



**UNIVERSITATEA DE STAT DE MEDICINĂ ȘI FARMACIE „NICOLAE TESTEMIȚANU”
CATEDRA MANAGEMENT ȘI PSIHOLOGIE**

ANALIZA CORELAȚIEI ȘI REGRESIEI. APLICAȚII ÎN CERCETAREA ȘTIINȚIFICĂ.

Nicov Irina, dr. in ec., asistent universitar

Globa Nina, asistent universitar

Corelația

- În medicină și biologie, ca de altfel în toate domeniile de activitate, există o interdependență între fenomene. Apariția și evoluția unui fenomen este în strânsă legătură cu o serie de alte fenomene sau factori care intervin în determinarea sau favorizarea acestuia.
- Corelația este o metodă care ne permite să cunoaștem fenomenele din natură și societate sub raportul interferenței lor, a conexiunilor în care se găsesc.

Definiție

- 1. Relație, legătură reciprocă între două sau mai multe lucruri sau fenomene; relație în care unul dintre termeni nu poate exista fără celălalt.**
- 2. Dependență reciprocă, relație a două fenomene sau procese între variațiile cărora există o anumită legătură.**

Corelația

- În statistică, pentru studierea legăturilor multiple ce au loc între diferite fenomene, se folosește noțiunea de funcție f , care constă în faptul că fiecărei valori a variabilei independente (X), numită argument, îi corespunde valoarea altei variabile numită funcție (Y).

Tipuri de corelații

- **Corelațiile funcționale sau matematice** sunt perfecte, rigide, exprimând legătura de la cauză la efect între fenomene unde unei valori determinate a unei variabile independente X (argument) îi corespunde strict o valoare a variabilei dependente Y (funcție). (ex. Corelația dintre viteză, timp și distanță)
- **Correlatiile statistice** sunt mai puțin perfecte, se evidențiază mai greu, exprimând legătura de dependență care există între fenomene. În cazul corelației statistice fiecărei 143 valori numerice a variabilei X corespund nu una ci mai multe valori a variabilei Y , adică o totalitate statistică a acestei valori, care se grupează în jurul mediei xY (de exemplu, persoanele cu aceeași înălțime pot avea valori diferite ale greutateii, care se grupează în jurul mediei lor).

Tipuri de corelații

După formă legătura de corelație poate fi :

1. Lineară - schimbărilor uniforme a valorilor medii a unei variabile, au loc schimbări egale a altei variabile (de exemplu, schimbările tensiunii arteriale sistolice și celei diastolice).
2. Nonlineară - schimbărilor uniforme a unei valori îi corespund valori medii a altei variabile, care poartă caracter de creștere ori de micșorare.

Tipuri de corelații

- **După direcție :**

- 1. directă (pozitivă) - se stabilesc între fenomene care evoluează în același sens, în aceeași direcție. Spre exemplu, mărirea înălțimii copiilor determină mărirea greutateii lor.**

- 2. indirectă (negativă) - se stabilesc între fenomene care evoluează în sens opus. Spre exemplu, cu cât e mai mare vârsta copiilor, cu atât e mai mică mortalitatea lor.**

Corelația directă

- **Corelațiile statistice directe presupun evoluția în același sens a fenomenelor ce se corelează, dar nu cu aceeași unitate de măsură.**
- **Spre exemplu, crește nivelul de trai al populației unei colectivități de două ori, crește și rezistența organismului la îmbolnăviri dar nu în aceeași măsură, pentru că în afară de nivelul de trai, receptivitatea organismului la îmbolnăviri este determinată și de alți factori (biologici, climatici, geografici etc.).**

Corelația statistică indirectă

- Corelațiile statistice inverse presupun creșterea unui fenomen și scăderea celui cu care se corelează, dar nu în aceeași măsură.
- Exemplu: dacă într-o colectivitate am efectuat un număr dublu de vaccinări anti-gripă, numărul persoanelor ce vor contracta boala, în anul următor, va fi mai scăzut, dar nu va fi de două ori mai mic decât în anul precedent.

