

d. Ipoteza de necorelare sau de independență a erorilor: $cov(\varepsilon_i, \varepsilon_i) = 0$.

1. Noțiuni

- a. Autocorelarea sau corelația serială :
- presupune existența unei autocorelări între erorile ε , altfel spus: $cov(\varepsilon_i, \varepsilon_i) \# 0$ sau $M(\varepsilon_i, \varepsilon_i) \# 0$.
- b. Coeficientul de autocorelație
- coeficientul de autocorelație între erorile ε_i și ε_{i-1} ale unui model de regresie se calculează după

relația:

$$\rho = \frac{cov(\varepsilon_{i}, \varepsilon_{i-1})}{\sigma_{i}\sigma_{i-1}} = \frac{cov(\varepsilon_{i}, \varepsilon_{i-1})}{\sigma^{2}}$$

- acest coeficient este un coeficient de autocorelație de ordinul 1.

c. Funcția de autocorelație

este definită de valorile coeficienților de autocorelare de ordinul k.

2. Sursa autocorelării erorilor

- neincluderea în modelul de regresie a uneia sau mai multor variabile explicative importante;
- inerția fenomenelor în timp și decalajul, în cazul seriilor de timp;
- modelul de regresie nu este corect specificat.

3. Testarea autocorelării erorilor

3.1. Testul Durbin-Watson

Ipoteze statistice:

 H_0 : erorile nu sunt autocorelate ($\rho = 0$)

 H_1 : erorile sunt autocorelate $(\rho \neq 0)$

Calculul statisticii test:

$$DW = d = \frac{\sum_{i=2}^{\infty} (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^{\infty} \hat{\varepsilon}_i^2}$$

Întrucât $\varepsilon_i = \rho \varepsilon_{i-1} + u_i$ itistica DW se mai poate scrie astfel:

$$d = \frac{\sum_{i} \hat{\varepsilon}_{i}^{2} - 2\sum_{i} \hat{\varepsilon}_{i} \hat{\varepsilon}_{i-1} + \sum_{i} \hat{\varepsilon}_{i-1}^{2}}{\sum_{i} \hat{\varepsilon}_{i}^{2}} \cong 2 \frac{\sum_{i} \hat{\varepsilon}_{i}^{2} - \sum_{i} \hat{\varepsilon}_{i} \hat{\varepsilon}_{i-1}}{\sum_{i} \hat{\varepsilon}_{i}^{2}} = 2 \cdot \left(1 - \frac{\sum_{i} \hat{\varepsilon}_{i} \hat{\varepsilon}_{i-1}}{\sum_{i} \hat{\varepsilon}_{i}^{2}}\right) = 2 \cdot (1 - \hat{\rho})$$

Deoarece $-1 \le \hat{\rho} \le 1$ alorile statisticii DW sunt date de intervalul: $0 \le d \le 4$

Interpretare:

- Dacă $\hat{\rho} = 1 \Rightarrow d = 0$ zunci există autocorelare pozitivă maximă a erorilor;
- Dacă $\hat{\rho} = -1 \Rightarrow d = 4$, atunci există autocorelare negativă maximă a erorilor;
- Dacă $\hat{\rho} = 0 \Rightarrow d = 2$ unci nu există autocorelare.

Regula de decizie:

- Valorile teoretice ale statisticii DW sunt calculate și tabelate în funcție de pragul de semnificație, de volumul eșantionului și de numărul de parametri ai modelului.
- În tabele se determină două valori critice, notate cu d_L (limita inferioară) și d_L (limita superioară).

- În funcție de aceste valori critice se determină următoarele intervale, care permit luarea deciziei de respingere sau acceptare a ipotezei nule:
- a. $(0 < d_{calc} < d_L)$ se respinge ipoteza H_o , erorile înregistrează o autocorelare pozitivă;
- b. $(d_L < d_{calc} < d_U)$ și $(4-d_u < d_{calc} < 4-d_L)$ sunt regiuni de nedeterminare, nu se poate decide asupra existenței autocorelării erorilor;

- c. $(d_u < d_{calc} < 4- d_u)$ se acceptă ipoteza Ho, erorile nu sunt autocorelate;
- d. (4-d_L <d_{calc}< 4) se respinge ipoteza H_o, erorile înregistrează o autocorelare negativă.

