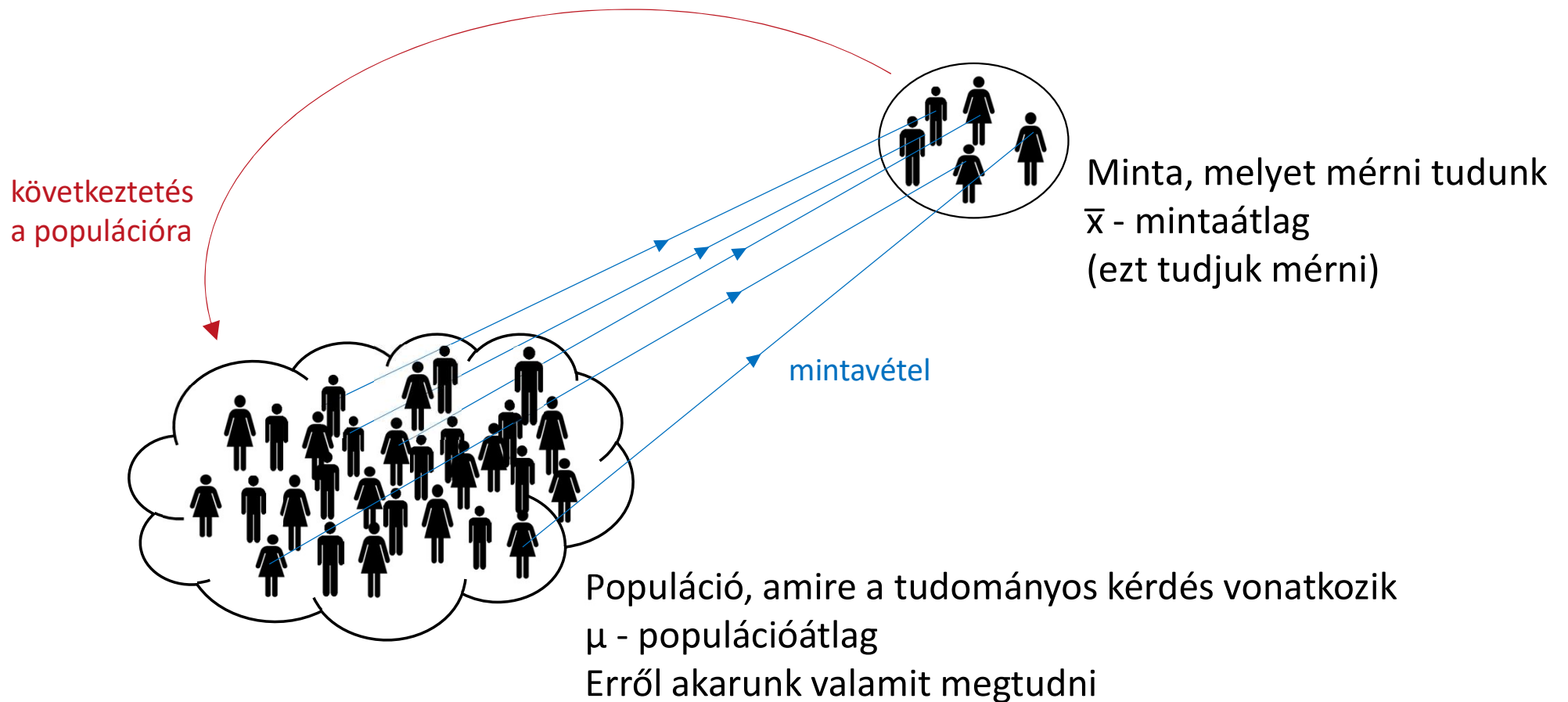
Populáció és minta

Miért csak valószínűségekről beszélhetünk?

“In God we Trust, all others bring data.”

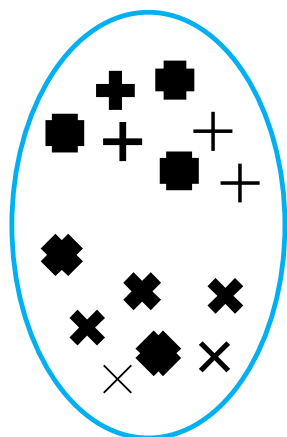
W. Edwards Deming

Populáció és minta

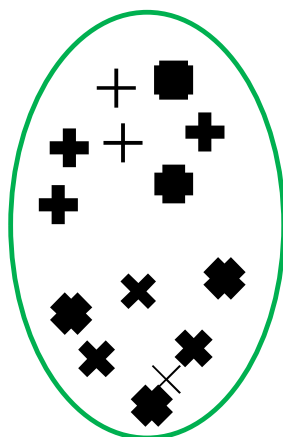


(az ábrát még kiegészítjük néhány diával később, ne ez alapján tanulj meg!)

