4

STUDIUL STATISTIC AL LEGĂTURILOR DINTRE FENOMENELE ŞI PROCESELE ECONOMICO-SOCIALE

Procesele şi fenomenele economice (şi nu numai) apar şi se dezvoltă ca urmare a unor cauze variate, care pot acţiona în acelaşi sens sau în sensuri opuse şi cu diferite grade de intensitate. Drept urmare acestea sunt legate între ele prin conexiuni, uneori foarte complexe, care nu sunt cunoscute sau observate de la bun început, ci, de regulă, sunt descoperite pe măsura studierii lor. Manifestarea unui sau altuia dintre procese sau fenomene generează efecte care pot provoca apariţia, modificarea sau încetarea altora determinând astfel relaţii de interdependenţă sau cauzalitate. Complexitatea interacţiunii dintre fenomene este cu atât mai mare cu cât acestea aparţin unor colectivităţi mai numeroase. De aici deducem că fenomenele şi procesele economico-sociale nu sunt univoc determinate fiind rezultatul conjugării influenţei mai multor fenomene-cauză, iar în sistemul acesta de conexiuni nu toate raporturile de dependenţă prezintă aceeaşi importanţă întrucât există factori ce se compensează reciproc.

Studiul statistic al raporturilor de dependenţă dintre procese şi fenomene se concentrează pe identificarea relaţiei care există între două sau mai multe caracteristici. Importanţa cunoaşterii legăturii dintre un fenomen sau proces şi a cauzelor care-l generează şi determină este deosebită întrucât numai în acest fel se creează posibilitatea reală de control şi influenţare a acestuia.

În context, devine necesară utilizarea unor metode, tehnici şi instrumente care să poată:

- indica existenţa sau absenţa legăturii;
- măsura intensitatea acesteia;
- preciza sensul în care acţionează;
- descrie, eventual, forma legăturii.

Statistica pune la dispoziţie astfel de metode, tehnici şi instrumente, unele simple, altele extrem de laborioase şi complexe.

4.1. Tipuri de legături dintre fenomenele şi procesele economice

Aşa cum am amintit, formele de manifestare a relaţiilor de interdependenţă dintre procese şi fenomene sunt extrem de variate şi cel mai adesea dificil de sesizat. Problema esenţială care trebuie rezolvată în analiza legăturii dintre o *variabilă dependentă* (rezultativă, efect, explicată) notată de regulă cu y şi una sau mai multe *variabile independente* (factoriale, cauzale, explicative) notate de regulă cu x_i se referă la răspunsul

întrebării: există o legătură între variabile sau modificarea variabilei rezultative-efect *y* este influenţată de modificarea variabilei (variabilelor) cauză (x_i).

În practică se întâlnesc următoarele situaţii:

a) variabila independentă *x* determină modificarea variabilei dependente *y*, caz în care între cele două există o *legătură univocă;*
b) cele două variabile se influenţează reciproc (*legătură reciprocă*);
c) variabilele evoluează similar independent, influenţate însă de o altă variabilă simultan;
d) variabilele au o evoluţie similară fără să existe vreo legătură între ele.

Din această cauză, pentru studiul sistematic al relaţiilor dintre cele două tipuri de variabile este necesară clasificarea lor în funcţie de anumite criterii:

a) după natura relaţiei de interdependenţă (de cauzalitate), distingem:

- *legături funcţionale (deterministe);*
- *legături statistice sau stochastice;*

Legăturile funcţionale sunt univoce şi se realizează direct între un fenomen-cauză şi un fenomen-efect. În această situaţie fenomenul-efect depinde de o singură cauză, ce poate fi identificată ori de câte ori se produce. De aici concluzia că dacă se menţin aceleaşi condiţii, atunci unei valori a caracteristicii fenomenului-cauză îi corespunde o singură valoare a caracteristicii fenomenului-efect (caracteristicii rezultative). Aceste legături se mai numesc şi legături de tip determinist şi relaţia matematică prin care putem descrie un astfel de tip de legătură este:

$$y = f(x),$$

unde: y – fenomenul-efect;
x – fenomenul-cauză;

Legăturile funcţionale de tipul $y = f(x)$ se întâlnesc rar în activitatea economică deoarece, de cele mai multe ori, modificarea variabilei efect *y* este rezultatul influenţei simultane a mai multor variabile-cauză (x_i).

Legăturile statistice sau stochastice, cel mai frecvent întâlnite în realitate, se caracterizează prin faptul că variabila rezultativă *y* este influenţată de una sau mai multe variabile cauză x_i, considerate ca esenţiale, dar pe lângă acestea există şi acţionează şi alte variabile neînregistrate sau nespecificate.

Influenţa variabilelor nespecificate este luată în calcul sub forma variaţiei reziduale (*e*) numită şi eroare aleatoare astfel că relaţia matematică ce descrie o astfel de legătură este:

$y = f(x) + e$ – cazul unei singure variabile cauză,
$y = f(x_i) + e$ – cazul mai multor variabile cauză.

b) după numărul variabilelor factoriale, distingem:

- *legături simple* – o variabilă efect *y* şi o singură variabilă cauză *x*. Exemplu: profitul (*y*), cifra de afaceri (*x*);
- *legături multiple* – o variabilă efect *y* şi două sau mai multe variabile cauză x_i. Exemplu: salariul (*y*), numărul de ore lucrate (x_1), vechimea în muncă (x_2) şi nivelul calificării (x_3);

c) după natura caracteristicilor, distingem:

- *legături de asociere* – se referă la raporturile de interdependenţă dintre caracteristicile calitative sau dintre o caracteristică numerică şi una calitativă. Exemplu: ramura de activitate – salariul mediu; calificare – productivitate; domeniul de activitate şi dimensiunea întreprinderii. Studiul statistic al legăturilor de asociere este posibil numai în situaţia în care variantele pot

exprimate numeric. De exemplu, clasele de calitate ale produselor: 0 – produse inferioare, 1 – produse medii, 2 – produse superioare;
- *legături de corelaţie* – corelaţia statistică intervine numai în cazul legăturilor de tip cauză-efect dintre două sau mai multe *variabile cantitative;*

d) după direcţia legăturii, distingem:
- *legături directe* – modificarea într-un sens a variabilei (variabilelor) cauză este însoţită de modificarea în acelaşi sens a variabilei efect. Exemplu: salariu-productivitatea muncii, ofertă-preţ;
- *legături inverse* – modificarea într-un sens a variabilei (variabilelor) cauză este însoţită de modificarea în sens opus a variabilei efect. Exemplu: cerere-preţ;

e) după forma funcţiei sau expresia analitică prin care se descrie legătura, distingem:
- *legături liniare* – când legătura este pusă în evidenţă printr-o funcţie liniară;
- *legături neliniare* – când legătura este pusă în evidenţă printr-o funcţie neliniară (parabolă, hiperbolă etc.);

f) după timpul realizării legăturii, distingem:
- *legături sincrone* – când modificarea variabilei efect se produce aproape în acelaşi timp cu cea a variabilei (variabilelor) cauză. Exemplu: modificarea preţurilor şi a cererii;
- *legături asincrone* – când modificarea variabilei efect se produce la un anumit timp (defazat) de la modificarea variabilei (variabilelor) cauză. Exemplu: modificarea investiţiilor în economie şi modificarea produsului intern brut (PIB).

4.2. Metode statistice utilizate în studiul legăturii dintre fenomenele şi procesele economice

Metodele statistice utilizate pentru studiul legăturii dintre două sau mai multe fenomene (puse în evidenţă de anumite caracteristici) pot fi grupate în:
- metode elementare;
- metode analitice.

Metodele elementare sunt cele prin care se poate determina existenţa legăturii dintre fenomene, a tăriei, a sensului şi a formei acesteia dar nu cu o precizie foarte mare, ele fiind de obicei folosite pentru orientarea către metode de altă natură, mai rafinate, pentru determinarea elementelor de mai sus foarte precis.

Metodele analitice sunt cele prin care se pot determina aceleaşi elemente ca şi prin metodele elementare, dar cu o precizie mult mai mare, ele permiţând, de asemenea, şi studiul legăturii dintre un fenomen efect şi mai multe fenomene cauză simultan.

4.2.1. Metode elementare utilizate în studiul legăturii dintre fenomenele şi procesele economice

În categoria metodelor elementare se includ:
- metoda seriilor paralele interdependente;
- metoda grupărilor;
- metoda tabelului de corelaţie;

- metoda grafică.

- **Metoda seriilor paralele interdependente**

Este o metodă relativ simplă ce se recomandă a fi aplicată în cazul existenţei unui număr redus de valori pentru variabilele *y* şi *x*.

Aplicarea acestei metode presupune parcurgerea următoarelor etape:

- ordonarea crescătoare a datelor ce caracterizează variabila independentă (cauză) *x;*
- ataşarea corespunzătoare a valorilor variabilei dependente (efect) *y;*
- desprinderea concluziilor referitoare la forma şi direcţia legăturii în raport de reacţia lui *y* la modificările *x* astfel:
  - datele se modifică în acelaşi sens → corelaţie directă;
  - datele se modifică în sensuri diferite → corelaţie inversă.

Mărimea modificării lui *y* funcţie de modificările lui *x* permite o apreciere empirică a intensităţii legăturii.

- **Metoda grupărilor**

Este în fapt o variantă a metodei precedente. Potrivit acestei metode se grupează în prealabil unităţile colectivităţii după caracteristica factorială *x*. Pentru fiecare grupă se calculează media caracteristicii dependente *y*. În coloane paralele, se înscriu grupările ordonate ale caracteristicii *x* şi mediile corespunzătoare ale lui *y*. Prin compararea variaţiei celor două caracteristici *x* şi *y* obţinem informaţii ce permit formularea de concluzii privind existenţa, sensul şi intensitatea legăturii.

În cazul în care se analizează o singură variabilă rezultativă în raport de mai multe variabile factoriale (corelaţie multiplă) se înregistrează pe grupe valorile caracteristicilor factoriale înscriindu-se valorile respective în coloane distincte, în ordinea importanţei lor pentru caracteristica rezultativă. Pentru caracteristica rezultativă se calculează valorile medii condiţionate pe grupe.

- **Metoda tabelului de corelaţie**

Permite evidenţierea tuturor elementelor necesare pentru confirmarea existenţei unei legături dintre două fenomene, pe baza observaţiei modului de manifestare. Pentru aceasta se utilizează măsurătorile unor variabile care caracterizează fenomenele supuse studiului.

Pentru utilizarea acestei metode este necesară distribuţia bidimensională obţinută prin prelucrarea perechilor de valori determinate prin măsurarea celor două variabile care caracterizează fenomenul cauză, respectiv fenomenul efect.

Modul în care se distribuie frecvenţele în interiorul acestei distribuţii (tabelul 4.1.) oferă toate elementele pentru evidenţierea unei eventuale legături între cele două fenomene.

Tabelul 4.1.

X \ Y	y_1	y_2	y_3	$...y_j...$	y_n	Fx
x_1	f_{11}	f_{12}	f_{13}	$...f_{1j}...$	f_{1n}	Fx_1
x_2	f_{21}	f_{22}	f_{23}	$...f_{2j}...$	f_{2n}	Fx_2
x_3	f_{31}	f_{32}	f_{33}	$...f_{3j}...$	f_{3n}	Fx_3
$\vdots$ x_i $\vdots$	$\vdots$ f_{i1} $\vdots$	$\vdots$ f_{i2} $\vdots$	$\vdots$ f_{i1} $\vdots$	... f_{ij} ...	$\vdots$ f_{ni} $\vdots$	$\vdots$ Fx_i $\vdots$
x_n	f_{n1}	f_{n2}	f_{n3}	$...f_{nj}...$	f_{nn}	Fx_n
Fy	Fy_1	Fy_2	Fy_3	$...Fy_j...$	Fy_n	F

unde: X - variabila cauză;
Y - variabila efect;
$x_1...x_n$ - valorile variabilei cauză;
$y_1...y_n$ - valorile variabilei efect;
f_{ij} - frecvenţa de apariţie a perechii de valori (x_i, y_j);
Fx_i - frecvenţa de apariţie a valorii x_i;
Fy_i - frecvenţa de apariţie a valorii y_i;
F - numărul total de perechi de valori (x_i, y_j).

Atunci când este posibil se recomandă folosirea intervalelor egale de grupare, un număr suficient de grupe şi acelaşi număr de grupe pentru ambele caracteristici.

Elementele care pot fi evidenţiate cu ajutorul acestei metode:

1. Existenţa legăturii dintre variabila X factorială şi Y rezultativă:

Dacă frecvenţele f_{ij} se distribuie într-o bandă grupată de-a lungul unei diagonale a tabelului (figurile 4.1., 4.2.);

2. Sensul legăturii:

Dacă banda în care sunt grupate frecvenţele f_{ij} se află pe diagonala tabelului care corespunde aceluiaşi sens de variaţie a valorilor corespunzătoare celor două variabile X,Y înseamnă că între cele două variabile există o legătură directă (figura 4.1.). Dacă se află pe cealaltă diagonală care corespunde sensului diferit de variaţie a celor două variabile X,Y atunci legătura dintre cele două variabile este inversă (figura 4.2.).

