

# Súlyozás

Tóth Gergely  
ELTE-TÁTK, Doktori iskola  
Statisztika tanszék

Notes:

---

---

---

---

---

---

## Mi az a súlyozás?

### • Egyszerű válasz

- Egy eljárás, amely segítségével a mintavételt korigáljuk, az eredmények reprezentativitását garantáljuk.
- Analóg példa: kreditpontokkal súlyozott átlag kiszámítása ... Mi a tétje?

### • Összetettebb válasz

- Egy olyan matematikai transzformáció, amely segítségével – az alapsokaság és a minta azonos paramétereinek ismeretében – korrekciót lehet végezni a mintavételből történő becslés javítása érdekében.

Notes:

---

---

---

---

---

## Egy példa a súlyozásra – I.

Alapsokaság eloszlása								
Nem			Nem					
	Férfi	Nő	TOTAL		Férfi	Nő	TOTAL	
kor	fiatal	16 667	16 667	<b>33 334</b>	fiatal	17%	17%	<b>33%</b>
	középkorú	16 667	16 667	<b>33 334</b>	középkorú	17%	17%	<b>33%</b>
	idős	16 667	16 667	<b>33 334</b>	idős	17%	17%	<b>33%</b>
	<b>TOTAL</b>	<b>50 001</b>	<b>50 001</b>	<b>100 002</b>	<b>TOTAL</b>	<b>50%</b>	<b>50%</b>	<b>100%</b>

Minta eloszlása								
Nem			Nem					
	Férfi	Nő	TOTAL		Férfi	Nő	TOTAL	
kor	fiatal	180	100	<b>280</b>	fiatal	18%	10%	<b>28%</b>
	középkorú	212	80	<b>292</b>	középkorú	21%	8%	<b>29%</b>
	idős	228	200	<b>428</b>	idős	23%	20%	<b>43%</b>
	<b>TOTAL</b>	<b>620</b>	<b>380</b>	<b>1 000</b>	<b>TOTAL</b>	<b>62%</b>	<b>38%</b>	<b>100%</b>

Notes:

---



---



---



---



---

## Egy példa a súlyozásra – II.

Eltérés				Eltérés százalékosan				
Nem				Nem				
	Férfi	Nő	TOTAL		Férfi	Nő	TOTAL	
kor	fiatal	-1%	7%	5%	fiatal	93%	167%	119%
	középkorú	-5%	9%	4%	középkorú	79%	208%	114%
	idős	-6%	-3%	-9%	idős	73%	83%	78%
TOTAL	-12%	12%	0%	TOTAL	81%	132%	100%	

Súlyok				Súlyozott értékek				
Nem				Nem				
	Férfi	Nő	TOTAL		Férfi	Nő	TOTAL	
kor	fiatal	0.93	1.67	1.19	fiatal	166.7	166.7	333.3
	középkorú	0.79	2.08	1.14	középkorú	166.7	166.7	333.3
	idős	0.73	0.83	0.78	idős	166.7	166.7	333.3
TOTAL	0.81	1.32	1.00	TOTAL	500.0	500.0	1 000.0	

Notes:

---



---



---



---



---

## Egy példa a súlyozásra – III.

Valódi jellemzők: hány % fogyaszt X terméket				Valódi jellemzők: hány Fő fogyaszt X terméket			
Nem				Nem			
				Férfi	Nő		
kor	fiatal	100%	100%	16667	16667		
	középkorú	50%	50%	8334	8334		
	idős	0%	0%	0	0		
				<b>50%</b>		<b>50001</b>	
<b>Termékhasználók aránya= 50 001/100 002 = 50%</b>							

Minta jellemzők: hány % fogyaszt X terméket				Minta jellemzők: hány Fő fogyaszt X terméket			
Nem				Nem			
				Férfi	Nő		
kor	fiatal	100%	100%	180	100		
	középkorú	50%	50%	106	40		
	idős	0%	0%	0	0		
						<b>426</b>	
<b>Becslés= 426/1000 = 42,6%</b>							

Notes:

---



---



---



---



---

## Egy példa a súlyozásra – IV.

Súlyok				Minta jellemzők: hány Fő fogyaszt X terméket				
		Nem				Nem		
		Férfi	Nő	TOTAL			Férfi	Nő
kor	fiatal	0.93	1.67	<b>1.19</b>	kor	fiatal	180	100
	középkorú	0.79	2.08	<b>1.14</b>		középkorú	106	40
	idős	0.73	0.83	<b>0.78</b>		idős	0	0
TOTAL		<b>0.81</b>	<b>1.32</b>	<b>1.00</b>			<b>426</b>	

Súlyozás				Súlyozott esetszámok			
		Nem				Nem	
		Férfi	Nő			Férfi	Nő
kor	fiatal	$0.93 \cdot 180 = 166.6$	$1.67 \cdot 100 = 167$	kor	fiatal	167	167
	középkorú	$0.79 \cdot 106 = 83.3$	$2.08 \cdot 40 = 83.3$		középkorú	83	83
	idős	0.00	0.00		idős	0	0
						<b>500</b>	

$$\text{Beclés} = 500/1000 = 50\%$$

Notes:

---



---



---



---



---

## Milyen mértékben lehet súlyozni?




- A súlyozás nem „csodaszer”

Ha az eltérés oka:

- hibás az alkalmazott a mintavétel
- nagy a válassz megtagadás
- egyéb adathiány

akkor valójában a hibák elfedését szolgálja!

- Súlyok értékei

- $=1$   tökéletes megfelelés
- $<1$   adott csoport felülreprezentált
- $>1$   adott csoport alulreprezentált

Notes:

---

---

---

---

---

---

## Milyen hibákat lehet korrigálni?

### Valószínűségi mintákat:

1. Teljes véletlen minta 
2. Rétegzett minta 

### Reprezentatívvá tenni bizonyos szempontból, ha

- I. Eltérés mértéke nem túl nagy (súlyok eloszlása)
- II. Véletlen mintavételnek betudható hibák korrigálására

### De nem alkalmas/Nem szabad használni:

- A minta valószínűtlenül nagy hibáinak „eltüntetésére” 
  - Ennek okai sokfélék lehetnek: rossz mintavétel, véletlen, válassz megtagadás, stb.

### További fontos megjegyzés:

- A súlyozás NEM garantálja a becslés jóságát

Notes:

---

---

---

---

---

---



KÖSZÖNÖM A FIGYELMET!

Tóth Gergely  
toth.gergo@gmail.com

Notes:

---

---

---

---

---