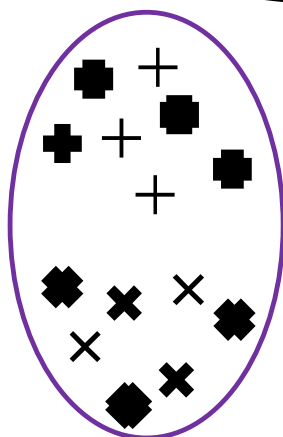
A populációban nincs különbség



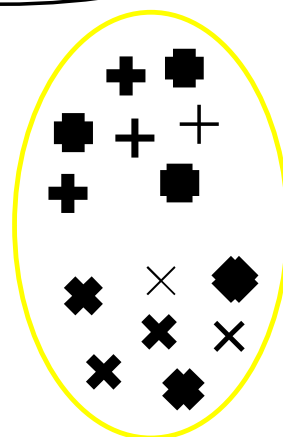
Nem találtunk
különbséget



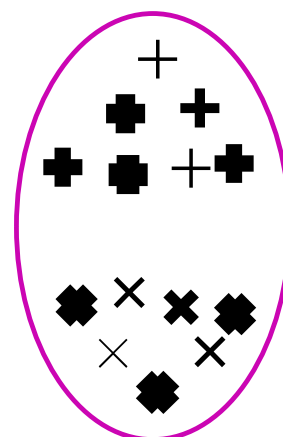
Nem találtunk
különbséget



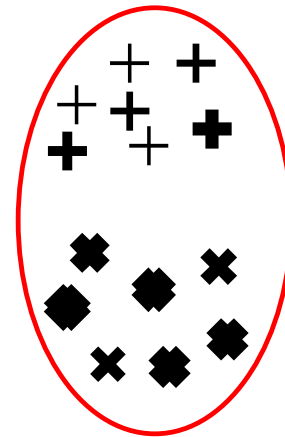
Nem találtunk
különbséget



Nem találtunk
különbséget

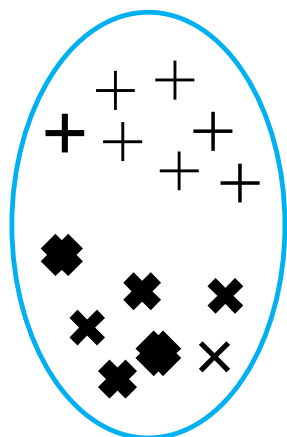
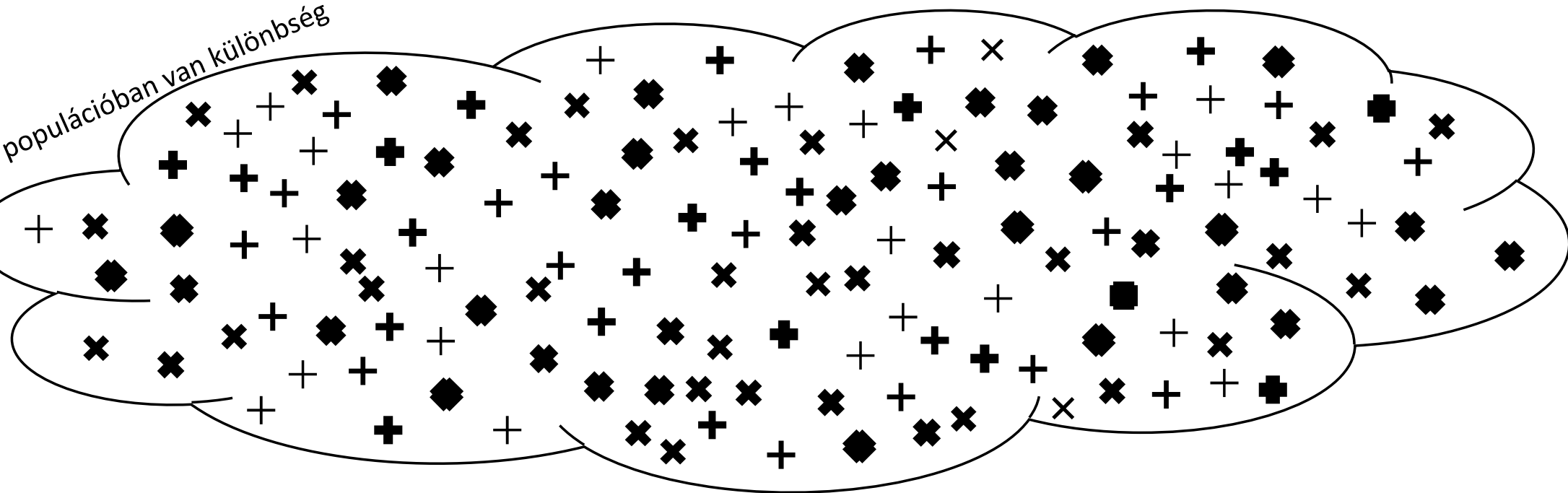