3. Intensitatea legăturii:

Este dată de lăţimea benzii în care sunt grupate frecvenţele f_{ij}. Cu cât banda este mai îngustă cu atât intensitatea legăturii creşte (figurile 4.3., 4.4.).

4. Forma legăturii:

Este dată de forma benzii, putând fi liniară dacă forma benzii este liniară (figura 4.5.) sau neliniară dacă banda are altă formă decât cea liniară (figura 4.6.).

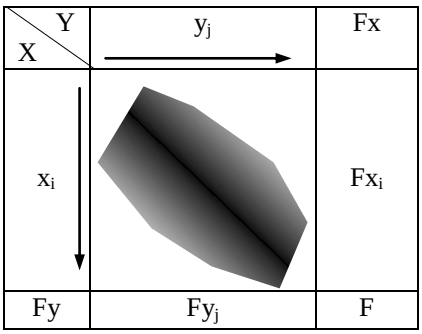


Figura 4.1. Corelaţie directă

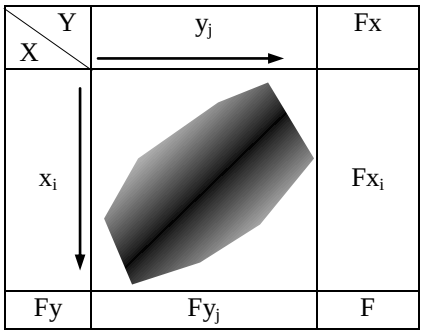


Figura 4.2. Corelaţie inversă

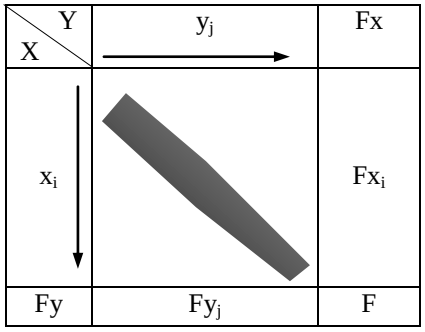


Figura 4.3. Corelaţie puternică

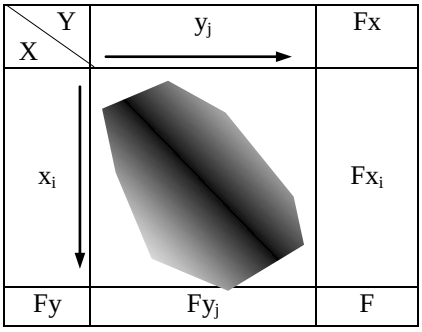


Figura 4.4. Corelaţie slabă

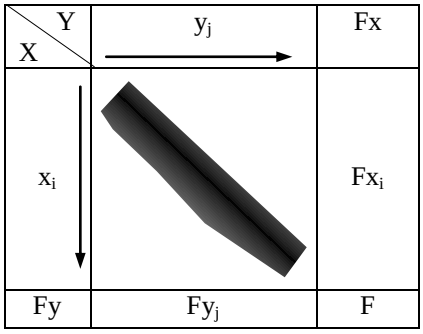


Figura 4.5. Corelaţie liniară

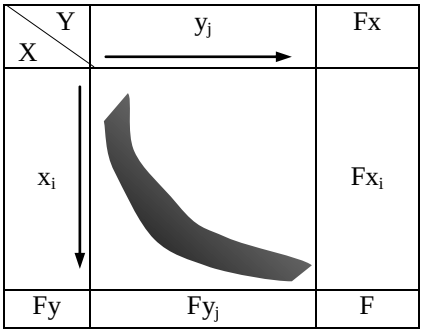


Figura 4.6. Corelaţie neliniară

Exemplul 4.1.
Pentru muncitorii unei firme se cunosc următoarele date (tabelul 4.2.):

Tabelul 4.2.

vechimea (ani) / vârsta (ani)	1-5	5-10	10-15	15-20	20-25	25-30	Total
18-25	**1**	-	-	-	-	-	1
25-32	-	**2**	-	-	-	-	2
32-39	-	-	**3**	**2**	-	-	5
39-46	-	-	-	**3**	**1**	-	4
46-53	-	-	-	-	**2**	**4**	6
53-60	-	-	-	-	**2**	**2**	4
Total	1	2	3	6	5	5	22

În tabelul anterior se poate observa existenţa unei corelaţii directe, între cele două variabile, de intensitate mare şi formă aproape liniară.

- **Metoda grafică**

Ca şi metoda precedentă permite evidenţierea prin apreciere vizuală a elementelor ce caracterizează legătura dintre două variabile.

În acest caz este necesară construirea corelogramei. Pe abscisă se trec valorile scării de reprezentare corespunzătoare variabilei cauză X, iar pe ordonată, valorile scării de reprezentare corespunzătoare variabilei Y. Prin unirea cu segmente de dreaptă a punctelor obţinute reprezentând grafic perechile de valori (x_i, y_j) se obţine corelograma (figura 4.7.).

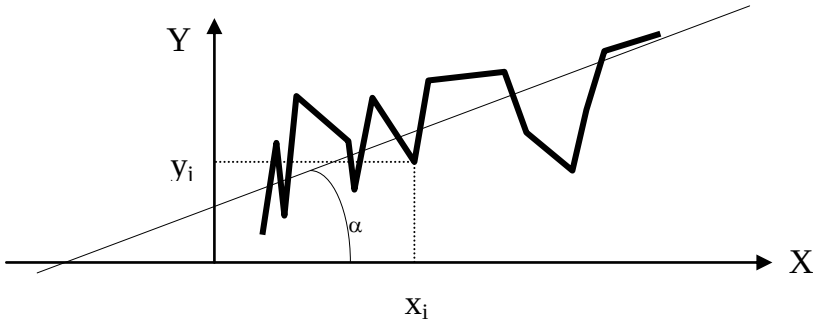


Figura 4.7. Corelograma.

Cu ajutorul acestei metode se pot evidenţia:

1. Existenţa legăturii:

Se determină prin existenţa unghiului α (diferit de 0) realizat de linia de tendinţă cu orizontala .

2. Sensul legăturii:

- legătură directă - atunci când linia de tendinţă este ascendentă (figura 4.8.);
- legătură inversă - atunci când linia de tendinţă este descendentă (figura 4.9.);

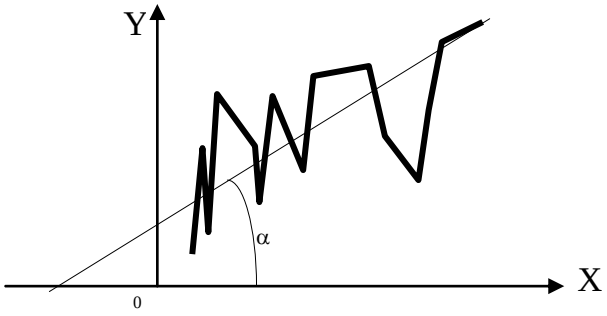


Figura 4.8. Corelaţie directă.

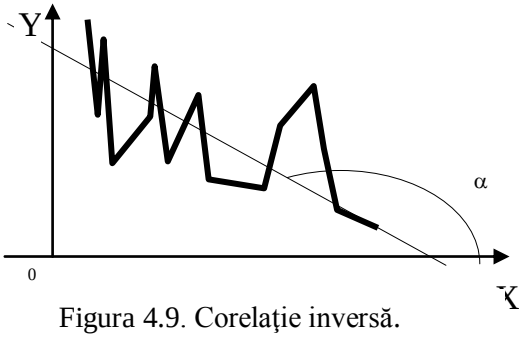


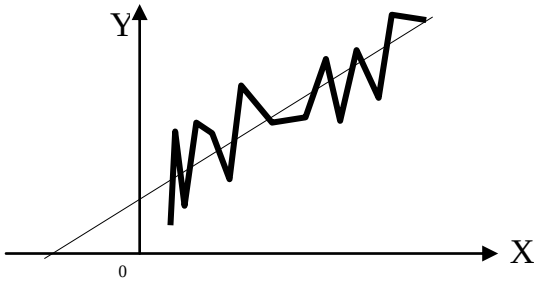
Figura 4.9. Corelaţie inversă.

3. Intensitatea legăturii:

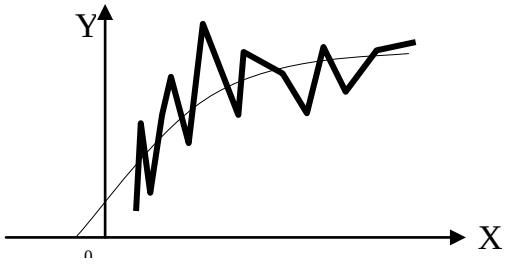
Este dată de mărimea unghiului α. Cu cât α este mai mare cu atât legătura este mai intensă (şi invers).

4. Forma legăturii:

Este dată de forma corelogramei (figura 4.10.).



Corelaţie liniară



Corelaţie neliniară

Figura 4.10.

Exemplul 4.2.

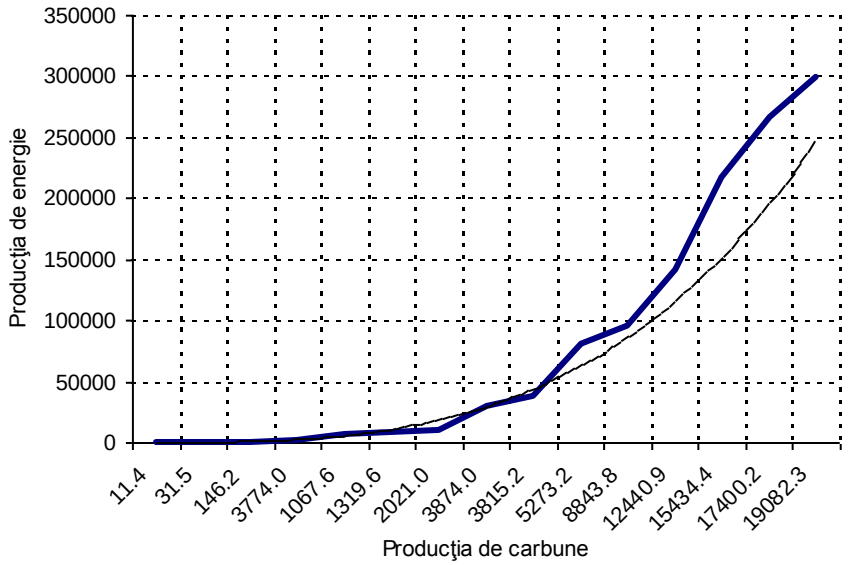
În urma unui studiu efectuat pentru dependenţa dintre producţia de energie electrică şi termică şi cea de cărbune, în perioada 1990-2004, s-au determinat datele (tabelul 4.3.):

Tabelul 4.3.

miliarde lei preţuri curente

Anul	**Producţia de cărbune**	**Producţia de energie electrică şi termică, gaze şi apă**
1990	11,4	61,2
1991	31,5	210,4
1992	146,2	735,3
1993	3774	1553,2
1994	1067,6	5877,0
1995	1319,6	7581,3
1996	2021,0	10506,5
1997	3874,0	28763,9
1998	3815,2	37689,4
1999	5273,2	79889,1
2000	8843,8	94826,7
2001	12440,9	141103,0
2002	15434,4	217040,6
2003	17400,2	266203,2
2004	19082,3	299615,9

Sursa datelor: Anuarul statistic al României 2005



Din corelogramă rezultă că între cele două variabile (producţia de cărbune şi producţia de energie) există o legătură directă, neliniară (exponenţială) cu o intensitate destul de ridicată.

4.2.2. Metode analitice (parametrice) de măsurare a legăturilor dintre fenomenele şi procesele economice

Metodele parametrice sunt cele care permit determinarea precisă atât a legăturii dintre două sau mai multe variabile cât şi a intensităţii acesteia. Metodele parametrice sunt:

- metoda regresiei;
- metoda coeficientului (raportului) de corelaţie.

4.2.2.1. Metoda regresiei

Se bazează pe utilizarea funcţiilor matematice pentru descrierea formei legăturii dintre variabile. Funcţia de regresie are forma generală:

$y = f(x_1, x_2, ..., x_n) + \varepsilon$,

unde: y – variabila dependentă (efect);

$x_1, x_2, ..., x_n$ – variabilele independente (factorii de influenţă);

n – numărul factorilor de influenţă (variabilelor independente);

ε – variabila aleatoare (perturbatoare) sau eroarea ce sintetizează influenţa factorilor nespecificaţi (de regulă greu de cuantificat sau nesemnificativi).

În raport de numărul factorilor de influenţă înregistraţi avem:

- regresie simplă (unifactorilă);
- regresie multiplă (multifactorilă).

Regresia simplă

Se bazează pe funcţia:

$y = f(x) + \varepsilon$

şi studiază variaţia unei caracteristici rezultative (dependente) y în raport cu un singur factor de influenţă x ceilalţi factori fiind consideraţi neglijabili şi cu acţiune constantă.