Exemplu:

În studiul legăturii dintre două variabile, X și Y, observate pentru un eșantion format din 25 unități statistice, s-a estimat un model de regresie liniară simplă și s-au obținut următoarele rezultate:

Calculul statisticii Durbin Watson în SPSS:

Model Summary^b

			Adjusted	Std. Error of	Durbin-
Model	R	R Square	R Square	the Estimate	Watson
1	,985 ^a	,970	,960	2,41523	1,429

a. Predictors: (Constant), X

b. Dependent Variable: Y

Să se testeze ipoteza de autocorelare a erorilor, considerând un risc de 0,05.

3.2. Runs Test

- se bazează pe ideea că valorile variabilei reziduale se constituie în secvențe sau seturi de valori pozitive sau negative numite *runs*, care se succed într-o anumită ordine sau aleator.
- ipoteza de bază a acestui test este aceea că, în cazul lipsei autocorelării erorilor, succesiunea acestor seturi (*runs*, notate *k*) este aleatoare sau numărul acestora este distribuit normal.

Ipoteze statistice

H₀: k este distribuit normal (nu există autocorelare a erorilor)

H₁: k nu este distribuit normal (ipoteza este încălcată)

Calculul statisticii test

- se foloseste statistica t Student, calculată după relația:

$$t_{calc} = \frac{k - M(k)}{S_k}$$

unde: k este numărul de runs caracterizat prin:

$$M(k) = 2\frac{n_1 n_2}{n_1 + n_2} + 1$$

$$s_k^2 = 2n_1n_2 \frac{2n_1n_2 - n_1 - n_2}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

- \square n_i este numărul de valori pozitive ale erorilor e_i ;
- \square n_2 este numărul de valori negative ale erorilor e_i , cu $n_1 + n_2 = n$.

Regula de decizie:

- dacă $|t_{calc}| \stackrel{\leq}{=} t_{\mathbf{a}/2,n-2}$, atunci se acceptă ipoteza H_0 .

Exemplu:

Pentru două variabile, X și Y, s-au obținut următoarele rezultate:

Runs Test 2

	Unstandardiz ed Residual
Test Value ^a	,0000000
Cases < Test Value	17
Cases >= Test Value	15
Total Cases	32
Number of Runs	3
Z	-4,849
Asymp. Sig. (2-tailed)	,000

a. Mean

Să se testeze ipoteza de autocorelare a erorilor, folosind testul Runs.

4. Efectele încălcării ipotezei de necorelare a erorilor

 \Box în aceste condiții, prin aplicarea MCMMP, se obține pentru parametrul β₀ un estimator neeficient.

5.2. Ipotezele asupra variabilelor independente

- 1. Variabilele independente sunt nestochastice sau deterministe.
- 2. Variabilele independente și variabila eroare sunt necorelate, cov $(X_i, \varepsilon_i)=0$.
- această ipoteză este îndeplinită dacă variabilele independente sunt nestochastice.

3. Ipoteza de necoliniaritate a variabilelor independente a. Definire

Coliniaritatea poate fi definită ca o legătură liniară funcțională existentă între două sau mai multe variabile independente ale unui model de regresie multiplă de forma:

a.1. Coliniaritate perfectă

$$Y_{X_1,\dots X_p} = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_p \cdot X_p + \varepsilon$$

apare atunci când între variabilele independente $X_1, X_2, ..., X_p$ există o legătură liniară perfectă, funcțională. Această legătură poate fi exprimată printr-o relație de forma:

$$\lambda_1 \cdot X_1 + \lambda_2 \cdot X_2 + \dots + \lambda_p \cdot X_p = 0$$

unde: λ_i , cu i=1, ..., p, valori constante care nu sunt toate, în mod simultan, nule.

a.2. Coliniaritatea imperfectă

Poate fi definită ca o relație liniară puternică existentă între două sau mai multe variabile independente.