Nem találtunk
különbséget

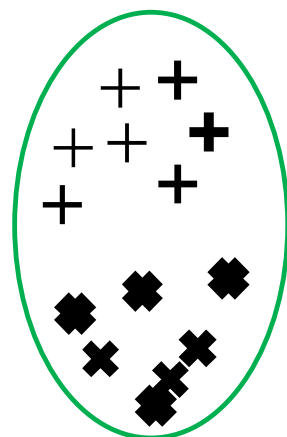


Különbséget találtunk,
és tévedtünk

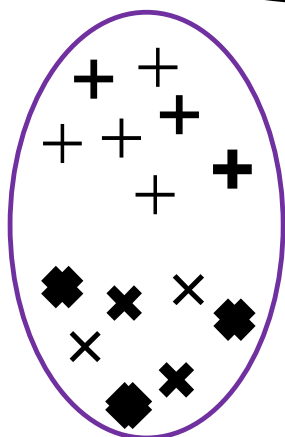
A populációban van különbség



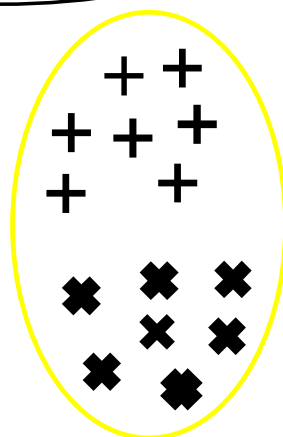
Kimutatjuk a különbséget



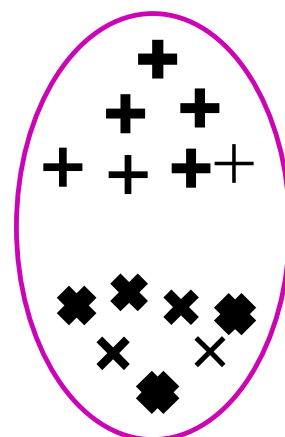
Kimutatjuk a különbséget



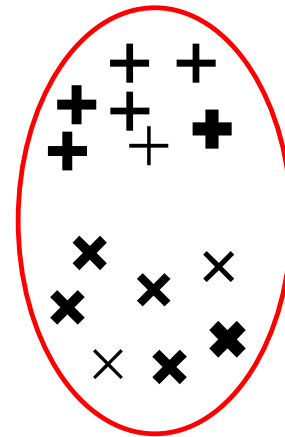
Kimutatjuk a különbséget



Kimutatjuk a különbséget

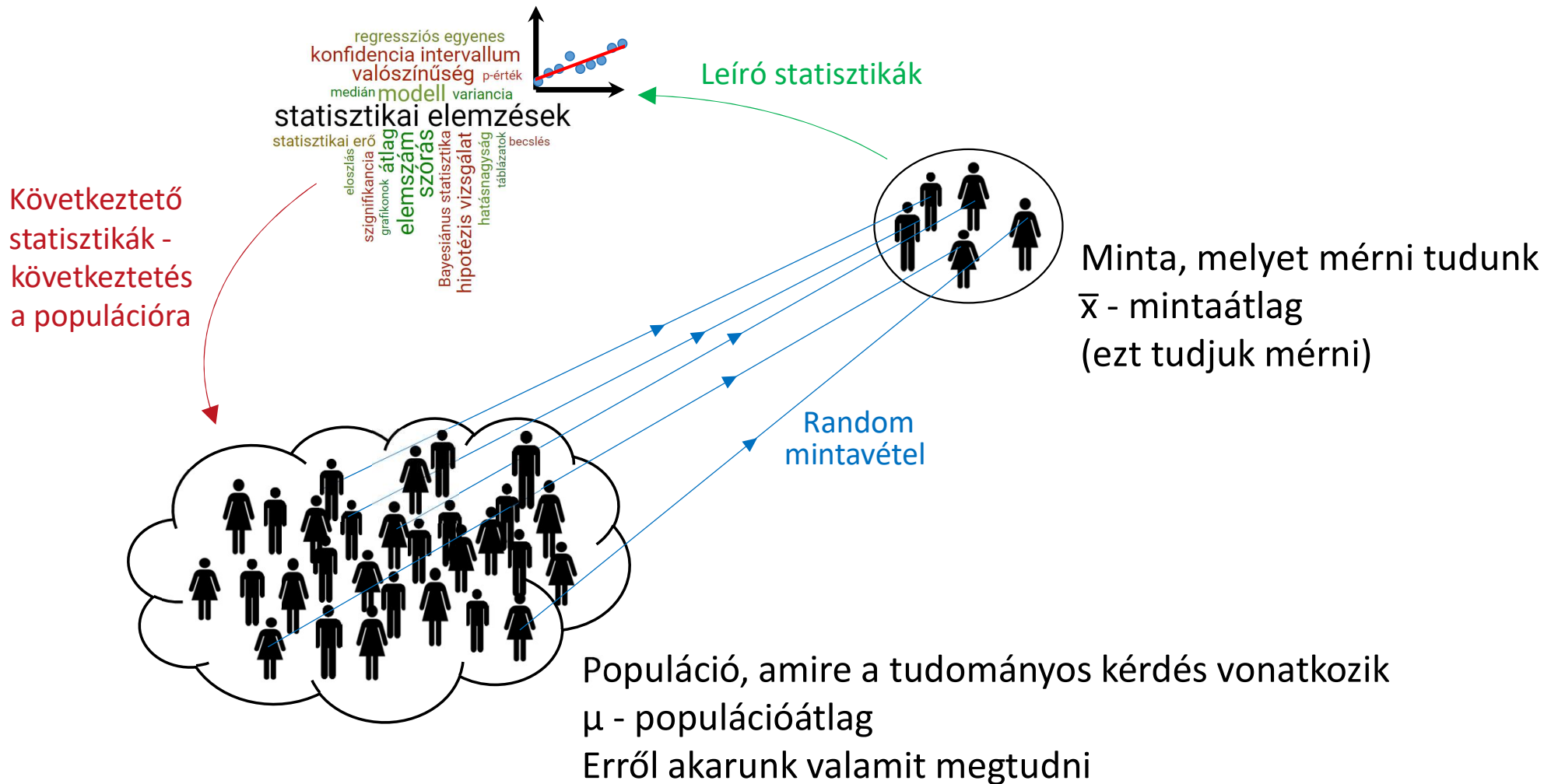


Kimutatjuk a különbséget?



Nem találunk különbséget, és tévedtünk

Populáció és minta



(Annette Kopp-Schneider ábrája alapján)

Populáció és minta

- **Fontos, hogy a minta jól reprezentálja a populációt.**
 - **A) Random mintavételezés**
 - A minta kiválasztásakor törekedni kell arra, hogy **a populáció minden tagjának egyforma esélye legyen a mintába való kerülésre.**
 - A random mintavételezés szinte **soha nem valósul meg tökéletesen**
 - *(brazil nő a mintámban)*
 - Törekedni kell a populáció arányos reprezentálására **a mérést meghatározó tényezők mentén!**
 - Törekedni kell, hogy a mintavételezéssel ne vigyünk a mérésbe **szisztematikus torzítást!**
 - *(férfiak és nők matematika tudása, nők pszichológia, férfiak az informatika szakról)*
 - Korlátozott mintavételezés esetén **át kell gondolni, valójában mi az a populáció, melyre következtethetünk!**
 - Magyarok, tinik, egyetemisták
 - *(férfiak magasabb empátia értéke – pszichológia szakos minta)*

Populáció és minta

- Fontos, hogy a minta jól reprezentálja a populációt.
- B) Statisztikai elemzések
 - *Igaz-e, hogy jelen van a hatás?*
 - Annak megállapítására, hogy a **mintában kapott hatásokat mennyire hihetjük el a populációban is létezőnek**.
 - Mennyire valószínű, hogy a mintában látott hatások (különbségek, összefüggések) a populációban is jelen vannak.
 - Pontosabban fogalmazva éppen megfordítva történik: **p érték annak valószínűsége** lesz, hogy **a mintában látott hatás** csak a **véletlen mintavételezésből** adódik, és a **populációban nincs jelen**.
 - *Igaz-e, hogy nincs jelen a hatás?*
 - Klasszikus hipotézisbecsléssel nem alkalmas annak eldöntésére, hogy a populációban nincs jelen a hatás vagy csak rosszul sikerült a mérés (később részletesen beszélünk róla).
 - Bayesiánus tesztelés
 - Hatásnagyság elemzés

Populáció és minta

- **Mi befolyásolja, mennyire hihetjük el a mintában kapott hatásokat a populációban is létezőnek?**

- Sok minden. Ebből egyelőre hármat emelünk ki:

- **Minta elemszáma**

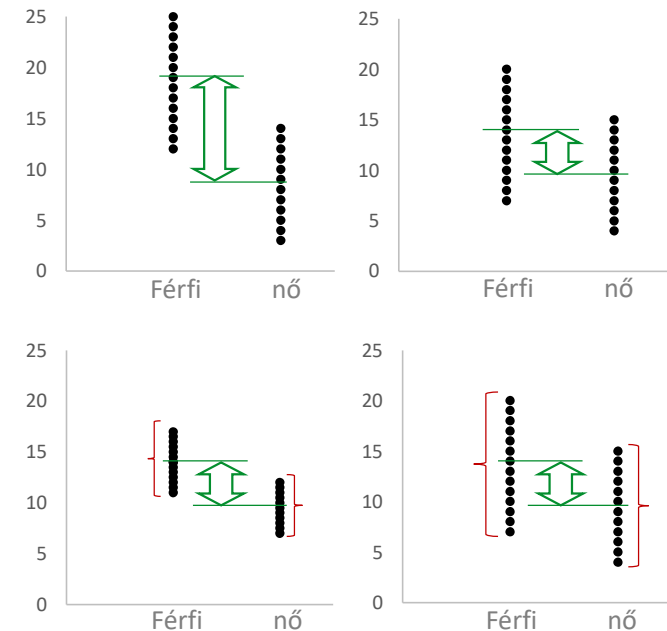
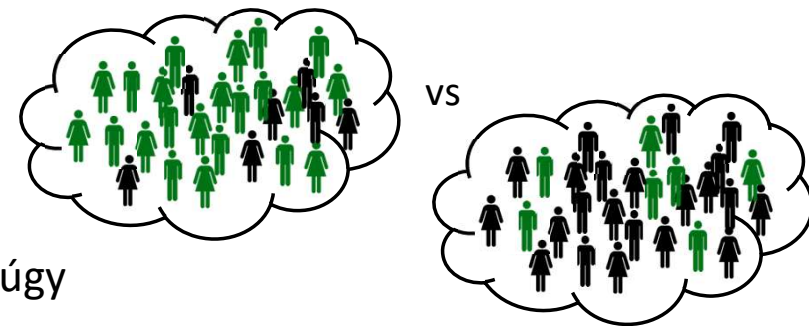
- evidens, minél jobban lefedi a minta a populációt, annál valószínűbb, hogy pontosan reprezentálja azt
- Annál kisebb a valószínűsége, hogy csak véletlenül állt úgy össze a minta, mintha lenne hatás benne

- **A hatás nagysága**

- minél nagyobb a hatás, annál kevésbé valószínű, hogy a véletlen mintavételezés eredménye.

- **A zaj/hiba nagysága**

- minél kisebb a zaj (pl. a csoporton belüli különbségek), annál jobban el tud különülni a két minta,
- annál valószínűbb, hogy a mintában talált hatás nem a véletlen mintavételezés eredménye.



Populáció és minta

- **Nagyobb mintaelemszámmal**

- Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
- Csökken a nem reprezentatív elemek hatása a számolt mutatókra

Gyártó azt írja: 50% figura, 50% összerakós

Ha 3 db tojás veszek, annak valószínűsége,
hogy csak összerakós lesz:
 $0,5 * 0,5 * 0,5 = 0,125$, azaz 12,5%

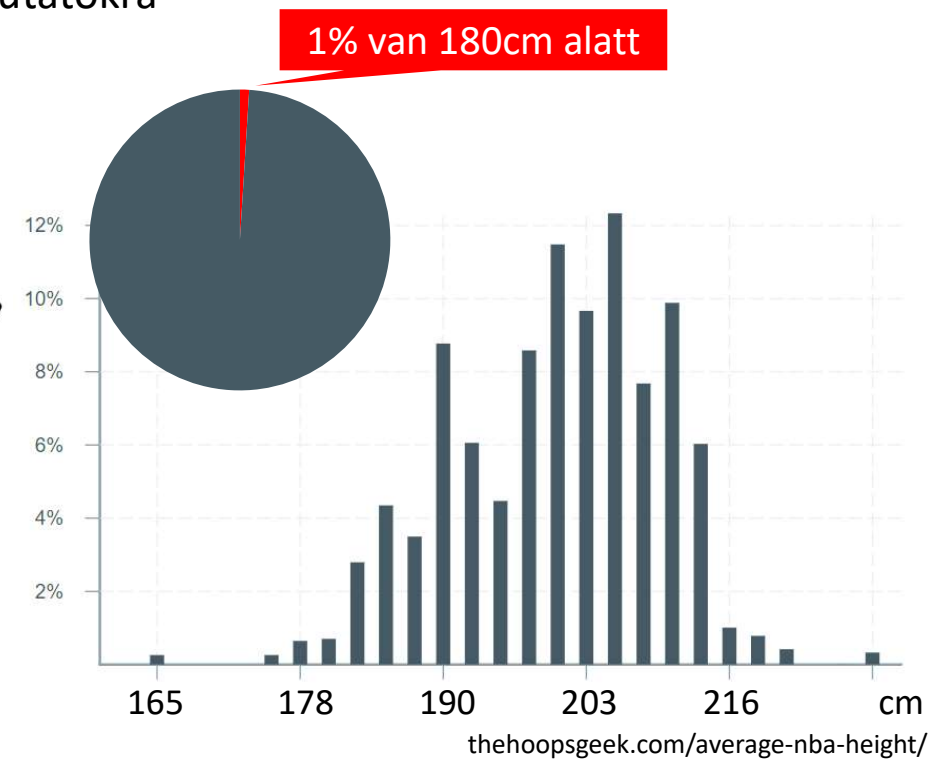
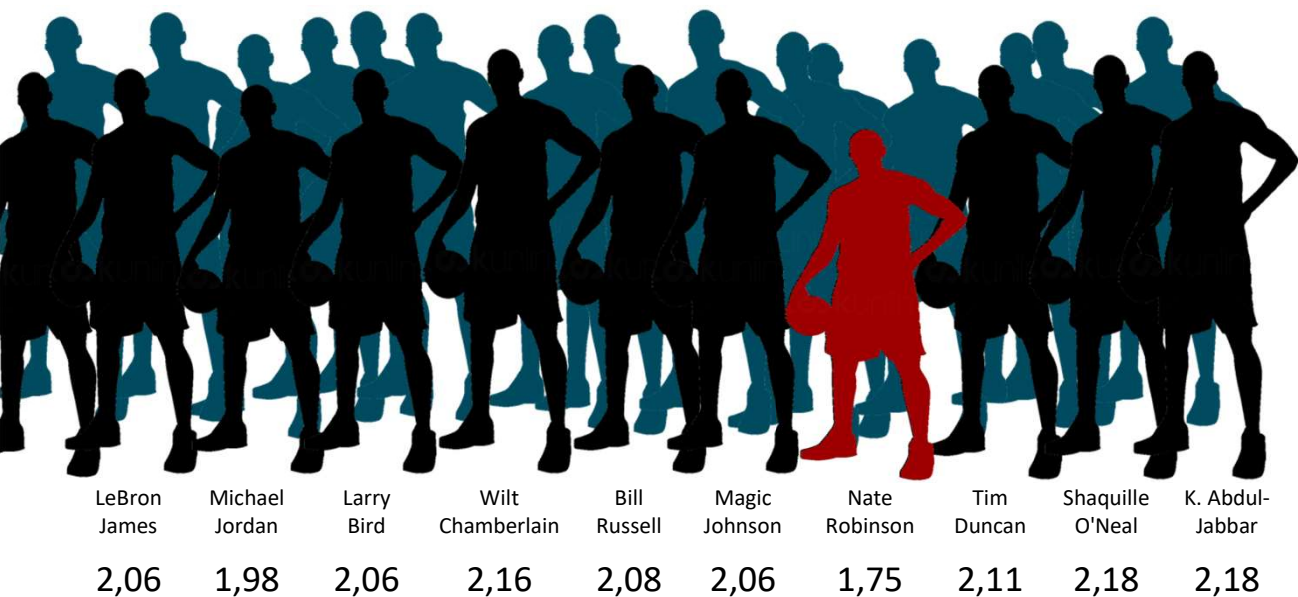