Alegerea funcţiei se face cu ajutorul graficului de corelaţie. Cele mai frecvent utilizate funcţii de corelaţie simplă sunt:

- funcţia liniară: $y = a + bx$;
- funcţia parabolică: $y = a + bx + cx^2$;
- funcţia exponenţială: $y = ab^x$;
- funcţia hiperbolică: $y = \frac{1}{a + bx}$ sau $y = a + b\frac{1}{x}$;
- funcţia logaritmică: $y = a \cdot lgx$ sau $y = a + b \cdot lgx$,

unde: y – variabila dependentă sau rezultativă;

x – variabila independentă sau factorială;

a, b, c – parametrii ce urmează a fi determinaţi.

Pentru determinarea concretă a valorilor numerice a parametrilor se utilizează, de obicei, metoda celor mai mici pătrate, conform căreia pentru ca funcţia de regresie aleasă să fie cu adevărat semnificativă trebuie să avem:

$$S = \sum_{i=1}^{n} \left(y_i - y_{x_i}\right)^2 \to min, \qquad (*)$$

unde: $i = 1, 2, ..., n$ – numărul unităţilor statistice observate;

y_i - valorile empirice (observate) ale variabilei dependente;

y_{x_i} – valorile teoretice exprimate prin ecuaţia de regresie.

Din condiţia de mai sus rezultă că suma pătratelor abaterilor valorilor reale observate (y_i)de la valorile exprimate prin ecuaţia de regresie y_{x_i} trebuie să fie minimă.

Pentru calculul concret al parametrilor funcţiilor se anulează derivatele parţiale în raport cu fiecare parametru (*a, b, c etc.*) al expresiei (*), obţinându-se un sistem cu un număr de ecuaţii egal cu numărul parametrilor funcţiei. Prin rezolvarea acestui sistem se obţin valorile concrete ale parametrilor.

În continuare vom ilustra metoda pentru funcţia liniară

$y = a + bx$.

Vom avea condiţia:

$$S = \sum [y_i - a - b \cdot x_i]^2 \to min$$

$$\frac{\partial S}{\partial a} = -2\sum (y_i - a - b \cdot x_i) \cdot 1 = 0$$

$$\frac{\partial S}{\partial b} = -2\sum (y_i - a - b \cdot x_i) \cdot x_i = 0$$

După calcule simple şi ţinând cont de faptul că *a* şi *b* sunt constante, rezultă sistemul:

$$\begin{cases} na + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i \end{cases}$$

care va avea soluţia:

$$b = \frac{n\sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \cdot \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} \text{ sau}$$

$$a = \bar{y} - b\bar{x}$$

Sistemele de ecuaţii normale specifice celor mai uzuale funcţii de regresie sunt prezentate în tabelul 4.4.

Tabelul 4.4.

Sistemele de ecuaţii normale ale principalelor funcţii de extrapolare

Tipul funcţiei	**Funcţia**	**Sistemul de ecuaţii normale corespunzător**
Liniară	$y = a \pm bx$	$\begin{cases} na \pm b\sum x = \sum y \\ a\sum x \pm b\sum x^2 = \sum xy \end{cases}$
Parabolică	$y = a + bx + cx^2$	$\begin{cases} na + b\sum x + c\sum x^2 = \sum y \\ a\sum x + b\sum x^2 + c\sum x^3 = \sum xy \\ a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum x^2 y \end{cases}$
Exponenţială	$y = ab^x$ (*)	$\begin{cases} n\,lg\,a + lg\,b\sum x = \sum lg\,y \\ lg\,a\sum x + lg\,b\sum x^2 = \sum x\,lg\,y \end{cases}$

Tipul funcţiei	**Funcţia**	**Sistemul de ecuaţii normale corespunzător**
Putere	$y = ax^b$ (**)	$\begin{cases} n\lg a + b\sum \lg x = \sum \lg y \\ \lg a\sum \lg x + b\sum (\lg x)^2 = \sum \lg x \cdot \lg y \end{cases}$
Hiperbolică	$y = \frac{1}{a+bx}$ (***)	$\begin{cases} na + b\sum x = \sum \frac{1}{y} \\ a\sum x + b\sum x^2 = \sum \frac{1}{y} \cdot x \end{cases}$
Logistică clasică	$y = \frac{k}{1+be^{-cx}}$ (****)	$\begin{cases} (n-1)A + B\sum y = \sum \frac{\Delta y}{y} \\ A\sum y + B\sum y^2 = \sum \Delta y \end{cases}$
Törnquist	$y = \frac{kx}{x+a}$ (*****)	$\begin{cases} n \cdot \frac{1}{k} + \frac{a}{k}\sum \frac{1}{x} = \sum \frac{1}{y} \\ \frac{1}{k}\sum \frac{1}{x} + \frac{a}{k}\sum \frac{1}{x^2} = \sum \frac{1}{xy} \end{cases}$

(*) se liniarizează mai întâi, prin logaritmare, şi $\Rightarrow$ *lgy = lga + x·lgb*.

(**) se liniarizează mai întâi, prin logaritmare, şi $\Rightarrow$ *lgy = lga + blgx*.

(***) se porneşte de la inversa sa: $\frac{1}{y} = a + bx$.

(****) Se scrie, în primul rând, forma transformată a acesteia $\frac{\Delta y}{y} = c - \frac{c}{k} \cdot y$ şi se parcurg două etape de lucru. În prima etapă se calculează parametrii *c* şi *k,* unde *k* reprezintă nivelul de saturaţie. În acest scop se notează *c=A* şi *–c/k = B* şi rezultă $\frac{\Delta y}{y} = A + By$. De aici obţinem sistemul de ecuaţii normale inserat în tabel. Observăm că parametrul *A* se înmulţeşte cu numărul termenilor seriei diminuat cu 1, deoarece Δy_i reprezintă diferenţa dintre y_i şi y_{i-1}, şi numărul termenilor Δy_i este mai mic cu 1 decât numărul termenilor y_i. În etapa a doua se calculează parametrul *b*, pornind de la relaţia

$$\frac{k}{y} - 1 = be^{-cx}.$$

Prin logaritmare (în acest caz folosim logaritmii naturali, pentru că modelul matematic conţine numărul *e*, adică baza acestor logaritmi) se obţine:

$$ln\left(\frac{k}{y} - 1\right) = \ln b - cx \Leftrightarrow \ln b = ln\left(\frac{k}{y} - 1\right) + cx.$$

Fiind vorba de o serie statistică cu *n* variabile x_i şi y_i relaţia devine:

$$\ln b = \frac{1}{n}\left[\sum ln\left(\frac{k}{y} - 1\right) + c\sum x\right].$$

Când variabila independentă este timpul, *b* se calculează după relaţia:

$$\ln b = \frac{1}{n}\left[\sum \ln\left(\frac{k}{y}-1\right) + c\cdot\frac{n(n+1)}{2}\right],$$

unde *n* reprezintă numărul anilor din perioada de analiză retrospectivă.

(*****) Ca şi la funcţia hiperbolică, se porneşte de la inversa sa:

$$\frac{1}{y} = \frac{1+a}{kx} = \frac{1}{k}\cdot\frac{x+a}{x} = \frac{1}{k}\left(1+\frac{a}{x}\right) = \frac{1}{k}+\frac{a}{k}\cdot\frac{1}{x}.$$

De regulă aceste sisteme se rezolvă cu ajutorul calculatorului utilizând programe adecvate.

Dacă totuşi se apelează la rezolvarea manuală se construieşte un tabel pentru determinarea constantelor ecuaţiilor respective. De exemplu, pentru funcţia parabolică de ordinul 2, tabelul va avea următorul conţinut (tabelul 4.5.):

Tabelul 4.5.

Anii	***y***	***x***	x^2	x^3	x^4	***xy***	x^2y
t_1	y_1	x_1	x_1^2	x_1^3	x_1^4	$x_1 y_1$	$x_1^2 y_1$
t_2	y_2	x_2	x_2^2	x_2^3	x_2^4	$x_2 y_2$	$x_2^2 y_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
t_i	y_i	x_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
t_n	y_n	x_n	x_n^2	x_n^3	x_n^4	$x_n y_n$	$x_n^2 y_n$
Σ	$\sum y$	$\sum x$	$\sum x^2$	$\sum x^3$	$\sum x^4$	$\sum xy$	$\sum x^2$

În cazul rezolvării manuale, pentru uşurarea calculelor, se poate proceda astfel: în locul valorilor x_i se consideră $x_i' = x_i - \bar{x}$ unde:

$\bar{x} = \frac{\sum x_i}{n}$ - media aritmetică simplă a valorilor x_i observate

În acest fel vom avea:

$$x_1' = x_1 - \frac{x_1 + x_2 + \ldots + x_n}{n}$$

...

$$x_i' = x_i - \frac{x_1 + x_2 + \ldots + x_n}{n}$$

...

$$x_n' = x_n - \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Sumând aceste relaţii vom obţine:

$$\sum x_i' = \sum x_i - n\cdot\frac{\sum x_i}{n} = 0$$

În acest fel, în locul variabilelor x_i se lucrează cu variabilele x_i' a căror sumă este 0. De aici rezultă că şi sumele de puteri impare ale lui x_i' devin tot 0.

Prin acest procedeu se simplifică mult calculele astfel:

a) în cazul funcţiei de regresie liniară $y = a + bx$, sistemul de ecuaţii normale devine:

$$\begin{cases} na = \sum y \\ b\sum x^2 = \sum x' y \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} a = \dfrac{\sum y}{n} = \bar{y} \\ b = \dfrac{\sum x' y}{\sum x^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} \cdot \dfrac{1}{\sigma_x^2} = cov(x,y) \cdot \dfrac{1}{\sigma_x^2} \end{cases}$$

b) în cazul funcţiei parabolice de ordinul II, $y = a + bx + cx^2$, vom avea:

$$\begin{cases} na + c\sum (x')^2 = \sum y \\ b\sum (x')^2 = \sum x' y \\ a\sum (x')^2 + c\sum (x')^4 = \sum (x')^2 y \end{cases}$$

După aflarea parametrilor funcţiei de regresie se calculează valorile teoretice (ajustate) ale variabilei y pe baza ecuaţiei explicitate (ecuaţia în care a şi b au valorile numerice rezultate din calcule).

Pentru verificarea calculului parametrilor funcţiei de regresie se utilizează relaţia $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_{x_i}$ ceea ce arată că prin ajustare nu se face decât o redistribuire a influenţei factorilor.

Exemplul 4.3.

În ultimii ani o firmă a obţinut următoarele rezultate economice (tabelul 4.6.):

Tabelul 4.6.

Anii	t_1	t_2	t_3	t_4
Cifra de afaceri (mii $)	2200	3400	3800	4700
Bugetul de publicitate (mii $)	45	54	76	81

Conducerea întreprinderii doreşte să ştie care va fi cifra de afaceri dacă bugetul de publicitate va fi majorat la 150 mii $.

Pentru rezolvarea acestei probleme ne propunem utilizarea unei funcţii de corelaţie simplă de forma:

$y = a + bx$, unde

y = cifra de afaceri (în mii $)

x = bugetul de publicitate (în mii $)

Aplicând metoda celor mai mici pătrate simplificată vom scrie

$minS = \sum_{i=1}^{n} (y_i - a - bx')^2$ de unde anulând derivatele parţiale în raport de a şi b obţinem:

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n} (y_i - a - bx') \cdot 1 = 0$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n} (y_i - a - bx') \cdot x' = 0$$

Dezvoltând va rezulta sistemul

$$\begin{cases} na + b\sum x' = \sum y \\ a\sum x' + b\sum (x')^2 = \sum x' y \end{cases} \quad \text{unde } x' = x_i - \bar{x}$$

Datele necesare rezolvării sistemului se vor obţine din tabelul 4.7.

Tabelul 4.7.

Anii	y	x	$x' = x_i - \bar{x}$	$(x')^2$	$x'y$	y_i'	$y_i - y'$	$\frac{y_i - y'}{y_i} \cdot 100$	$\left(\frac{y_i - y'}{y_i} \cdot 100\right)^2$	y^2
t_1	2200	45	-19	361	-41800	2468,79	-268,79	-12,22	149,33	4840000
t_2	3400	54	-10	100	-34000	2969,10	430,90	12,67	160,53	11560000
t_3	3800	76	12	144	45600	4192,08	-392,08	10,32	106,50	14440000
t_4	4700	81	17	289	79900	4470,03	229,97	4,89	23,91	22090000
Σ	14100	256	0	894	49700	*	*	*	440,27	52930000

$$\bar{x} = \frac{256}{4} = 64$$

Rezultă sistemul:

$$\begin{cases} 4a = 14110 \\ 894b = 49700 \end{cases}$$

De unde $a=3525$ şi $b=55,59$.

