

Problema liniaritatii in econometrie

Deseori modelele liniare simple nu sunt satisfacatoare motiv pentru care in econometrie pot fi utilizate frecvent modele alternative fie mai complexe, cu mai multe variabile, fie de o alta forma, modele exponentiale, parabolice, inverse.

Cu toate acestea exista o categorie destul de mare de modele care desi sunt de alta forma ele pot fi liniarizate prin aplicarea unei transformari sau a unei schimbari de notatie, abordare care simplifica foarte mult utilizarea acestora.

Ex.:

Modelul reciproc (invers) cunoscut in economie si sub denumirea de curba lui Philips: $Y = \beta_0 + \beta_1(1/X) + \varepsilon \Leftrightarrow Y = \beta_0 + \beta_1 X' + \varepsilon$

Model de crestere: $Y = e^{\beta_0 + \beta_1 X + \varepsilon} \Leftrightarrow \ln Y = \beta_0 + \beta_1 X + \varepsilon \Leftrightarrow Y' = \beta_0 + \beta_1 X + \varepsilon$

REGRESIA LINIARĂ MULTIPLĂ

C5.

1. Prezentarea modelului liniar multiplu
2. Estimarea parametrilor modelului liniar multiplu
3. Testarea parametrilor modelului liniar multiplu
4. Testarea modelului de regresie liniara multipla

Modelul liniar multiplu

Forma generala a modelului liniar multiplu este data prin relatia:

$$Y = M(Y / X_1, \dots, X_i, \dots, X_p) + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

unde:

Y - variabila dependentă;

$X_1, X_2, \dots, X_i, \dots, X_p$ - variabile independente (predictori);

ε - variabilă reziduu de modelare (variabila aleatoare);

β_i - parametrii modelului de regresie

k - numărul de parametri din model, $k=p+1$.

Exemplu: Pentru un esantion de 50 de marci de cereale, se poate studia legatura dintre **ratingul acordat de consumatori (Y)** unei marci de cereale si factorii de influenta (**nr. de calorii (X_1)**, **cantitatea (g/100g) de grasimi (X_2)**, **cantitatea (g/100g) de zahăr (X_3)**, **cantitatea (g/100g) de fibre (X_4)**, etc.)

Cei k parametri ai modelului liniar multiplu au următoarea semnificație:

β_0 – **valoarea medie a variabilei dependente Y** , în condițiile în care *influența variabilelor independente ar fi nula*;

$\beta_i = \frac{\partial Y}{\partial X_i}, i = \overline{1, p}$ - **variația absolută a variabilei dependente** la o variație absolută cu o unitate a variabilei independente X_i , în condițiile în care *influența celorlalte variabile independente este menținută constantă*. Arată **influența parțială** a fiecărei variabile independente asupra variabilei dependente.

Ipotezele modelului clasic de regresie (IIN):

- variabilele independente sunt *nestochastic*
- **normalitatea erorilor, media zero :** $\varepsilon_i \sim N(0, \sigma^2)$
- **homoscedasticitate:** $V(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2$
- **necorelarea erorilor:** $cov(\varepsilon_i, \varepsilon_j) = 0$
- **lipsa corelației dintre variabilele independente si variabila eroare**
- **lipsa coliniarității** sau a unei legături liniare între variabilele independente

Estimarea parametrilor modelului multiplu liniar

Se consideră modelul de **regresie liniara multipla cu doua variabile independente**:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

La nivelul unui esantion, modelul devine:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\varepsilon}_i \quad \text{sau} \quad y_i = \hat{y}_i + \hat{\varepsilon}_i$$

Rezultă $\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}$

Estimarea parametrilor modelului prin metoda celor mai mici pătrate presupune respectarea conditiei:

$$\sum_{i=1}^n \hat{\varepsilon}_i^2 = \text{minim} \quad , \text{ adica } \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2 = \mathbf{minim}$$

Estimarea punctuala

Pentru satisfacerea conditiei MCMMP trebuie ca derivatele partiale de ordin I în raport cu coeficientii modelului sa se anuleze. Astfel se va obtine un sistem de $2+1=3$ ecuatii cu 3 necunoscute. La nivelul unui esantion de date, sistemul de ecuatii devine:

$$\left\{ \begin{array}{l} nb_0 + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i \\ b_0 \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{1i} x_{2i} = \sum_{i=1}^n y_i x_{1i} \\ b_0 \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{1i} x_{2i} + b_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n y_i x_{2i} \end{array} \right.$$

Prin rezolvarea sistemului, se obtin relatiile pentru estimatiile parametrilor modelului de regresie.

Exemplu: Rating = 61.1 - 3.07 Grasimi - 2.21 Zahar + error

Estimarea parametrilor prin interval de încredere

Intervalele de încredere sunt de forma:

$$\beta_i \in [\hat{\beta}_i \pm t_{\alpha/2, n-k} \sigma_{\hat{\beta}_i}]$$

La nivelul unui eșantion de date se obține un interval de forma:

$$\beta_i \in [b_i - t_{\alpha/2, n-k} s_{\hat{\beta}_i}, b_i + t_{\alpha/2, n-k} s_{\hat{\beta}_i}]$$

Testarea individuală parametrilor modelului liniar multiplu

Testarea parametrilor modelului multiplu liniar se face la fel ca în cazul modelului simplu liniar:

1. Formularea ipotezelor $i=0, p$:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

2. Alegerea pragului de semnificație α

De regula, se asuma un risc $\alpha = 0,05$.