Ha 100 db tojás veszek, annak valószínűsége,
hogy csak összerakós lesz:
 $0,5^{100} = 0,0009765625$, azaz nincs 0,01% sincs



Populáció és minta

- **Nagyobb mintaelemszámmal**
 - Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
 - Csökken a nem reprezentatív elemek hatása a számolt mutatókra

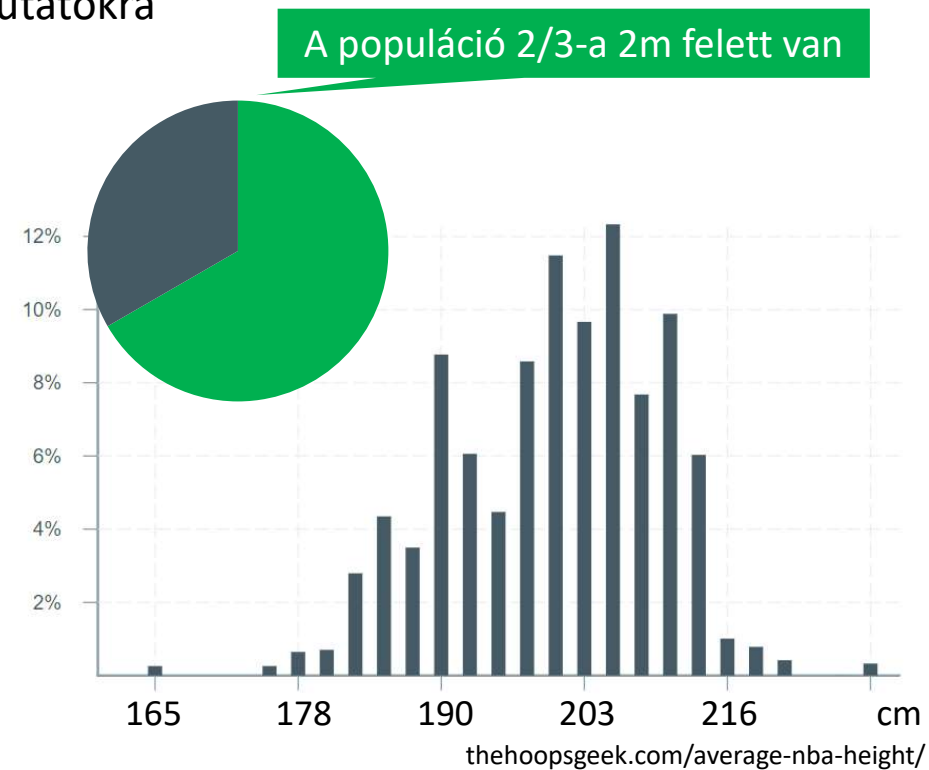


Populáció és minta

- **Nagyobb mintaelemszámmal**
 - Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
 - Csökken a nem reprezentatív elemek hatása a számolt mutatókra



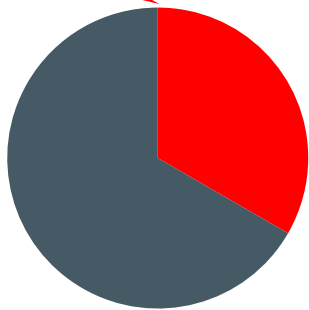
Wilt Chamberlain	Bill Russell	Magic Johnson
2,16	2,08	2,06



Populáció és minta

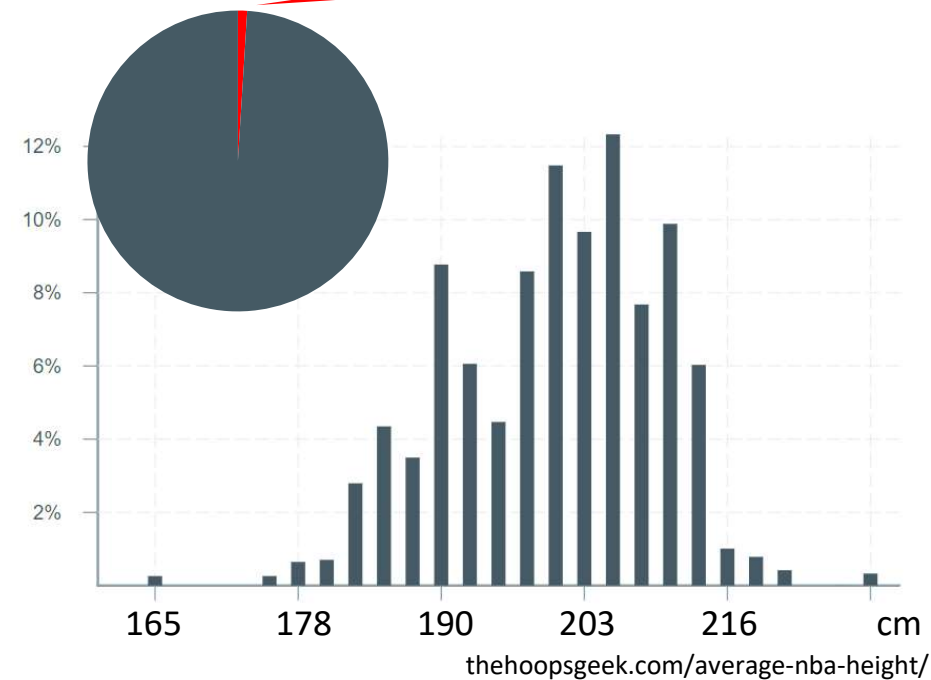
- **Nagyobb mintaelemszámmal**
 - Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
 - Csökken a nem reprezentatív elemek hatása a számolt mutatókra