Rezultă ecuaţia dreptei de regresie:

$Y = 3525 + 55,59x'$ şi

$y_1' = 2468,79$

$y_2' = 2969,1$

$y_3' = 4192,08$

$y_4' = 4470,03$

Calculăm abaterea medie pătratică procentuală cu relaţia:

$$\sigma_{\%} = \sqrt{\frac{\sum\left(\frac{y_i - y'}{y_i} \cdot 100\right)^2}{n}} = \sqrt{\frac{440,27}{4}} = 10,50\%$$

Coeficientul de corelaţie care măsoară legătura dintre y şi x se poate calcula cu relaţia:

$$r = \frac{n\sum x' y - \sum x' \cdot \sum y}{\sqrt{\left[n\sum y^2 - \sum (y)^2\right]\left[n\sum (x')^2 - (\sum x')^2\right]}} =$$

$$r = \frac{4 \cdot 49700}{\sqrt{\left[4 \cdot 52930000 - 19881000\right]\left[4 \cdot 894 - 0^2\right]}} =$$

$$r = \frac{198800}{\sqrt{46166160000}} = 0,9252$$

Cum coeficientul de corelaţie este apropiat de 1 putem utiliza ecuaţia

$y = 3552 + 55,59x'$

pentru estimarea cifrei de afaceri.

Cifra de afaceri *y* pentru un buget de publicitate *x* de 150 mii $ va fi
y = 3552 + 55,59(150 - 64) = 8305,74 mii $.

Exemplul 4.4.

Pentru ultimele cinci luni se cunosc datele statistice privind preţurile unitare practicate (x) şi cantitatea corespunzătoare de produse vândute (y), conform tabelului 4.8. (coloanele 1 şi 2). Să se studieze relaţia dintre cele două variabile şi să se estimeze vânzările pentru un preţ x = 9,5 cu o probabilitate de 95 %.

Tabelul 4.8.

Luna	x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	x^2	xy	y^2	$(y-\bar{y})^2$
1	13	25	1,5	2,25	-14	-21	169	325	625	196
2	12	30	0,5	0,25	-9	-4,5	144	360	900	81
3	11,5	45	0	0	6	0	132,25	517,5	2025	36
4	11	45	-0,5	0,25	6	-3	121	495	2025	36
5	10	50	-1,5	2,25	11	-16,5	100	500	2500	121
Total	57,5	195	0	5	0	-45	666,25	2197,5	8075	470
Media	11,5	39	-	-	-	-	-	-	-	-

În acest scop vom parcurge următoarele etape:

1) Se reprezintă grafic datele statistice *x* şi *y* (vezi figura 4.11., linie discontinuă)

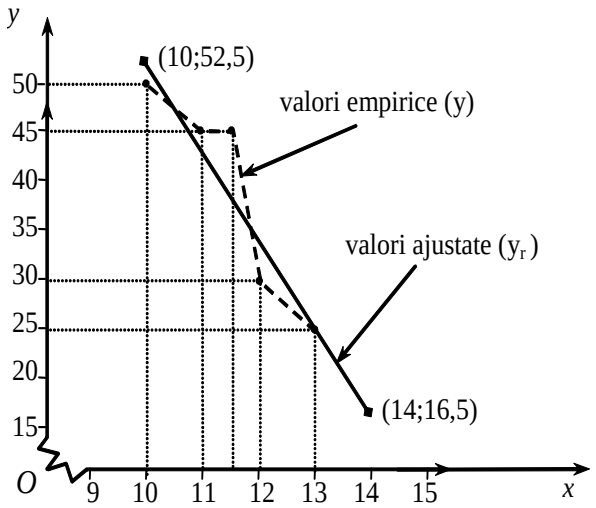


Figura 4.11. Evoluţia vânzărilor funcţie de preţ.

Dacă punctele marcate sunt relativ aliniate înseamnă că între *x* şi *y* există o legătură liniară. Dacă punctele rezultate au un mare grad de împrăştiere sau au o alură curbilinie înseamnă că între cele două variabile există alt tip de dependenţă şi, în consecinţă, nu se mai parcurg etapele cerute de regresia simplă.

2) Se calculează coeficientul de corelaţie (r), astfel:

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2} \cdot \sqrt{\sum (y-\bar{y})^2}}$$

Cu datele din tabelul 4.8. rezultă $r = \frac{-45}{\sqrt{5} \cdot \sqrt{470}} = -0{,}9282791.$

Interpretare:

- dacă r este pozitiv, între cele două variabile există o legătură directă, iar dacă *r* este negativ (cazul de mai sus), variabilele se află într-o legătură inversă (când *x* creşte, *y* scade şi invers);
- valoarea absolută a lui *r* constituie un indiciu al corelaţiei (legăturii) dintre *x* şi *y*: legătura este foarte strânsă când r tinde spre *1* şi foarte redusă când *r* tinde spre zero.

3) Se calculează coeficientul de determinare (R^2) , astfel:

$R^2=r^2$

în cazul dat, $R^2 =(-0{,}928279\ 1)^2 =0{,}8617021$

Interpretare:

- dacă R^2 ia valoarea minimă $R^2 = 0$ înseamnă că între cele două variabile *x* şi *y* nu există nici o legătură liniară, iar dacă $R^2 = 1$ – maxim, înseamnă că variabilele sunt perfect legate una de alta.

Complementul lui R^2, respectiv ($1-R^2$), se numeşte coeficient de nedeterminare şi arată proporţia în care *y* nu este explicat de *x* ci de alţi factori neluaţi în considerare.

În cazul de mai sus se poate afirma că factorul preţ explică volumul vânzărilor în proporţie de $0{,}8617021 = 86{,}17\ \%$, în timp ce restul $(1-R^2) = 0{,}1382979 = 13{,}83\ \%$ se datorează influenţei altor factori.

4) Ajustarea datelor empirice

Convinşi fiind de mărimea lui R^2 că între cele două variabile există o indiscutabilă legătură liniară,se caută acea dreaptă numită „dreaptă de regresie" de forma $y_r = a+bx$ care să reprezinte cât mai fidel datele reale care au o evoluţie greu de exprimat matematic.

În acest scop, pe baza metodei celor mai mici pătrate care minimizează suma pătratelor diferenţelor dintre y şi y_r, coeficienţii a şi b rezultă din sistemul:

$$\begin{cases} na + b\Sigma x = \Sigma y \\ a\Sigma x + b\Sigma x^2 = \Sigma xy \end{cases}$$

unde: n = numărul datelor empirice

Cu datele problemei (vezi şi tabelul 4.8.), avem:

$$\begin{cases} 5a + 57{,}5b = 195 \\ 57{,}5a + 666{,}25b = 2197{,}5 \end{cases} \Rightarrow \quad \begin{matrix} a = 142{,}5 \\ b = -9 \end{matrix}$$

Coeficienţii a şi b mai pot fi calculaţi şi direct, astfel:

$$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{-45}{5} = -9$$

$$a = \bar{y} - b\bar{x} = 39-(-9 \cdot 11{,}5) = 39 + 103{,}5 = 142{,}5$$

Ca urmare, pentru cazul dat,

$y_r = 142{,}5 - 9x$ (vezi şi reprezentarea grafică în figura 4.11.)

Observaţii:

- coeficientul *b* (panta dreptei de regresie) are următoarea semnificaţie: o scădere cu o unitate a lui *x* (preţul) conduce la o creştere a lui *y* (vânzări) cu *b* unităţi şi invers.
- dacă coeficienţii a şi b au fost bine calculaţi trebuie să existe relaţia:

$\bar{y} = a + b\bar{x}$

5) Calculul erorii dintre *y* şi y_r

Datorită ajustării efectuate, între valorile reale ale lui *y* şi valorile corespunzătoare de pe dreapta $y_r = 142{,}5 - 9x$, vor exista în mod firesc deosebiri (vezi figura 4.11. şi tabelul 4.9., coloanele 2 şi 3).

Tabelul 4.9.

Luna	*x*	*y*	y_r	$(y_r - \bar{y}_r)^2$	$(y - y_r)^2$
1	13	25	25,5	182,25	0,25
2	12	30	34,5	20,25	20,25
3	11,5	45	39	0	36
4	11	45	43,5	20,25	2,25
5	10	50	52,5	182,25	6,25
Total	57,5	195	195	405	65
Medie	11,5	39	39	-	-

Important este ca aceste abateri să fie cât mai mici, respectiv, diferenţele (erorile) să fie cuantificate pentru a fi luate în considerare atunci când se va face previziunea. Deoarece întotdeauna $\bar{y} = \bar{y}_r$, rezultă că abaterile dintre valorile lui *y* şi y_r nu pot fi măsurate decât de dispersie, în termenii analizei dispersionale, se pot face următoarele asocieri:

- variaţia totală a datelor y reale = $\sum (y - \bar{y})^2$ =*470* (vezi col. 10, tabelul 4.8.);
- variaţia totală a datelor y_r estimate = $\sum (y_r - \bar{y}_r)^2$ = *405* (vezi col.4, tabelul 4.9.);
- variaţia dintre datele reale şi cele estimate = $\sum (y - y_r)^2$ = *65* (vezi col.5, tabelul 4.9.).

Cele de mai sus se interpretează astfel: din cele 470 de unităţi ale variaţiei datelor reale, 450 sunt explicate de funcţia y_r, iar restul de 65 rămân neexplicate.

În aceeaşi optică, R^2 şi $(1-R^2)$ se pot calcula astfel:

$$R^2 = \frac{\sum (y_r - \bar{y}_r)^2}{\sum (y - \bar{y})^2} = \frac{\sum (y_r - \bar{y})^2}{\sum (y - \bar{y})^2}$$

$$\left(1 - R^2\right) = \frac{\sum (y - y_r)^2}{\sum (y - \bar{y})^2}$$

Cu datele problemei, $R^2 = \frac{405}{470} = 0{,}8617021$ şi $\left(1 - R^2\right) = \frac{65}{470} = 0{,}1382979$, adică exact rezultatele obţinute anterior.

Variaţia dintre datele reale şi cele estimate $\sum (y - y_r)^2$ se mai numeşte şi variaţie reziduală sau variaţie neexplicată sau eroare de estimaţie. Se exprimă sub forma abaterii medii pătratice (e) astfel:

$$e = \sqrt{\frac{\sum (y - y_r)^2}{\upsilon}},$$ unde υ = numărul gradelor de libertate

Observaţii:

- deoarece se cunosc doi parametri (coeficienţii *a* şi *b*), în cazul regresiei simple $\upsilon = n-2$;
- pentru un anumit nivel de semnificaţie (α) şi un număr dat de grade de libertate (υ), R^2 trebuie să aibă o valoare minimă teoretică. Pentru ca dreapta de regresie să poată fi

considerată ca reprezentativă (în cazul de faţă, unde $\upsilon = (n - 2) = (5 - 2) = 3$, pentru $\alpha = 0,05$, $R^2_{calculat} = 0,8617021 > R^2$ *teoretic = 0,7714*).

În situaţia concretă presupusă,

$$e = \sqrt{\frac{65}{5-2}} = 4,65$$

6) Se estimează cererea (volumul vânzărilor)

Prin similitudine cu eşantionajul, unde $m = \bar{x} \pm t \cdot S_{\bar{x}}$

$P = (a+bx) \pm t \cdot e$

Eşantionul fiind mic, pentru $\alpha = 0,05$ şi $\upsilon = n-1 = 5-1 = 4$, $t = 2,778$ şi, ca urmare:

$P = (142,5-9-9,5) \pm 12,9$, respectiv $44,1 < y < 69,9$

Concluzie. Cu o probabilitate de 95%, în cazul practicării unui preţ de 9,5 unităţi băneşti / produs, în luna următoare cererea (vânzările) se va situa între 44 şi 70 bucăţi.

Regresia multifactorială

În viaţa economică reală se înregistrează influenţe multiple chiar şi asupra celor mai simple fenomene şi procese. De aceea, modelul regresiei unifactoriale este un model simplificat. Mult mai apropiat de realitate este modelul regresiei multiple (multifactoriale) bazat pe ecuaţia de tipul:

$$y_{x_i} = f(x_1, x_2, ..., x_n) + \varepsilon .$$

În acest model variaţia fenomenului efect (y_{x_i}) se datorează influenţei unei multitudini de factori cauză (independenţi) explicitaţi prin intermediul variabilelor x_1, x_2, ..., x_n şi aleatori explicitaţi prin valoarea reziduală ε.

Având în vedere complexitatea abordării multifactoriale cel mai accesibil şi utilizat model este cel liniar de forma:

$$y_{x_i} = a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n ,$$

unde: y_{x_i} – valorile ajustate (teoretice) ale variabilei efect (dependente);

$x_1, x_2, ..., x_n$ – factorii de influenţă înregistraţi;

a_0 – parametru ce exprimă influenţa factorilor neînregistraţi;

$a_1, a_2, ..., a_n$ – parametri, coeficienţi parţiali de regresie care arată cu cât se modifică y_{x_i}, atunci factorii de influenţă înregistraţi $x_1, x_2, ..., x_n$ se modifică cu o unitate, iar toate celelalte variabile rămân constante (a_1 arată cu cât se modifică y_{x_i} dacă x_1 se modifică cu o unitate ceilalţi factori rămânând neschimbaţi ş.a.m.d.).

