

# Kalibrálás nem-determinisztikus sarokszámok esetén

Mihályffy László, KSH (ny.)

XII. Gazdaságmodellezési Szakértői Konferencia  
Budapest, 2012. június 4.

## Kalibrálás – a feladat

Valószínűségi minta

$$s = \{1, 2, \dots, n\}$$

Célváltozók:  $y_j, z_j$ , stb.

Segédváltozók:  $x_{1j}, x_{2j}, \dots, x_{mj}$

„Design” súlyok:  $w_1^0, w_2^0, \dots, w_n^0$

## Kalibrálás – a feladat

Valószínűségi minta

$$s = \{1, 2, \dots, n\}$$

Célváltozók:  $y_j, z_j$ , stb.

Segédváltozók:  $x_{1j}, x_{2j}, \dots, x_{mj}$

„Design” súlyok:  $w_1^0, w_2^0, \dots, w_n^0$

|  
*Kalibrálás*

|  
|-----> Valószínűségi minta  
|  
|----->  $s = \{1, 2, \dots, n\}$   
|-----> Célváltozók:  $y_j, z_j$ , stb.  
|-----> Segédváltozók:  $x_{1j}, x_{2j}, \dots, x_{mj}$   
|-----> Kalibrált súlyok:  $w_1, w_2, \dots, w_n$

## A kalibrálás célja

Segédváltozók értékösszege

- a sokaságban:  $X_1, X_2, \dots, X_m$  ← ismert  
(kontrollok v. sarokszámok,  $m < n$ )
- becslés a design súlyokkal:  $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_m$
- becslés a kalibrált ’’ :  $\hat{X}_1^{\text{kal}}, \hat{X}_2^{\text{kal}}, \dots, \hat{X}_m^{\text{kal}}$

**Cél** (egyben követelmény):

1.  $\hat{X}_1^{\text{kal}} = X_1, \hat{X}_2^{\text{kal}} = X_2, \dots, \hat{X}_m^{\text{kal}} = X_m$
2. A design- és a kalibrált súlyok legyenek „közel” egymáshoz

## **A legegyszerűbb kalibrálási feladat:**

Minimalizáljuk a

$$\sum_{j=1}^n (w_j - w_j^0)^2 / w_j^0$$

**távolságfüggvényt** (2 .követelmény)

a 
$$\sum_{j=1}^n x_{ij} w_j = X_i, \quad i = 1, 2, \dots, m$$

**kalibrálási feltételek** mellett (1. követelmény)

## A legegyszerűbb kalibrálási feladat tulajdonságai

- $w_j$  zárt alakban felírható
- egy  $Y$  értékösszeg (célváltozó) kal. becslése  
**regressziós becslés:**

$$\hat{Y}^{\text{kal}} = \hat{Y} + \sum_{i=1}^m b_i (X_i - \hat{X}_i), \quad b_1, b_2, \dots, b_m$$

regressziós együtthatók,  $b_i$  helyett  $\hat{b}_i$

- *segédváltozók esetén:*

$$\hat{X}_i^{\text{kal}} = X_i$$

A regressziós becslés szórásnégyzetének becslése:

$$\text{var}(\hat{Y}^{\text{kal}}) = \text{var}(\hat{Y} - \sum_{i=1}^m \hat{b}_i \hat{X}_i) =$$

$$\text{var}\left(\sum_{j=1}^n w_j^0 (y_j - \sum_{i=1}^m \hat{b}_i x_{i,j})\right) \quad (\text{V})$$

a „var” operátor argumentuma a reziduumoknak a design súlyokkal súlyozott összege

## A kalibrálás célja

„**Filozófia**”: szoros kapcsolat a **cél-** és a **segédváltozók** között → kalibrált becslések mintavételi hibája kisebb

Alkalmazások:

(1) mintavételi hiba csökkentése, (2) meghiusulás kompenzálása

Megjegyzés: lakossági / háztartásstatisztikai felvételeknél a segédváltozók között dominálnak a demográfiai változók



## **Kalibrálás – mit nem szabad?**

- (1) célváltozó  $\neq$  segédváltozó
- (2) a sarokszámok *sokaságbeli* jellemzők

## Módosított kalibrálás

~~(2) a sarokszámok sokaságbeli jellemzők~~

Legyenek a sarokszámok:  $X_1, \dots, X_{m-1}, \check{X}_m^{\text{kal}}$

Speciális eset:

(i)  $\check{X}_m^{\text{kal}}$  egy olyan  $s_1$  mintából származik,  
melyre

$$s \subset s_1 \subseteq U$$

ahol  $U = \{1, 2, \dots, N\}$  a sokaság ( $n \ll N$ );

(ii)  $x_{1j}, x_{2j}, \dots, x_{mj} \in U \quad \forall j$

## Értékösszeg becslése módosított kalibrálási eljárás esetén

A sarokszámok:  $X_1, \dots, X_{m-1}, \check{X}_m^{\text{kal}}$

$$\hat{Y}^{\text{kal}} = \hat{Y} + \sum_{i=1}^{m-1} \hat{b}_i (X_i - \hat{X}_i) + \hat{b}_m (\check{X}_m^{\text{kal}} - \hat{X}_m),$$

$$\check{X}_m^{\text{kal}} = \check{X}_m + \sum_{k=1}^{m-1} \hat{b}'_k (X_k - \check{X}_k)$$