A minta 33% van 180cm alatt



Magic Johnson	Nate Robinson	Tim Duncan
2,06	1,75	2,11

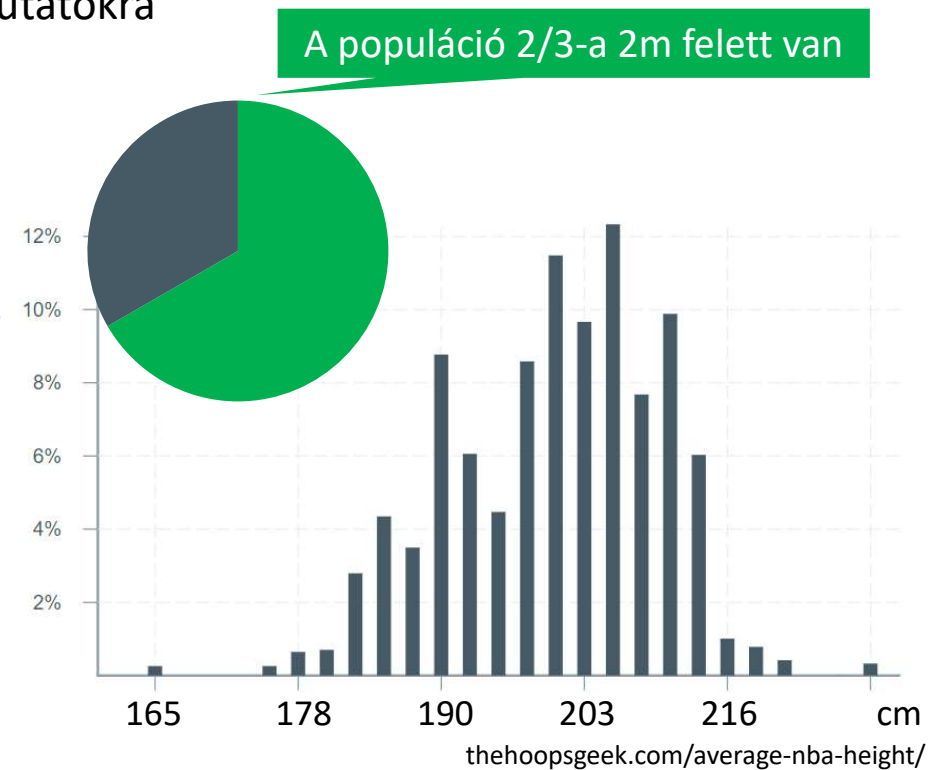
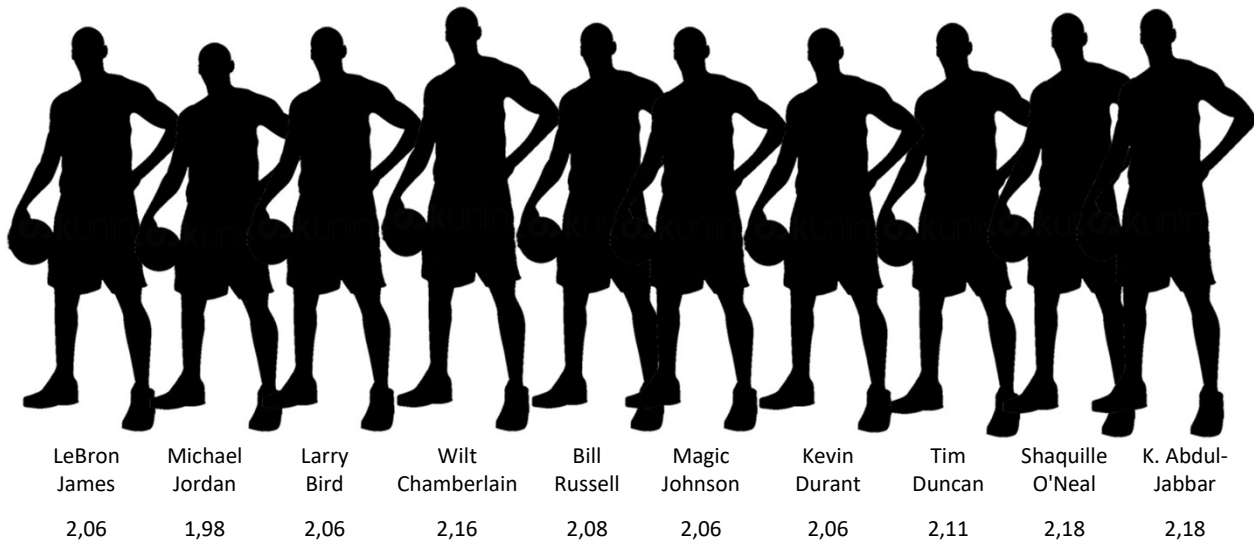
A populáció 1% van 180cm alatt



Populáció és minta

- **Nagyobb mintaelemszámmal**

- Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
- Csökken a nem reprezentatív elemek hatása a számolt mutatókra



