

# Súlyozás

Tóth Gergely  
ELTE-TÁTK, Doktori iskola  
Statisztika tanszék

# Mi az a súlyozás?

## • Egyszerű válasz

- Egy eljárás, amely segítségével a mintavételt korrigáljuk, 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.

# Egy példa a súlyozásra – I.

## Alapsokaság eloszlása

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

		Nem		
		Férfi	Nő	TOTAL
kor	fiatal	17%	17%	<b>33%</b>
	középkorú	17%	17%	<b>33%</b>
	idős	17%	17%	<b>33%</b>
	TOTAL	<b>50%</b>	<b>50%</b>	<b>100%</b>

## Minta eloszlása

		Nem		
		Férfi	Nő	TOTAL
kor	fiatal	180	100	<b>280</b>
	középkorú	212	80	<b>292</b>
	idős	228	200	<b>428</b>
	TOTAL	<b>620</b>	<b>380</b>	<b>1 000</b>

		Nem		
		Férfi	Nő	TOTAL
kor	fiatal	18%	10%	<b>28%</b>
	középkorú	21%	8%	<b>29%</b>
	idős	23%	20%	<b>43%</b>
	TOTAL	<b>62%</b>	<b>38%</b>	<b>100%</b>

# Egy példa a súlyozásra – II.

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

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

# Egy példa a súlyozásra – III.

Valódi jellemzők: hány % fogyaszt X terméket

		Nem	
		Férfi	Nő
kor	fiatal	100%	100%
	középkorú	50%	50%
	idős	0%	0%

50%

Valódi jellemzők: hány Fő fogyaszt X terméket

		Nem	
		Férfi	Nő
kor	fiatal	16667	16667
	középkorú	8334	8334
	idős	0	0

50001

**Termékhasználók aránya= 50 001/100 002 = 50%**

Minta jellemzők: hány % fogyaszt X terméket

		Nem	
		Férfi	Nő
kor	fiatal	100%	100%
	középkorú	50%	50%
	idős	0%	0%

Minta jellemzők: hány Fő fogyaszt X terméket

		Nem	
		Férfi	Nő
kor	fiatal	180	100
	középkorú	106	40
	idős	0	0

426

**Beclslés= 426/1000 = 42,6%**

# Egy példa a súlyozásra – IV.

Súlyok				Minta jellemzők: hány Fő fogyaszt X terméket				
		Nem		TOTAL			Nem	
		Férfi	Nő		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		TOTAL			Nem	
		Férfi	Nő		Férfi	Nő		
kor	fiatal	$0.93 \cdot 180 = 166$	$1.67 \cdot 100 = 166$		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>	

**Becslés =  $500/1000 = 50\%$**

# Milyen mértékben lehet súlyozni?




- **A súlyozás nem „csodaszer”**

**Ha az eltérés oka:**

- **hibás az alkalmazott a mintavétel**
- **nagy a válasszmettagadás**
- **egyéb adathiány**

**akkor valójában a hibák elfedését szolgálja!**

- **Súlyok értékei**

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

# 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:



KÖSZÖNÖM A FIGYELMET!

Tóth Gergely  
toth.gergo@gmail.com