3. Alegerea statisticii test si calculul acesteia

$$t_{\text{calc}} = \frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}} \cong \frac{b_i}{s_{\hat{\beta}_i}}$$

4. Valoarea critica/teoretică a statisticii test

Pentru pragul de semnificație α ales și $U=n-k$ grade de libertate, se citește valoarea teoretică din tabela Student: $t_{\alpha/2; n-k}$

5. Regula de decizie/ compararea valorilor si interpretarea rezultatelor:

Dacă $|t_{\text{calc}}| > t_{\alpha/2, n-k} \Leftrightarrow \text{Sig } t < \alpha$ - se respinge H_0 cu un risc asumat de α

Dacă $|t_{\text{calc}}| \leq t_{\alpha/2, n-k} \Leftrightarrow \text{Sig } t \geq \alpha$ - se acceptă H_0 /nu se respinge H_0 cu o probabilitate de $1 - \alpha$

**Testarea modelului de regresie/ testarea
simultană a parametrilor modelului de regresie
multipla**

Testarea modelului de regresie se realizează cu ajutorul testului F, după următorul demers:

1. Formularea ipotezelor

H₀: $\beta_0 = \beta_1 = \dots = \beta_p = 0$ (modelul nu este semnificativ)

H₁: nu toți coeficienții sunt simultan zero

2. Alegerea pragului de semnificație α

3. Alegerea statisticii test și calculul acesteia:

$$F_{calc} = \frac{\hat{V}_E}{\hat{V}_R} \cdot \frac{n-k}{k-1} \cong \frac{ESS}{RSS} \cdot \frac{n-k}{k-1} \left(= \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1} \right) \sim F(\alpha, k-1, n-k)$$

4. Determinarea valorii critice a statisticii test: $F_{\alpha, k-1, n-k}$

5. Regula de decizie, compararea valorilor și interpretarea rezultatului:

Dacă $F_{calc} > F_{\alpha, k-1, n-k} \Leftrightarrow Sig F < \alpha \Rightarrow$ se **RH₀**, pentru risc asumat de α

Dacă $F_{calc} \leq F_{\alpha, k-1, n-k} \Leftrightarrow Sig F \geq \alpha \Rightarrow$ se **AH₀/nu se RH₀**, cu o probabilitate de **1- α**

EXEMPLU

Pentru un eșantion de marci de cereale, se studiază legătura dintre *ratingul acordat de consumatori* (Y) unei mărci de cereale și *cantitatea de grasimi* (grame/100g) (X_1), *cantitatea de zahar* (grame/100g) (X_2) și *cantitatea de fibre* (grame/100g) (X_3).

$n=77$ \rightarrow din tabelul ANOVA pentru $df_T = 76 = n-1 \Rightarrow n=77$

$\alpha=0.05=5\%$

$k=3+1=4$

$t_{0.025;77-4} = t_{0.025;73} \rightarrow z_{0.025} = 1.96$ (deoarece $v=n-k = 77-4 = 73 \gg 30$)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,789 ^a	,622	,612	8,75456

a. Predictors: (Constant), sugars, fat

$$Y = b_0 + b_1X_1 + b_2X_2 + e$$

$$Y = 61.089 - 3.066X_1 - 2.213X_2 + e$$

$$M(\Delta Y / X_1 = 0; X_2 = 0) = b_0 = 61.089$$

$$M(\Delta Y / \Delta X_1 = 1; \Delta X_2 = 0) = b_1 = -3.066$$

$$M(\Delta Y / \Delta X_2 = 1; \Delta X_1 = 0) = b_2 = -2.213$$

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9325,268	2	4662,634	60,836	,000 ^a
	Residual	5671,533	74	76,642		
	Total	14996,800	76			

a. Predictors: (Constant), sugars, fat

b. Dependent Variable: rating

$$IC_{\beta_0} = b_0 + / - t_{\alpha/2, n-k} \cdot s_{\beta_0} = 61.089 + / - t_{0.025; 77-4} \cdot 1.953 = 61.089 + / - 1.96 \cdot 1.953$$

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	61,089	1,953		31,284	,000
	fat	-3,066	1,036	-,220	-2,958	,004
	sugars	-2,213	,235	-,700	-9,428	,000

a. Dependent Variable: rating

Model Summary

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e$$

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,930 ^a	,865	,859	5,35086

a. Predictors: (Constant), fat, fiber, sugars

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12503,728	3	4167,909	145,570	,000 ^a
	Residual	1946,958	68	28,632		
	Total	14450,686	71			

a. Predictors: (Constant), fat, fiber, sugars

b. Dependent Variable: rating

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	53,673	1,389		38,637	,000
	fiber	2,938	,261	,507	11,265	,000
	sugars	-1,992	,150	-,622	-13,238	,000
	fat	-3,347	,656	-,238	-5,103	,000

a. Dependent Variable: rating