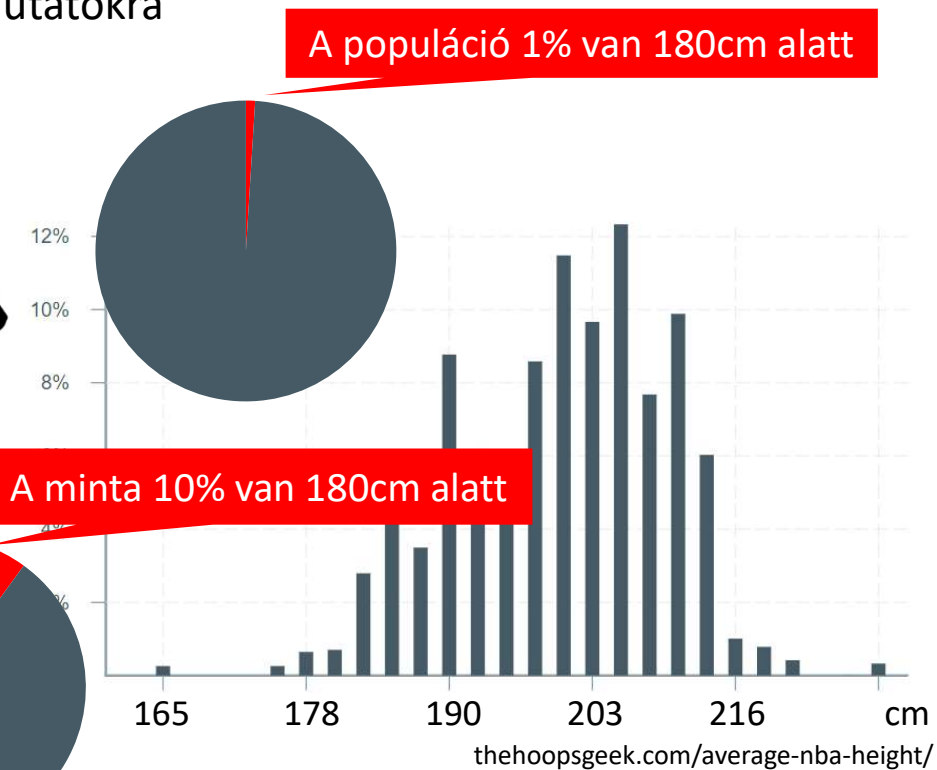
Populáció és minta

- **Nagyobb mintaelemszámmal**

- Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
- Csökken a nem reprezentatív elemek hatása a számolt mutatókra

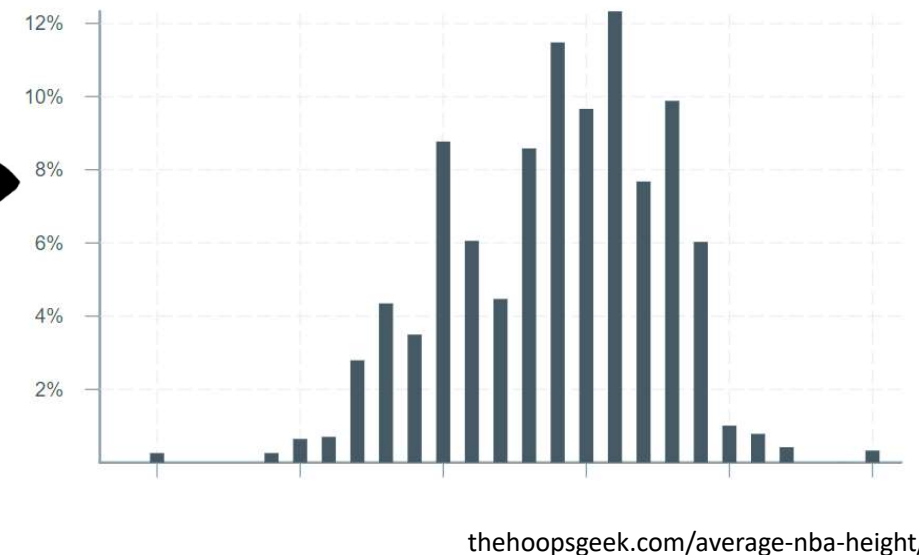
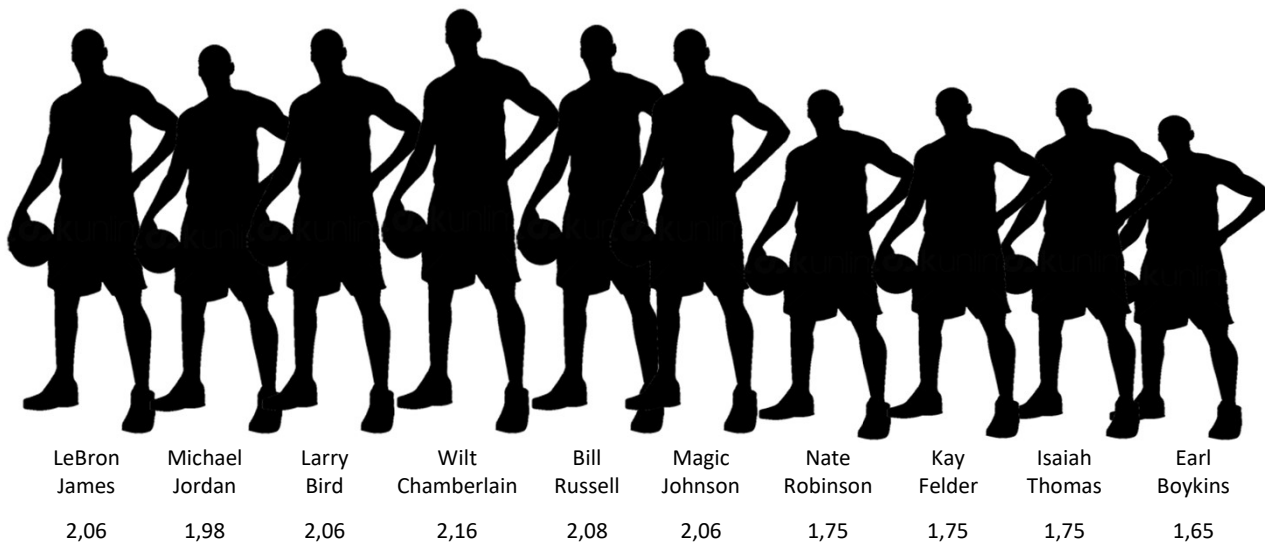


LeBron James	Michael Jordan	Larry Bird	Wilt Chamberlain	Bill Russell	Magic Johnson	Nate Robinson	Tim Duncan	Shaquille O'Neal	Kevin Durant
2,06	1,98	2,06	2,16	2,08	2,06	1,75	2,11	2,16	2,05