Pentru determinarea parametrilor funcţiei se utilizează metoda celor mai mici pătrate pornind de la relaţia:

$$S = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_1 - a_2 x_2 - ... - a_n x_n)^2 \to min$$

Anulând derivatele parţiale ale expresiei de mai sus în raport cu parametrii a_0, a_1, $a_2, ..., a_n$ se obţine un sistem de ecuaţii normale care rezolvat conduce la valorile numerice ale parametrilor $a_0, a_1, a_2, ..., a_n$. Acest sistem este:

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 + \ldots + a_n\sum x_n = \sum y \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + \ldots + a_n\sum x_1x_n = \sum x_1y \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 + \ldots + a_n\sum x_2x_n = \sum x_2y \\ \vdots \\ a_0\sum x_i + a_1\sum x_1x_i + a_2\sum x_2x_i + \ldots + a_n\sum x_ix_n = \sum x_iy \\ \vdots \\ a_0\sum x_n + a_1\sum x_1x_n + a_2\sum x_2x_n + \ldots + a_n\sum x_n^2 = \sum x_ny \end{cases}$$

Dacă avem în vedere modelul liniar cu numai doi factori de influenţă $y_{x_i} = a_0 + a_1x_1 + a_2x_2$ va rezulta sistemul

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 = \sum y \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 = \sum x_1y \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 = \sum x_2y \end{cases}$$

În cazul regresiei multifactoriale este posibilă existenţa unor interdependenţe între factorii de influenţă (fenomenul de multicoliniaritate) caz în care efectele acestuia afectează şi influenţează concluziile analizei.

Calitatea ajustării prin intermediul funcţiei de regresie se apreciază cu ajutorul indicatorilor:

a) eroarea standard calculată prin intermediul relaţiei:

$$S_{y_i/y_{x_i}} = \sqrt{\frac{\sum\left(y_i - y_{x_i}\right)}{n}}$$

Relaţia de mai sus este în fapt abaterea medie pătratică a valorilor reale y_i faţă de cele teoretice y_{x_i}.

b) coeficientul de eroare *e* (sau abaterea medie pătratică procentuală $\sigma_{\%}$) calculat pe baza relaţiei:

$$e = \frac{S_{y_i/y_{x_i}}}{\bar{y}} \cdot 100$$

$$\sigma_{\%} = \sqrt{\frac{\sum\left(\frac{y_i - y_{x_i}}{y_i} \cdot 100\right)^2}{n}}$$

Funcţia aleasă este cu atât mai reprezentativă cu cât valorile celor doi indicatori sunt mai apropiate de 0.

c) coeficientul de determinaţie calculat pe baza relaţiei:

$$R = \left(1 - \frac{\sum\left(y_i - y_{x_i}\right)^2}{\sum\left(y_i - \bar{y}\right)^2}\right) \cdot 100$$ care arată proporţia în care variabila (variabilele) independentă *x* (x_i) explică variaţia caracteristicii variabilei dependente y_{x_i}.

Variaţia totală a lui *y* faţă de media sa $\bar{y}$ se poate scrie:

$y_i - \bar{y} = \left(y_{x_i} - \bar{y}\right) + \left(y_i - y_{x_i}\right)$,

unde: $y_i - \bar{y}$ - variaţia totală;

$\left(y_{x_i} - \bar{y}\right)$ - variaţia explicată de regresie;

$\left(y_i - y_{x_i}\right)$ - variaţia neexplicată de regresie.

Pornind de la relaţia de mai sus şi calculând:

- *dispersia totală a lui y:* $\sigma_y^2 = \frac{\sum\left(y_i - \bar{y}\right)^2}{n}$ care exprimă influenţa tuturor factorilor asupra variabilei efect *y* putem descompune această dispersie (influenţă) în două:
  - dispersie explicată de factorii cuprinşi în ecuaţia de regresie

    $\sigma_{y/x_i}^2 = \frac{\sum\left(y_{x_i} - \bar{y}\right)^2}{n}$ şi care exprimă influenţa tuturor factorilor explicitaţi asupra variaţiei lui *y;*
  - dispersie neexplicată prin modelul ecuaţiei de regresie (reziduală)

    $\sigma_{y/r}^2 = \frac{\sum\left(y_i - y_{x_i}\right)^2}{n}$ care exprimă influenţa factorilor reziduali (neexplicaţi în model).

Pentru validarea modelului de regresie se utilizează testul Fisher-Snedecor (testul F) conform căruia

$$F_{calc} = \frac{\frac{\sum\left(y_{x_i} - \bar{y}\right)^2}{k-1}}{\frac{\sum\left(y_i - y_{x_i}\right)^2}{n-k}}$$

unde: k – numărul parametrilor funcţiei de regresie (modelului);

n – numărul perechilor de valori $x_i \cdot y_i$.

Valoarea calculată F_{calc} se compară cu valoarea teoretică a lui F obţinută din tabel $F_{\alpha,k-1;n-k}$ pentru un prag de semnificaţie (probabilitate) α şi $k-1$, $n-k$ grade de libertate.

Modelul de regresie se validează dacă:

$F_{calc} > F_{\alpha,k-1;n-k}$

Modelul bazat pe regresie constituie numai o ipoteză statistică prin care se exprimă tendinţa medie a legăturii dintre variabila dependentă *y* şi variabila (variabilele) independentă *x* (x_i) şi reprezintă primul pas pentru măsurarea intensităţii legăturii, lucru ce se realizează prin metoda corelaţiei.

4.2.2.2. Metoda corelaţiei

Pentru măsurarea intensităţii legăturii dintre variabila dependentă *y* şi variabila (variabilele) independentă *x* (x_i) se utilizează metoda corelaţiei. În funcţie de natura legăturii dintre variabila dependentă *y* şi variabila (variabilele) independentă *x* (x_i) – legătura directă sau inversă – corelaţia poate să fie pozitivă (în cazul legăturii directe) sau negativă (în cazul legăturii inverse). În cadrul acestei metode se utilizează indicatorii: covarianţa, coeficientul de corelaţie şi raportul de corelaţie.

Covarinţa surprinde existenţa şi direcţia legăturii dintre variabila dependentă *y* şi o variabilă independentă *(x)*. Se calculează sub forma mediei aritmetice simple a produselor

abaterilor celor două variabile corelate *y* şi *x* de la mediile lor aritmetice $\bar{y}$ şi $\bar{x}$ cu ajutorul relaţiei:

$$cov(x,y) = \frac{1}{n} \cdot \sum (x_i - \bar{x})(y_i - \bar{y}).$$

Valorile pozitive ale acestui indicator reflectă o legătură directă, iar cele negative o legătură inversă.

Valorile mari ale indicatorului arată o legătură puternică, în timp ce valorile apropiate de zero semnifică lipsa de legătură între variabilele *y* şi *x*.

Coeficientul de corelaţie simplă măsoară intensitatea *legăturii liniare* dintre două variabile *y* – rezultativă (endogenă) şi *x* – factorială.

Deoarece fenomenele aflate în relaţii de interdependenţă prezintă, în mod normal, o poziţie similară a valorilor individuale (y_i şi x_i) faţă de media corespunzătoare ($\bar{y}$ şi $\bar{x}$) va rezulta că şi abaterile normal normate $\frac{(y_i - \bar{y})}{\sigma_y}$, respectiv $\frac{(x_i - \bar{x})}{\sigma_x}$ au mărimi apropiate pentru valorile perechi (y_i, x_i).

Pentru a obţine mărimea sintetică a abaterilor normal normate la nivelul întregii colectivităţi se calculează *coeficientul de corelaţie* $r_{y/x}$ cu ajutorul relaţiei:

$$r_{y/x} = \frac{\sum \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right)}{n}.$$

numit şi *coeficientul de corelaţie liniară al lui Pearson.*

Se observă că $r_{y/x} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$.

Ţinând cont de relaţiile de calcul pentru σ_x şi σ_y se obţine relaţia de calcul simplificat:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{\left[n \sum x_i^2 - (\sum x_i)^2 \right] \left[n \sum y_i^2 - (\sum y_i)^2 \right]}}.$$

Examinând elementele din formula de calcul simplificat se observă că exceptând $\sum y_i^2$ toate celelalte se găsesc în tabelul de calcul al parametrilor funcţiei de regresie liniară. De aceea, în tabelele de lucru pentru calculul parametrilor funcţiei de regresie se recomandă includerea coloanei y_i^2.

În cazul cunoaşterii sumei abaterilor de la medie pentru cele două variabile se recomandă utilizarea relaţiei echivalente:

$$r_{y/x} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}.$$

Între coeficientul de corelaţie liniară simplă $r_{y/x}$, coeficientul de regresie (*b*) al funcţiei liniare $y=a+bx$ şi abaterile medii pătratice ale lui *x* şi *y* există relaţia:

$$b = r_{y/x} \cdot \frac{\sigma_y}{\sigma_x}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Deoarece $\sum y_i = \sum y_{x_i}$ se poate înlocui $\bar{y}$ cu $\bar{y}_{x_i}$ şi cum cele două medii sunt cunoscute ($\bar{y}$ şi $\bar{x}$) rezultă:

$a = \bar{y} - b\bar{x}$.

Cum legătura dintre cele două variabile este liniară rezultă:

$r_{y/x} = \sqrt{b_{y/x} \cdot b_{x/y}}$,

unde: $b_{y/x}$ – coeficientul de regresie corespunzător dependenţei lui *y* faţă de *x* (coeficientul *b* din ecuaţia $y=a+bx$);

$b_{x/y}$ – coeficientul de regresie corespunzător dependenţei lui *x* faţă de *y* (coeficientul *b* din ecuaţia $x=a+by$).

Coeficientul de corelaţie poate lua valori între -1 şi 1 astfel că:

$r_{y/x} \le |1|$.

Dacă: $r_{y/x} \in [-1, 0)$ – legătură inversă;

$r_{y/x} \in (0, 1]$ – legătură directă;

$r_{y/x} = 0$ – cele două variabile nu se corelează liniar.

În practică, funcţie de valorile lui $r_{y/x}$ avem:

- $|r_{y/x}| \in (0;0,2)$ – nu există legătură;
- $|r_{y/x}| \in [0,2;0,5)$ – există o legătură slabă;
- $|r_{y/x}| \in [0,5;0,75)$ – legătură de intensitate medie;
- $|r_{y/x}| \in [0,75;0,95)$ – legătură puternică;
- $|r_{y/x}| \in [0,95;1]$ – relaţie (legătură) deterministă.

Formula anterioară este utilizată în cazul unui număr relativ redus de valori individuale pentru variabilele y_i şi x_i .

În situaţia unor observaţii mai ample datele statistice pot fi sistematizate prin grupări simple sau combinate.

Dacă s-a utilizat gruparea simplă şi variabilele y_i şi x_i au frecvenţe comune f_i formula de calcul a coeficientului de corelaţie simplă devine:

$$r_{y/x} = \frac{\sum f_i \cdot \sum x_i y_i f_i - \sum x_i f_i \cdot \sum y_i f_i}{\sqrt{\left[\sum f_i \cdot \sum x_i^2 f_i - \left(\sum x_i f_i\right)^2\right]\left[\sum f_i \cdot \sum y_i^2 f_i - \left(\sum y_i f_i\right)^2\right]}}$$

În situaţia unei distribuţii bidimensionale variabilele y_i şi x_i au frecvenţe distincte f_j cât şi frecvenţe comune f_{ij} . Din acest considerent formula de calcul a coeficientului de corelaţie simplă devine:

$$r_{y/x} = \frac{\left(\sum_i \sum_j f_{ij} \cdot \sum_i \sum_j x_i y_j f_{ij}\right) - \sum_i x_i f_i \cdot \sum_j y_j f_j}{\sqrt{\left[\sum_i f_i \cdot \sum_i x_i^2 f_i - \left(\sum_i x_i f_i\right)^2\right]\left[\sum_j f_j \cdot \sum_j y_j^2 f_j - \left(\sum_j y_j f_j\right)^2\right]}} .$$

Verificarea semnificaţiei coeficientului de corelaţie liniară simplă se face cu ajutorul testului „t” (STUDENT) parcurgând următorii paşi:

1) Se determină $t_{calculat}$ cu relaţia:

$$t_{calc} = \frac{|r_{y/x}|}{\sqrt{1-r_{y/x}^2}} \cdot \sqrt{n-2},$$

unde: $r_{y/x}$ – coeficientul de corelaţie liniară simplă;
n – numărul observaţiilor;
$n-2$ – numărul gradelor de libertate.

2) Se compară valoarea rezultată din calcul (t_{calc}) cu valoarea teoretică din tabelul repartiţiei Student (t_{tab}) – $t(\alpha,f)$ în raport cu α (probabilitatea cu care se doreşte garantarea rezultatului) şi $f=n-2$ (numărul gradelor de libertate).
3) Dacă $t_{calc} \geq t_{tab}$ – coeficientul de corelaţie liniară simplă este semnificativ;
Dacă $t_{calc} < t_{tab}$ – influenţa caracteristicii factoriale x asupra lui y nu este reală sau nu este garantată cu probabilitatea dorită.