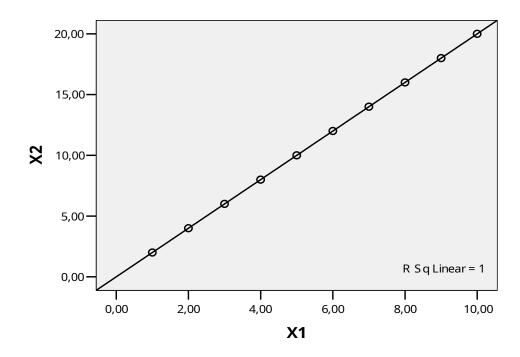
$$\lambda_1 \cdot X_1 + \lambda_2 \cdot X_2 + \dots + \lambda_p \cdot X_p + u = 0$$

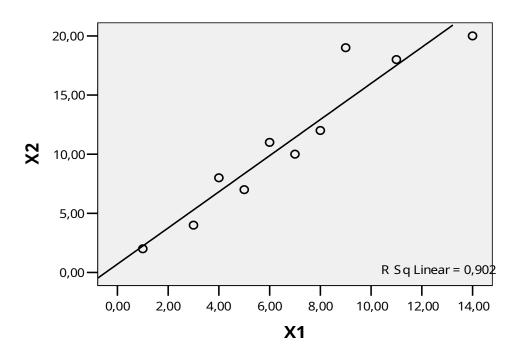
b. Efectele coliniarității

- □ Varianța estimatorilor parametrilor modelului de regresie crește, adică estimatorii pierd proprietatea de eficiență.
- În cazul unei coliniarități perfecte, varianța estimatorilor este infinită (parametrii modelului nu pot fi estimați).
- În cazul unei coliniarități imperfecte, varianțele estimatorilor sunt mari.

c. Testarea coliniarității

c.1. Folosind procedee grafice





c.2. Folosind procedee numerice

1. Factorul varianței crescute (variance-inflated factor, VIF)

$$VIF_{j} = \frac{1}{1 - R_{j}^{2}}$$

unde: R_j^2 este raportul de determinație din modelul de regresie auxiliar, respectiv dintre variabila X_j și celelalte variabile independente.

Interpretare:

- Atunci când legăturile dintre variabilele independente sunt puternice, valoarea raportului de determinație se apropie de unu, iar raportul VIF este infinit.
- Atunci când între variabilele independente nu există corelație (R²_i=0), valoarea raportului VIF este egală cu unu.

- În practică, o valoare VIF>10 indică prezența coliniarității.
- 2. Toleranța (Tolerance)
- se calculează după relația: $TOL=1/VIF=1-R_{j}^{2}$

Interpretare:

- Dacă valoarea TOL=1, atunci nu există coliniaritate;
- Dacă valoarea TOL=0, atunci există coliniaritate perfectă.
- Existența coliniarității este sugerată de valorile mici ale indicatorului *TOL*.

d. Corectarea coliniarității

- metodele de corecție țin cont de tipul de coliniaritate, de numărul de variabile din model și de informațiile suplimentare despre fenomenul studiat.
- □ Cea mai simplă metodă constă în eliminarea variabilei care introduce coliniaritatea.
- Altă metodă constă în construirea unui model cu variabile transformate prin diverse funcții sau operatori (de exemplu, operatorul decalaj, diferență).

Exemplu:

În urma analizei legăturilor dintre variabilele independente ale unui model de regresie, s-au obținut următoarele rezultate:

Coefficientsa

		Unstandardized Coefficients		Standardized Coefficients			Collinearity	Statistics
Model		В	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	65,705	27,731		2,369	,037		
	X1	48,979	10,658	,581	4,596	,001	,950	1,052
	X2	59,654	23,625	,359	2,525	,028	,753	1,328
	X3	-1,838	,814	-,324	-2,258	,045	,738	1,355

a. Dependent Variable: Y