Populáció és minta

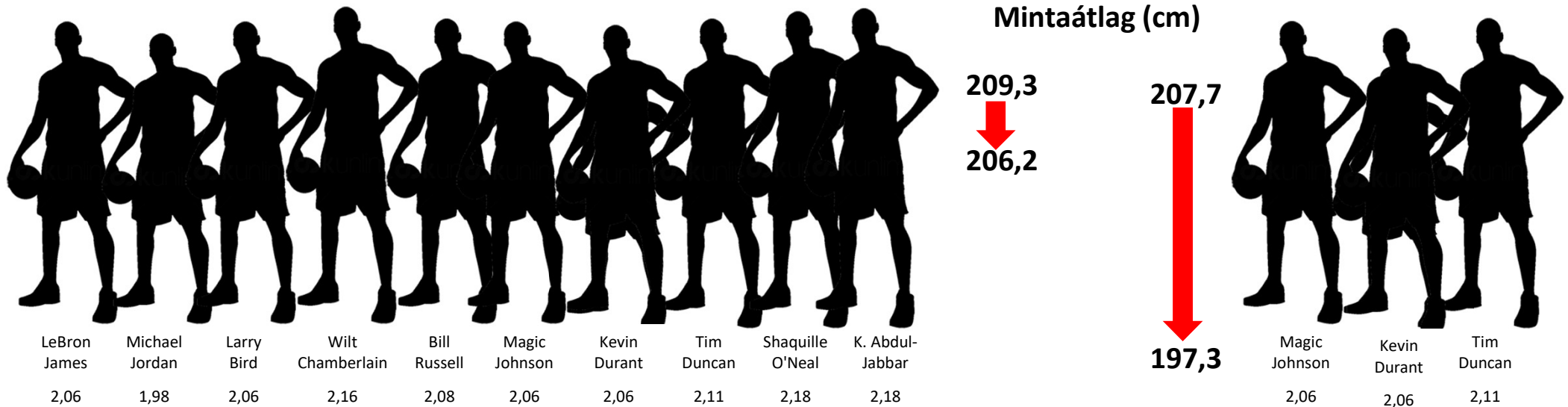
- **Nagyobb mintaelemszámmal**
 - Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
 - Csökken a nem reprezentatív elemek hatása a számolt mutatókra



Populáció és minta

- **Nagyobb mintaelemszámmal**

- Csökken annak valószínűsége, hogy a minta összetétele nem reprezentatív a populációra, azaz a minta összetétele nem hasonlít a populáció összetételéhez
- Csökken a nem reprezentatív elemek hatása a számolt mutatókra



Populáció és minta

- **Mi befolyásolja, mennyire hihetjük el a mintában kapott hatásokat a populációban is létezőnek?**

- Sok minden. Ebből egyelőre hármat emelünk ki:

- **Minta elemszáma**

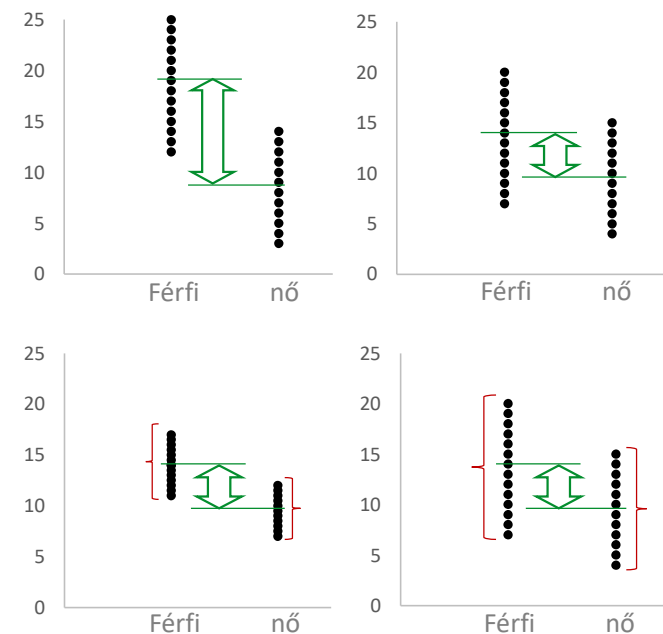
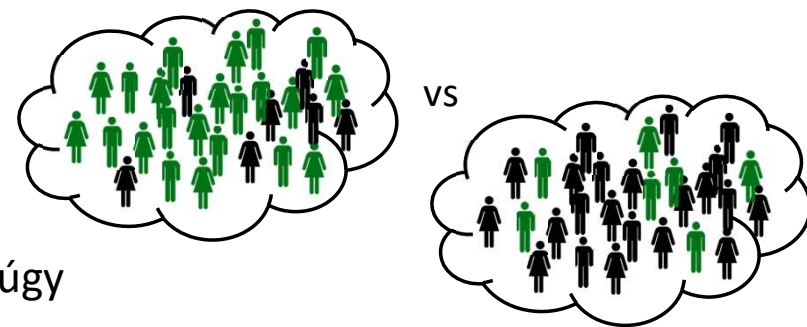
- evidens, minél jobban lefedi a minta a populációt, annál valószínűbb, hogy pontosan reprezentálja azt
- Annál kisebb a valószínűsége, hogy csak véletlenül állt úgy össze a minta, mintha lenne hatás benne

- **A hatás nagysága**

- minél nagyobb a hatás, annál kevésbé valószínű, hogy a véletlen mintavételezés eredménye.

- **A zaj/hiba nagysága**

- minél kisebb a zaj (pl. a csoporton belüli különbségek), annál jobban el tud különülni a két minta,
- annál valószínűbb, hogy a mintában talált hatás nem a véletlen mintavételezés eredménye.



Populáció és minta

- **Populáció**

- A kutatási kérdésben meghatározott **tulajdonságoknak megfelelő összes elemek halmaza**
- Hipotézis: Különbség van férfiak és nők magassága között -> Populáció: A világ összes férfinak és nőjének magassága
- **A populációt általában nem lehet vizsgálni**, mert
 - **túl nagy** ahhoz, hogy minden elemét lemérjük
 - **nem minden eleme elérhető** számunkra

- **Minta**

- **Populációból választott kezelhető méretű rész**
- **A minta reprezentálja a populációt:** a mintában mért hatásokból (különbségek, összefüggések) következtetünk a populációban meglévő hatásokra
- Fontos, hogy a minta jól reprezentálja a populációt

- **Egyed**

- A populáció egyedeiből épül fel
- A populációból kiválasztott egyedek alkotják a mintát, őket mérjük le