În cazul corelaţiei multiple de tip liniar se recomandă calculul coeficienţilor de corelaţie liniară simplă luând pe rând câte un factor x_i pentru a măsura intensitatea legăturii sale cu variabila dependentă y. Se va obţine astfel câte un coeficient de corelaţie liniară simplă pentru fiecare caracteristică factorială înregistrată: $r_{y/x_1}, r_{y/x_2}, ..., r_{y/x_n}$. Aceştia vor putea fi utilizaţi pentru calculul coeficientului de corelaţie multiplă $R_{y/x_1,x_2,...,x_n}$.

Raportul de corelaţie

Pentru măsurarea intensităţii legăturii dintre variabila dependentă y şi variabila independentă x în cazul funcţiilor de regresie neliniare se utilizează raportul de corelaţie simplă. În scopul determinării modelului de calcul al raportului de corelaţie se porneşte de la ideea că variaţia totală a caracteristicii rezultative y are două componente:

1) o componentă esenţială (determinantă) explicată prin influenţa caracteristicii factoriale x (variabila cauză esenţială);
2) o componentă neesenţială (reziduală) explicată prin influenţa factorilor aleatori (neînregistraţi).

Adâncind analiza, putem pune în evidenţă trei feluri de abateri:

- abaterea valorilor empirice ale lui y_i de la medie: $(y_i - \bar{y})$ sintetizată la nivelul seriei în *dispersia totală* $\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$, care reflectă influenţa tuturor factorilor esenţiali şi neesenţiali (întâmplători);
- abaterea valorilor calculate pe baza funcţiei de regresie (valori teoretice) $(y_i - y_{x_i})$ exprimată pe total prin *dispersia reziduală* $\sigma_{y/r}^2 = \frac{\sum (y_i - y_{x_i})^2}{n}$ şi care măsoară influenţa factorilor aleatori;
- abaterea valorilor teoretice de la medie $(y_{x_i} - \bar{y})$ sintetizată la nivelul seriei de *dispersia sistematică* $\sigma_{y/x}^2 = \frac{\sum (y_{x_i} - \bar{y})^2}{n}$ şi care arată influenţa variabilei independente x, considerată ca factor determinant al variaţiei y.

În acest fel putem scrie:

$y_i - \bar{y} = (y_i - y_{x_i}) + (y_{x_i} - \bar{y})$ la nivelul fiecărui y_i şi ţinând cont de câteva transformări elementare şi de relaţiile de calcul ale dispersiilor obţinem:

$\sigma_y^2 = \sigma_{y/r}^2 + \sigma_{y/x}^2$.

Împărţim prin σ_y^2 şi găsim:

$1 = \frac{\sigma_{y/r}^2}{\sigma_y^2} + \frac{\sigma_{y/x}^2}{\sigma_y^2} \Rightarrow$ notând $N_{y/x}^2 = \frac{\sigma_{y/r}^2}{\sigma_y^2}$ şi $R_{y/x}^2 = \frac{\sigma_{y/x}^2}{\sigma_y^2}$

$1 = N_{y/x}^2 + R_{y/x}^2$ (**)

$N_{y/x}^2 = \frac{\sigma_{y/r}^2}{\sigma_y^2}$ – coeficient de nedeterminaţie;

$R_{y/x}^2 = \frac{\sigma_{y/x}^2}{\sigma_y^2}$ – coeficient de determinaţie.

Coeficientul de determinaţie arată ponderea (proporţia) din variaţia totală a fenomenului reprezentat de *y* explicată prin variaţia factorului înregistrat *x*.

Coeficientul de nedeterminaţie arată ponderea (proporţia) factorilor aleatori în variaţia totală a fenomenului reprezentat de *y*.

Din relaţia (**) deducem:

$R_{y/x}^2 = 1 - N_{y/x}^2 = 1 - \frac{\sigma_{y/r}^2}{\sigma_y^2}$.

De aici deducem relaţia de calcul a *raportului de corelaţie* extrăgând rădăcina pătrată din coeficient de determinaţie:

$$R_{y/x} = \sqrt{1 - \frac{\sigma_{y/r}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\frac{\sum (y_i - y_{x_i})^2}{n}}{\frac{\sum (y_i - \bar{y})^2}{n}}} = \sqrt{1 - \frac{\sum (y_i - y_{x_i})^2}{\sum (y_i - \bar{y})^2}}$$

Din relaţia de calcul observăm că valorile raportului de corelaţie sunt totdeauna pozitive şi cuprinse între 0 şi 1.

În cazul legăturilor de tip liniar raportul de corelaţie trebuie să fie egal cu coeficientul de corelaţie.

Relaţia de mai sus se utilizează pentru un volum mic de date negrupate. În situaţia unei grupări simple în care *x* şi respectiv *y* au frecvenţe egale vom avea:

$$R_{y/x_i} = \sqrt{1 - \frac{\frac{\sum (y_i - y_{x_i})^2 \cdot f_i}{n}}{\frac{\sum (y_i - \bar{y})^2 \cdot f_i}{n}}},$$

iar în situaţia unei distribuţii bidimensionale

$$R_{y/x} = \sqrt{1 - \frac{\sum_i \sum_j (y_j - y_{x_i})^2 \cdot f_{ij}}{\sum_j (y_i - \bar{y})^2 \cdot f_i}}.$$

Atunci când este analizată modificarea variabilei dependente *y* în raport de variaţia mai multor factori de influenţă intensitatea legăturii se măsoară cu ajutorul *raportului de corelaţie multiplă* determinat cu relaţia:

$$R_{y/x_1,x_2,...,x_n} = \sqrt{1 - \frac{\sum\left(y_i - y_{x_1,x_2,...x_n}\right)^2}{\sum\left(y_i - \bar{y}\right)^2}}$$

Coeficientul de corelaţie multiplă

Coeficientul de corelaţie multiplă, simbolizat cu R_{y,x_i} sau simplu R, măsoară intensitatea legăturii dintre variabila dependentă y şi două sau mai multe variabile independente x_i. În acest caz, trebuie să se ţină seama de faptul că fiecărei variabile independente îi revine numai o fracţiune din influenţa totală asupra variabilei dependente. Relaţia de calcul este următoarea:

$R_{y,x_i} = R = \sqrt{1 - \frac{\sum\left(y_i - y_{x_i}\right)^2}{\sum\left(y_i - \bar{y}\right)^2}}$, unde $\bar{y}$ reprezintă media aritmetică simplă a valorilor empirice ale variabilei dependente pe perioada de analiză statistică, sau, dacă se ţine seama de funcţia de corelaţie multiplă, relaţia devine:

$$R = \sqrt{\frac{a_0\sum y + a_1\sum y_{x_1} + a_2\sum y_{x_2} + ... + a_n\sum y_{x_n} - \frac{\left(\sum y\right)^2}{n}}{\sum y^2 - \frac{\left(\sum y\right)^2}{n}}}.$$

Acest indicator are întotdeauna valoare pozitivă şi este mai mare decât oricare coeficient de corelaţie simplă dintre variabila dependentă şi cele independente, luat în valoare absolută. Pătratul coeficientului de corelaţie multiplă este cunoscut în literatura de specialitate sub denumirea de coeficient de determinaţie multiplă (R^2). Acesta exprimă ponderea cu care variabilele independente influenţează concomitent asupra variabilei dependente. Ponderea influenţei celorlalţi factori, neincluşi în model, se calculează ca diferenţă între unitate şi R^2, adică $1\text{-}R^2$.

Coeficientul de corelaţie multiplă în cazul a doi factori de influenţă în ipoteza unei legături liniare de forma:

$y = a + b_1x_1 + b_2x_2$

se poate determina cu relaţia:

$$R_{y/x_1,x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

unde: r_{yx_1} - coeficientul de corelaţie simplă dintre y şi x_1;

r_{yx_2} - coeficientul de corelaţie simplă dintre y şi x_2;

$r_{x_1x_2}$ - coeficientul de corelaţie simplă dintre x_1 şi x_2.

Pornind de la coeficienţii de corelaţie simplă se pot calcula coeficienţii de corelaţie parţială dintre variabila dependentă y şi un factor de influenţă cu excluderea influenţei celuilalt factor pe baza relaţiilor:

$$R_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}$$

$$R_{yx_2/x_1} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}}$$

unde: R_{yx_1/x_2} – coeficientul de corelaţie parţială dintre variabila dependentă y şi variabila independentă x_1 cu excluderea influenţei lui x_2;

R_{yx_2/x_1} – coeficientul de corelaţie parţială dintre variabila dependentă y şi variabila independentă x_2 cu excluderea influenţei lui x_1.

Valorile coeficienţilor de corelaţie parţială sunt mai mici decât valoarea coeficientului de corelaţie multiplă dar mai mari decât valoarea coeficientului de corelaţie simplă.

Testarea semnificaţiei raportului de corelaţie calculat se poate face cu ajutorul testului „F" de analiză dispersională.

Mărimea lui F_{calc} se determină cu ajutorul relaţiei:

$$F_{calc} = \frac{\frac{\sum(y_x - y)^2}{r-1}}{\frac{\sum(y_i - y_x)^2}{n-r}}$$

unde: n – numărul unităţilor statistice;

r – numărul grupelor.

Se compară valoarea lui F_{calc} cu F_{tab} identificată funcţie de nivelul de semnificaţie ales α (probabilitatea cu care se doreşte obţinerea rezultatului) şi de numărul gradelor de libertate $f_1=r-1$ şi $f_2=n-r$.

Dacă $F_{calc} \geq F_{tab}$ – raportul de corelaţie este semnificativ.

Dacă $F_{calc} \leq F_{tab}$ – raportul de corelaţie este respins.

Exemplul 4.5.

Regresia multiplă cercetează şi stabileşte existenţa unei legături liniare între o variabilă dependentă (y) şi mai multe variabile explicative ($x_1, x_2, \ldots x_n$), de tipul:

$y = a + b_1x_1 + b_2x_2 + \ldots b_nx_n + e$

unde: a = media variabilei y când $x_1=x_2= \ldots =x_n=0$

$b_1, b_2 \ldots b_n$= variaţia (±) a lui y când $x_1, x_2 \ldots x_n$ variază cu o unitate

e = eroarea care rezultă independent de $x_1, x_2 \ldots x_n$ şi care are o distribuţie normală.

Modul de lucru se prezintă pe baza exemplului următor care este o extensie a exemplului precedent.

Pentru ultimele 5 luni se cunosc datele statistice privind preţurile unitare (x_1) bugetul de publicitate alocat (x_2) şi cantitatea corespunzătoare (y) de produse vândute (vezi tabelul 4.10., col.1,2 şi 3). Să se studieze relaţia dintre cele trei variabile şi să se estimeze vânzările pentru $x_1=9,5$ şi $x_2=60$ cu o probabilitate de 95 %.

Tabelul 4.10.

Luna	x_1	x_2	y	x_1^2	$x_1 \cdot x_2$	x_2^2	$x_1 \cdot y$	$x_2 \cdot y$	$(y-\bar{y})(x_1-\bar{x}_1)$	$(y-\bar{y})(x_2-\bar{x}_2)$
0	1	2	3	4	5	6	7	8	9	10
1	13	4	25	169	52	16	325	100	-21	154
2	12	6	30	144	72	36	360	180	-4,5	81
3	11,5	15	45	132,25	172,5	225	517,5	675	0	0
4	11	20	45	121	220	400	495	900	-3	30
5	10	30	50	100	300	900	500	1550	-16,5	165
Total	57,5	75	195	666,25	816,5	1577	2197,5	3.355	-45	430
Media	11,5	15	39	-	-	-	-	-	-	-

Pentru rezolvarea acestei aplicaţii vom parcurge următoarele etape:

1) Ajustarea datelor empirice printr-o dreaptă de regresie

$$y_r = a + b_1 x_1 + b_2 x_2$$

Aplicându-se metoda celor mai mici pătrate, se demonstrează că parametrii a, b_1, b_2,...,b_n se determină rezolvând sistemul de ecuaţii normale corespunzător, conform modelului următor:

$$\begin{cases} a\cdot n + b_1\Sigma x_1 + b_2\Sigma x_2 + b_3\Sigma x_3 + \ldots + b_m\Sigma x_m = \Sigma y \\ a\Sigma x_1 + b_1\Sigma x_1 x_1 + b_2\Sigma x_1 x_2 + b_3\Sigma x_1 x_3 + \ldots + b_m\Sigma x_1 x_m = \Sigma y x_1 \\ a\Sigma x_2 + b_1\Sigma x_2 x_1 + b_2\Sigma x_2 x_2 + b_3\Sigma x_2 x_3 + \ldots + b_m\Sigma x_2 x_m = \Sigma y x_2 \\ a\Sigma x_3 + b_1\Sigma x_3 x_1 + b_2\Sigma x_3 x_2 + b_3\Sigma x_3 x_3 + \ldots + b_m\Sigma x_3 x_m = \Sigma y x_3 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a\Sigma x_m + b_1\Sigma x_m x_1 + b_2\Sigma x_m x_2 + b_3\Sigma x_m x_3 + \ldots + b_m\Sigma x_m x_m = \Sigma y x_m \end{cases}$$

unde: n = numărul datelor empirice, iar m = numărul necunoscutelor (b_1, b_2, b_3,...,b_m)

Notă: $x_1, x_2, \ldots, x_m$ pot fi şi $\frac{1}{x}$,x^2,$\log x$,....