## Becsült variancia módosított kalibrálás esetén

$$\begin{aligned}\text{var } \hat{Y}^{\text{kal}} &= \\ &\text{var}(\hat{Y} - \sum_{i=1}^m b_i \hat{X}_i + b_m (\bar{X}_m - \sum_{k=1}^{m-1} b'_k \bar{X}_k) + D)) \\ &= \text{var}(Z_o + Z + D) = \text{var}(Z_o + Z),\end{aligned}$$

ahol

$$D = \sum_{i=1}^{m-1} \hat{b}_i X_i + \hat{b}_m \sum_{k=1}^{m-1} \hat{b}'_k X_k \text{ mintától független tagok összege}$$

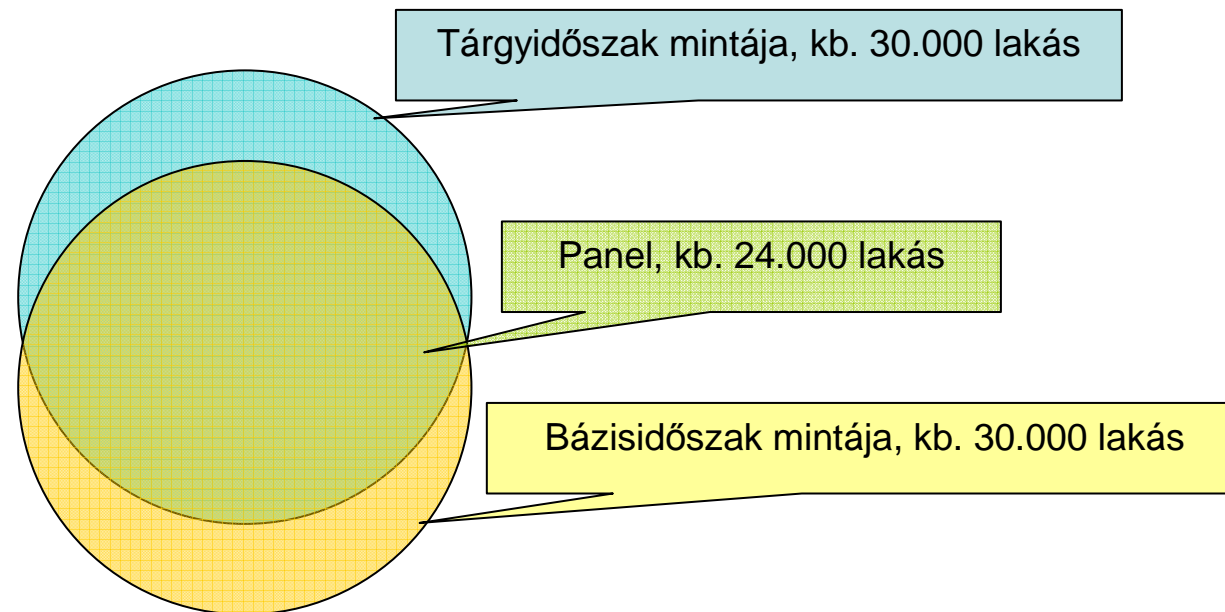
$Z_o$  és  $Z$ : reziduumoknak design súlyokkal súlyozott összegei

Alkalmazás:

**F: foglalkoztatott, M: munkanélküli, I: inaktív,**  
**MF: áramlás, munkanélküliből foglalkoztatott,**  
**FM: foglalkoztatottból munkanélküli —**  
*sokaságbeli adatok*

				<i>Tárgyidőszak(t<sub>2</sub>)</i>	
				Új	Állomány
<i>Bázis- időszak(t<sub>1</sub>)</i>	FF	MF	IF	F <sub>be</sub>	F
	FM	MM	IM	M <sub>be</sub>	M
	FI	MI	II	I <sub>be</sub>	I
Kilépő	F <sub>ki</sub>	M <sub>ki</sub>	I <sub>ki</sub>		
Állomány	F <sub>0</sub>	M <sub>0</sub>	I <sub>0</sub>		

## Egymás utáni negyedévek mintái és a panel a munkaerő-felmérésben (MEF)



## Állapotváltozások és állományok becslése a munkaerő-felmérés (MEF) alapján

				<i>Tárgyidőszak(t<sub>2</sub>)</i>	
				Új	Állomány
<i>Bázis- időszak(t<sub>1</sub>)</i>	$\hat{F}F$	$\hat{M}F$	$\hat{I}F$	$\hat{F}_{be}$	$\check{F}$
	$\hat{F}M$	$\hat{M}M$	$\hat{I}M$	$\hat{M}_{be}$	$\check{M}$
	$\hat{F}I$	$\hat{M}I$	$\hat{I}I$	$\hat{I}_{be}$	$\check{I}$
Kilépő	$\hat{F}_{ki}$	$\hat{M}_{ki}$	$\hat{I}_{ki}$		
Állomány	$\check{F}_0$	$\check{M}_0$	$\check{I}_0$		

## Állapotváltozások és állományok becslése a munkaerő-felmérés (MEF) alapján (*folyt.*)

Változók funkciója

Minta	Demográfiai változók	$\check{F}_0, \check{M}_0, \check{I}_0$	$\check{F}, \check{M}, \check{I}$
Bázisidőszak	segédváltozó	célváltozó	—
Tárgyidőszak	segédváltozó	—	célváltozó
Panel	segédváltozó	segédváltozó	segédváltozó

Célváltozók a panelben: állapotváltozások  
( $\hat{F}M, \hat{M}F, \hat{I}F$ , stb.)



*Köszönöm a figyelmet!*