În cazul considerat, sistemul cu trei necunoscute este următorul:

$$\begin{cases} a\cdot n + b_1\Sigma x_1 + b_2\Sigma x_2 = \Sigma y \\ a\Sigma x_1 + b_1\Sigma x_1 x_1 + b_2\Sigma x_1 x_2 = \Sigma y x_1 \\ a\Sigma x_2 + b_1\Sigma x_2 x_1 + b_2\Sigma x_2 x_2 = \Sigma y x_2 \end{cases}$$

Cu datele concrete, acesta are forma (vezi şi tabelul 4.10., col. 1-8)

$$\begin{cases} 5a + 57{,}5b_1 + 75b_2 = 195 \\ 57{,}5a + 666{,}25b_1 + 816{,}5b_2 = 2197{,}5 \\ 75a + 816{,}5b_1 + 1577b_2 = 3355 \end{cases} \Rightarrow \begin{array}{l} a = 75{,}388... \\ b_1 = -3{,}888... \\ b_2 = 0{,}555... \end{array}$$

Rezolvând sistemul, ecuaţia de regresie căutată are forma:

$y_r = 75{,}388 - 3{,}888x_1 + 0{,}555x_2$

Verificare: $\bar{y} = a + b_1\bar{x}_1 + b_2\bar{x}_2$

$$39 = 75{,}388 - 3{,}888 \cdot 11{,}5 + 0{,}555 \cdot 15$$

2) Se calculează abaterile dintre y şi $y_r = 75{,}3888 - 3{,}888x_1 + 0{,}555x_2$

Pentru uşurinţa calculelor datele sunt sintetizate în tabelul 4.11.

Tabelul 4.11.

Luna	x_1	x_2	y	y_r	$(y - \bar{y})^2$	$(y_r - \bar{y}_r)^2$	$(y-y_r)^2$
1	13	4	25	27,055..	196	142,67	4,23
2	12	6	30	22,055..	81	48,23	4,23
3	11,5	15	45	39	36	0	36
4	11	20	45	43,722..	36	22,30	1,63
5	10	30	50	53,166..	121	200,68	10,02
Total	57,5	75	195	195	470	413,88..	56,11..
Medie	11,5	15	39	39	-	-	-

- variaţia totală a datelor y = $\sum(y-\bar{y})^2$ = 470 (vezi col.5, tabelul 4.11.)
- variaţia totală a datelor y_r = $\sum(y_r-\bar{y}_r)^2$ =413,88 (vezi col.6, tabelul 4.11.)
- variaţia dintre y şi y_r = $\sum(y-y_r)^2$ =56,11 (vezi col. 7, tabelul 4.11.)

Interpretare: din cele 470 de unităţi ale variaţiei totale, 413,88 sunt explicate de funcţia y_r (adică de x_t şi x_2), iar restul de 56,11 rămân neexplicate.

La rândul său, variaţia explicată se poate descompune pe cei doi factori, astfel:

- variaţia explicată de $x_1 = b_1\sum(y-\bar{y})(x_1-\bar{x}_1)$
- variaţia explicată de $x_2 = b_2\sum(y-\bar{y})(x_2-\bar{x}_2)$

Cu datele problemei avem:

- $b_1\sum(y-\bar{y})(x_1-\bar{x}_1)$ = *(-3,888)(-45) = 175* (vezi şi tabelul 4.10., col.9)
- $b_2\sum(y-\bar{y})(x_2-\bar{x}_2) = (0{,}555)(430) = \frac{238{,}88}{413{,}88}$ (vezi şi tabelul 4.10., col. 10)

3) Se calculează eroarea standard (e)

În acest scop se utilizează relaţia (*) în care υ = *n-3* deoarece, în acest caz se cunosc trei parametri (*a, b_1* şi *b_2*)

$$e = \sqrt{\frac{\sum(y-y_r)^2}{n-3}} = \sqrt{\frac{56{,}11}{5-3}} = 5{,}3$$

4) Se compară R^2 calculat cu R^2 teoretic

$$R^2 = \frac{\sum(y_r-\bar{y}_r)^2}{\sum(y-\bar{y})^2} = \frac{413{,}88}{470} = 0{,}8806146$$

Pentru α = 0,05, R^2 minim trebuie să fie egal cu 0,9025 pentru υ = n-3 = 5-3 = 2 grade de libertate. Deoarece R^2 observat < R^2 teoretic deducem că funcţia y_r *= 75,388 - 3, 888x_1 + 0,555x_2* nu poate fi socotită acceptabilă decât dacă ne mulţumim cu o probabilitate a previziunii < 95 %.

5) Se estimează vânzările.

Ca şi în cazul regresiei simple,

$P = y_r \pm t\cdot e = (a + b_1x_1 + b_2x_2) + t\cdot e$

Pentru cazul dat, unde x_1 = 9,5, x_2 =60 şi t = 2,78,

$P = (75{,}388 - 3{,}888 \cdot 9{,}5 + 0{,}555 \cdot 60) \pm 2{,}78 \cdot 5{,}3 = 7\,1{,}777 \pm 14{,}73$, respectiv $57{,}65 < y_r < 86{,}51$

Concluzie: în cazul practicării unui preţ x_1 = *9,5* şi a alocării uni buget publicitar x_2 *=60,* volumul vânzărilor se va situa între 57 şi 87 unităţi, dar cu o probabilitate < 95 %.

Exemplul 4.6.

O întreprindere a observat pe un eşantion de 10 magazine că cifra de afaceri variază funcţie de suprafaţa magazinului şi numărul de salariaţi, conform datelor din tabelul 4.12.

Tabelul 4.12.

Suprafaţa magazinului (m^2)	100	600	600	700	700	500	800	300	200	200
Numărul de salariaţi	24	28	20	23	26	21	28	28	20	25
Cifra de afaceri a magazinului (mil \$)	11	23	20	21	21	13	30	18	7	18

Întreprinderea doreşte să construiască un magazin cu o suprafaţă de 600 m^2 şi un număr de 24 salariaţi. Care va fi cifra de afaceri previzibilă a acestui magazin?

Pentru rezolvarea acestei probleme vom folosi un model liniar de regresie multiplă de forma:

$y = a + b_1x_1 + b_2x_2$, unde:

y = cifra de afaceri a magazinului (în mil. \$)

x_1 = suprafaţa magazinului (m^2)

x_2 = numărul de salariaţi

Aplicând metoda celor mai mici pătrate

$minS = \sum_{i=1}^{n}\left(y_i - a - b_1x_1 - b_2x_2\right)^2$ şi anulând derivatele parţiale în raport cu a, b_1 şi b_2 vom obţine sistemul de ecuaţii normale, care va permite calculul parametrilor.

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{10}\left(y_i - a - b_1x_1 - b_2x_2\right)\cdot 1 = 0$$

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{10}\left(y_i - a - b_1x_1 - b_2x_2\right)\cdot x_1 = 0$$

$$\frac{\partial S}{\partial b_2} = -2\sum_{i=1}^{10}\left(y_i - a - b_1x_1 - b_2x_2\right)\cdot x_2 = 0$$

de unde:

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 = \sum y \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 = \sum x_1y \\ a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 = \sum x_2y \end{cases}$$

Pentru rezolvarea sistemului vom organiza datele ca în tabelul 4.13.

Tabelul 4.13.

Număr magazin	y (mil. lei)	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y	y^2
1	11	100	24	10000	576	2400	1100	264	121
2	23	600	28	360000	784	16800	13800	644	529
3	20	600	20	360000	400	12000	12000	400	400
4	21	700	23	490000	529	16100	14700	483	441
5	21	700	26	490000	676	18200	14700	546	441
6	13	500	21	250000	441	10500	6500	273	169
7	30	800	28	640000	784	22400	24000	840	900
8	18	300	28	90000	784	8400	5400	504	324
9	7	200	20	40000	400	4000	1400	140	49
10	18	200	25	40000	625	5000	3600	450	324
Σ	182	4700	243	2770000	5999	115800	97200	4544	3698

Rezultă sistemul:

$$\begin{cases} 10a + 4700b_1 + 243b_2 = 182 \\ 4700a + 2770000b_1 + 115800b_2 = 97200 \\ 243a + 115800b_1 + 5999b_2 = 4544 \end{cases}$$

Prin rezolvarea sistemului obţinem:

$a = -14,21$

$b_1 = 0,0179$

$b_2 = 0,99$

de unde ecuaţia modelului liniar de corelaţie:

$y = -14,21 + 0,0179x_1 + 0,99x_2$

Fiind vorba de o corelaţie multiplă liniară vom calcula coeficientul de corelaţie cu relaţia:

$$R_{y/x_1,x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

unde: r_{yx_1} = coeficientul de corelaţie simplă dintre y şi x_1

r_{yx_2} = coeficientul de corelaţie simplă dintre y şi x_2

$r_{x_1x_2}$ = coeficientul de corelaţie simplă dintre x_1 şi x_2

$$r_{yx_1} = \frac{n\sum x_1y - \sum x_1 \cdot \sum y}{\sqrt{\left[n\sum x_1^2 - \left(\sum x_1\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} =$$

$$= \frac{10 \cdot 97200 - 4700 \cdot 182}{\sqrt{[10 \cdot 2770000 - 2209000][10 \cdot 3698 - 33124]}} =$$

$$= \frac{116600}{\sqrt{5610000 \cdot 3856}} = \frac{116600}{147079} = 0,793$$

$$r_{yx_2} = \frac{n\sum x_2y - \sum x_2 \cdot \sum y}{\sqrt{\left[n\sum x_2^2 - \left(\sum x_2\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} =$$

$$= \frac{10 \cdot 4544 - 243 \cdot 182}{\sqrt{[10 \cdot 5999 - 59049][10 \cdot 3698 - 33124]}} =$$

$$= \frac{1214}{\sqrt{941 \cdot 3856}} = \frac{1214}{1905} = 0{,}637$$

$$r_{x_1x_2} = \frac{n\sum x_1x_2 - \sum x_1 \cdot \sum x_2}{\sqrt{\left[n\sum x_1^{\,2} - \left(\sum x_1\right)^2\right]\left[n\sum x_2^{\,2} - \left(\sum x_2\right)^2\right]}} =$$

$$= \frac{10 \cdot 115800 - 4700 \cdot 243}{\sqrt{5610000 \cdot 941}} = \frac{1158000 - 1142100}{72657} = 0{,}219$$

Rezultă

$$R_{y/x_1,x_2} = \sqrt{\frac{0{,}628849 + 0{,}405769 - 2 \cdot 0{,}793 \cdot 0{,}637 - 0{,}219}{1 - 0{,}047961}} =$$

$$= \sqrt{\frac{1{,}034618 - 0{,}221252}{0{,}952039}} = 0{,}9243$$

Rezultă că variaţia volumului cifrei de afaceri se datorează în proporţie de circa 92,43% suprafeţei magazinului şi numărului de salariaţi.

Calculând coeficienţii de corelaţie parţială cu excluderea influenţei celuilalt factor rezultă:

$$R_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_2}^2\right)\left(1 - r_{x_1x_2}^2\right)}} = \frac{0{,}793 - 0{,}637 \cdot 0{,}219}{\sqrt{\left(1 - 0{,}637^2\right)\left(1 - 0{,}219^2\right)}} = 0{,}869$$

$$R_{yx_2/x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_1}^2\right)\left(1 - r_{x_1x_2}^2\right)}} = \frac{0{,}637 - 0{,}793 \cdot 0{,}219}{\sqrt{\left(1 - 0{,}793^2\right)\left(1 - 0{,}219^2\right)}} = 0{,}779$$

În consecinţă construirea unui magazin cu o suprafaţă de 600 m^2 şi 24 salariaţi va conduce la obţinerea de către acesta a unei cifre de afaceri

Y = -14,21 + 0,0179 · 600 + 0,99 · 24 = 20,29 mil. $

4.2.3. Metode neparametrice de măsurare a legăturilor

Sunt metode care se utilizează atunci când:

- nu se cunoaşte forma legăturii;
- caracteristicile sunt calitative şi nu se pot exprima numeric dar este posibilă ierarhizarea lor;
- în cazul distribuţiilor asimetrice;
- când dispunem de un număr mic de observaţii.

Cele mai utilizate metode neparametrice sunt:

- metoda tabelului de asociere şi a coeficientului de asociere;
- metoda corelaţiei rangurilor.

A. Metoda tabelului de asociere se utilizează în cazul caracteristicilor alternative care admit numai două forme de manifestare sau valori. Tabelul de asociere este de fapt un caz particular al tabelului cu dublă intrare. Pe linii se înscrie variaţia caracteristicii factoriale *x* (variabila independentă), iar pe coloane variaţia caracteristicii rezultative *y*

(variabila efect). La intersecţia liniilor cu coloanele se trec frecvenţele cu care unităţile colectivităţii se înscriu în cele patru grupe formate prin intersecţia variaţiei caracteristicilor *x* şi *y* (tabelul 4.14.).

Tabelul 4.14.

Y / *X*	y_1	y_2	*Total*
x_1	*a*	*b*	*a+b*
x_2	*c*	*d*	*c+d*
Total	*a+c*	*b+d*	*a+b+c+d*

Pentru măsurarea intensităţii legăturii dintre *x* şi *y* se utilizează *coeficientul de asociere* propus de Yulle:

$$Q = \frac{ad - bc}{ad + bc}.$$

Valorile acestui indicator sunt cuprinse între –1 şi +1.

Dacă $Q>0$ – asociere directă.

Dacă $Q<0$ – asociere inversă.

Dacă $Q=0$ – nu există asociere.

Dacă *Q* tinde la ±1 avem asociere foarte puternică.

Dacă o frevenţă din cele patru (*a, b, c, d*) este nulă, atunci asocierea este completă şi $Q=\pm 1$.

B. Metoda corelaţiei rangurilor

După cum s-a văzut până aici, în calculul coeficientului de corelaţie liniară sau a raportului de corelaţie, se folosesc valorile variabilelor care alcătuiesc cuplul corelativ fapt care evidenţiază oarecum că nivelul acestor indicatori depinde de nivelul variabilelor pentru care se realizează studiul.

O posibilă soluţie pentru eliminarea acestei dependenţe este utilizarea unor metode de studiu a corelaţiei care nu folosesc direct în calculul respectivilor coeficienţi valorile variabilelor din cuplul corelativ.

Corelaţia rangurilor este o metodă de studiu a corelaţiei care presupune folosirea în calculul indicatorilor pentru măsurarea intensităţii corelaţiei, a rangurilor corespunzătoare variabilelor din cuplul corelativ.

Rang – poziţia pe care o ocupă valorile variabilelor X şi Y din cuplul corelativ în cadrul şirului din care fac parte, ordonat crescător sau descrescător.

Pentru ranguri se folosesc următoarele notaţii:

- u_i - rangurile valorilor xi, din cadrul şirului ordonat $x_1, x_2, \ldots, x_n$;
- w_i - rangurile valorilor yi, din cadrul şirului ordonat $y_1, y_2, \ldots, y_n$.

Coeficientului lui Spearman

Folosind în relaţia de calcul a coeficientului de corelaţie liniară (*r*) rangurile corespunzătoare, în locul valorilor variabilelor *X* respectiv *Y*, utilizând o serie de proprietăţi a seriilor de ranguri şi după câteva calcule elementare se ajunge la:

$$\theta = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

unde: d - diferenţa dintre rangurile u_i şi w_i;
n - numărul termenilor seriei.

Relaţia dă rezultate corecte atâta timp cât sunt îndeplinite premisele folosite la obţinerea ei, şi anume:

1. $\sum u_i = \sum w_i$;
2. $\bar{u}_i = \frac{\sum u_i}{n} = \frac{\sum w_i}{n} = \frac{n+1}{2} = \bar{w}_i$;

3. rangurile u_i şi w_i, sunt unice, nu se repetă în cadrul şirului din care fac parte. Dacă această ultimă condiţie nu este îndeplinită atunci se poate proceda astfel: valoarea lui x_i sau y_i care se repetă se va trece o singură dată în şirul din care face parte, iar ca valoare corespondentă se va trece media valorilor celeilalte variabile, corespunzătoare valorii care se repeta.

Coeficientul lui Spearman ia valori în intervalul **[-1;1]**.

Semnificaţia acestui coeficient este similară cu cea a coeficientului de corelaţie liniară (r).

Coeficientul lui Kendall

Presupune îndeplinirea aceloraşi condiţii ca şi coeficientul lui Spearman şi se poate calcula cu ajutorul relaţiei:

$$\tau = \frac{2S}{n(n-1)} = \frac{2(P-Q)}{n(n-1)},$$

unde: P - suma rangurilor w_i mai mari decât rangul curent.
Q - suma rangurilor w_i mai mici decât rangul curent.

Pentru determinarea lui P şi lui Q se procedează astfel:

- se ordonează crescător perechile de valori (x_i, y_i);
- se elimină eventualele repetiţii de valori pentru variabila X şi variabila Y şi dacă este cazul, se reordonează crescător perechile de valori (x_i, y_i);
- se determină rangurile u_i ataşate valorilor variabilei X;
- se determină rangurile w_i ataşate valorilor variabilei Y;
- pornind de la prima valoare spre sfârşitul şirului rangurilor w_i, se determină succesiv pentru fiecare rang:
  - câte ranguri w_i sunt mai mari de cât rangul curent;
  - câte ranguri w_i sunt mai mici de cât rangul curent;

Notă: Numărul rangurilor mai mari sau mai mici decât rangul curent se face pornind numărătoarea de la următoarea poziţie faţă de rangul curent, către sfârşitul şirului.

- se determină P şi Q.

Şi coeficientul lui Kendall $\in$ **[-1;1]** şi are o interpretare similară cu cea de la coeficientul lui Spearman.

Dacă pentru acelaşi set de date se calculează ambii coeficienţi se va observa că:
$\theta > \tau$.

Exemplul 4.3.

Presupunând că avem un set de perechi de valori (x_i, y_i) corespunzătoare a două variabile X şi Y, pentru care se doreşte evidenţierea existenţei unei legături cauzale, procedăm astfel (tabelul 4.15.):

Tabelul 4.15.

Iniţial		**După eliminare repetiţii**		**După ordonare**	
x_i	y_i	x_i	y_i	x_i	y_i
100	28	100	27	51	9
98	24	98	24	54	8
51	9	51	9	69,5	12
100	26	86	14	86	14
86	14	69,5	12	98	24
70	12	54	8	100	27
69	12	108	31	102	35
54	8	102	35	108	31
108	31				
102	35				
n = 10		n = 8		n = 8	

Pentru calculul celor doi coeficienţi determinăm mai întâi rangurile u_i şi w_i (tabelul 4.16.).

Tabelul 4.16.

Valori ordonate		**Ranguri**		**Indicatori**			
x_i	y_i	u_i	w_i	d	d^2	P	Q
51	9	1	2	-1	1	6	1
54	8	2	1	1	1	6	0
69,5	12	3	3	0	0	5	0
86	14	4	4	0	0	4	0
98	24	5	5	0	0	3	0
100	27	6	6	0	0	2	0
102	35	7	8	-1	1	0	1
108	31	8	7	1	1	0	0
n = 10		36	36	0	4	26	2

$$\theta = 1 - \frac{6 \cdot 4}{8\left(8^2 - 1\right)} = 0{,}994$$

$$\tau = \frac{2 \cdot (26 - 2)}{8(8 - 1)} = 0{,}857$$

Între cele două variabile există o corelaţie directă foarte intensă.

Tot în studiul corelaţiei rangurilor poate fi utilizată şi corelograma rangurilor.

Modul de construcţie al acesteia este similar cu cel al corelogramei prezentate la metodele elementare de studiu a corelaţiei cu deosebirea că la corelograma rangurilor se reprezintă perechile de ranguri (u_i, w_i).

Diagonalele reţelei reprezintă corelaţie maximă directă respectiv inversă. Interpretarea corelogramei rangurilor este similară cu cea a corelogramei prezentate la metode elementare în studiul corelaţiei.

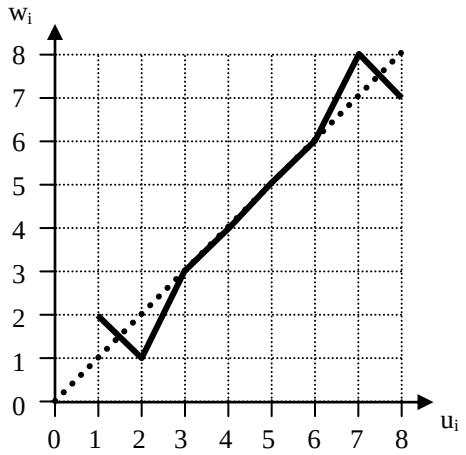


Figura 4.13. Corelograma rangurilor

Exemplul 4.4.

Exporturile (FOB) şi importurile (CIF) de mărfuri ale României în / din ţările Uniunii Europene în anul 2006 au fost (tabelul 4.17.).

Tabelul 4.17.

Nr. crt.	*Ţara*	*Export (x)*	*Import (y)*	*Rang după*		*Diferenţa de rang*
				Export (x)	*Import (y)*	
1.	Austria	685,2	1535,3	20	22	-2
2.	Belgia	426,3	626,5	15	15	0
3.	Cehia	294,5	1007,4	14	19	-5
4.	Cipru	50,9	15,2	8	4	4
5.	Danemarca	63,6	164,9	9	8	1
6.	Estonia	6,3	6,7	3	3	0
7.	Franţa	1938,3	2664,7	23	23	0
8.	Finlanda	31,2	185,1	5	9	-4
9.	Germania	4060,2	6176,8	24	25	-1
10.	Grecia	507,5	481,4	17	14	3
11.	Irlanda	42,1	240,4	7	11	-4
12.	Italia	4637,2	5954,9	25	24	1
13.	Letonia	5,8	3,8	2	2	0
14.	Lituania	15,5	23,8	4	5	-1
15.	Luxemburg	5,4	29,9	1	6	-5
16.	Malta	76,6	3,7	10	1	9
17.	Olanda	637,4	739,6	19	16	3
18.	Polonia	468,3	1128,4	16	18	-2
19.	Portugalia	33,5	128,7	6	7	-1

Nr. crt.	*Ţara*	*Export (x)*	*Import (y)*	*Rang după*		*Diferenţa de rang*
				Export (x)	*Import (y)*	
20.	Spania	605,8	953,1	18	17	-1
21.	Suedia	142,3	419,4	12	12	0
22.	Regat Unit M. Britanii	1216,8	1011,7	21	20	1
23.	Slovacia	180,4	461,7	13	13	0
24.	Slovenia	92,6	192,0	11	10	1
25.	Ungaria	1276,0	1331,6	22	21	1
	Total	**17499,7**	**25486,7**	-	-	-

Sursa: *Buletin statistic lunar*, nr.12, 2006, Institul Naţional de Statistică.

Să se analizeze legătura dintre cele două variabile.

Rezolvare

$$\theta = 1 - \frac{6 \cdot 223}{25\left(25^2 - 1\right)} = 0{,}914$$

Pentru calculul coeficientului lui Kendall (tabelul 4.18.) se ordonează crescător ţările UE după variabila *x* (export) înscriind în coloana alăturată rnagurile corespunzătoare după *y* (import). Se determină apoi pentru fiecare ţară, pe baza rangurilor *y* (import):

- P_i – numărul de ţări (de la rândul *i* până la sfârşitul seriei) având la importuri ranguri superioare rangului ţării *i;*
- Q_i – numărul de ţări (de la rândul *i* până la sfârşitul seriei) având la importuri ranguri inferioare rangului ţării *i.*

Tabelul 4.18.

Nr. crt.	*Ţara*	*Export (x)*	*Import (y)*	*Rang după*		P_i	Q_i	$P_i - Q_i$
				x	*y*			
1.	Austria	5,4	29,9	1	6	19	5	14
2.	Belgia	5,8	3,8	2	2	22	1	21
3.	Cehia	6,3	6,7	3	3	21	1	20
4.	Cipru	15,5	23,8	4	5	19	2	17
5.	Danemarca	31,2	185,1	5	9	16	4	12
6.	Estonia	33,5	128,7	6	7	17	2	15
7.	Franţa	42,1	240,4	7	11	14	4	10
8.	Finlanda	50,9	15,2	8	4	16	1	15
9.	Germania	63,6	164,9	9	8	15	1	14
10.	Grecia	76,6	3,7	10	1	15	0	15
11.	Irlanda	92,6	192	11	10	14	0	14
12.	Italia	142,3	419,4	12	12	13	0	13
13.	Letonia	180,4	461,7	13	13	12	0	12
14.	Lituania	294,5	1007,4	14	19	6	4	2
15.	Luxemburg	426,3	626,5	15	15	9	1	8
16.	Malta	468,3	1128,4	16	18	6	3	3
17.	Olanda	507,5	481,4	17	14	8	0	8
18.	Polonia	605,8	953,1	18	17	6	1	5
19.	Portugalia	637,4	739,6	19	16	6	0	6
20.	Spania	685,2	1535,3	20	22	3	2	1

Nr. crt.	*Ţara*	*Export (x)*	*Import (y)*	*Rang după*		P_i	Q_i	$P_i - Q_i$
				x	*y*			
21.	Suedia	1216,8	1011,7	21	20	4	0	4
22.	Regat Unit M. Britanii	1276	1331,6	22	21	3	0	3
23.	Slovacia	1938,3	2664,7	23	23	2	0	2
24.	Slovenia	4060,2	6176,8	24	25	0	1	-1
25.	Ungaria	4637,2	5954,9	25	24	0	0	0
	Total	17499,7	25486,7	-	-	266	33	233

$$\tau = \frac{2 \cdot (266 - 33)}{25(25 - 1)} = 0{,}777$$

Valorile celor doi coeficienţi indică o legătură directă puternică între exportul şi importul României în / din ţările Uniunii Europene.