Coeficientul de corelație lineară Bravais - Pearson

- Coeficientul de corelație Pearson este o măsură a relației dintre două caracteristici numerice, simbolizate prin "X" și "Y".
- Coeficientul de corelație este notat cu "r", se calculează cu ajutorul formulei:

Formula:

$$r_{xy} = \pm \frac{\Sigma d_x d_y}{\sqrt{\Sigma d_x^2 \Sigma d_y^2}}$$

r_{xy} – coeficientul de corelație

$\Sigma d_x d_y$ – suma produselor dintre abaterile de la media aritmetică a valorilor frecvențelor Σxy celor două fenomene (x și y) ce se corelează;

Σdx^2 - suma pătratelor abaterilor de la media aritmetică a valorilor frecvențelor Σd^2 fenomenului x;

Σdy^2 - suma pătratelor abaterilor de la media aritmetică a valorilor frecvențelor Σd^2 fenomenului y.

Exemplu: De determinat legătura de corelație dintre timpul trecut din momentul accesului de pancreatită acută (x) și numărul complicațiilor postoperatorii (y):

Nr	Timpul x	Număr de complicații y	dx (X-9)	dy (Y-14,8)	dx ²	dy ²	dx dy
1	3	6	-6	-8,8	36	77,4	52,8
2	5	8	-4	-6,8	16	46,2	27,2
3	7	12	-2	-2,8	4	7,8	5,6
4	10	19	1	4,2	1	17,6	4,2
5	13	20	4	5,2	16	27	20,8
6	16	84	7	9,2	49	84	64,4
Total	54	89			122	260	175
Media	54/6= 9	89/6= 14,8					

$$r_{xy} = \pm \frac{\sum dx dy}{\sqrt{\sum dx^2 \sum dy^2}} = \pm \frac{175}{\sqrt{122 \times 260}} = \pm \frac{175}{178,3} = \pm 0,98$$

Interpretarea coeficientului de corelație

- Valoarea maximă a lui "r" este 1, iar minimă -1.
- Valoarea coeficientului de corelație: $-1 < r < 1$
- **Corelația pozitivă (+), când coeficientul de corelație este $0 < r < 1$: "Y" tendea să crească în amplitudine cu "X" crește;;**
- **Corelația negativă (-), când coeficientul de corelație este $-1 < r < 0$: "Y" descrește "X" crește;**

Interpretarea coeficientului de corelație

- valoarea coeficientului de corelație ± 1 denotă o corelație perfectă între fenomene;
- valoarea coeficientului de corelație cuprinsă între:
- $\pm 0,99$ și $\pm 0,70$ denotă o corelație puternică;
- $\pm 0,69$ și $\pm 0,30$ denotă o corelație medie;
- $\pm 0,0$ și $\pm 0,29$ denotă o corelație slabă ;
- valoarea coeficientului de corelație 0 denotă că legătura lineară dintre fenomene în mod practic o considerăm inexistentă.

Interpretarea semnificației corelației

Coeficientul de determinare (r^2 – r pătrat):

- Prin intermediul coeficientului de determinare se determină partea de asociere comună a factorilor care influențează cele două variabile. Indică partea din dispersia totală a unei variabile care poate fi explicată sau justificată de dispersia valorilor din cealaltă variabilă.
- Pentru exemplul anterior, pentru “ r ”=0,98, „ r^2 ” este de 0,96. Uzual coeficientul de determinare se înmulțește cu 100 și exprimarea se transformă în procente din dispersie (96%). Prin urmare, 96% din variația numărului de complicații în cazul pancreatitei este determinată de variația timpului trecut din momentul accesului.

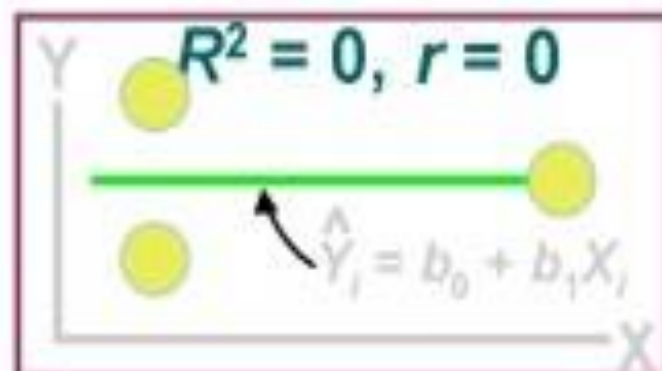
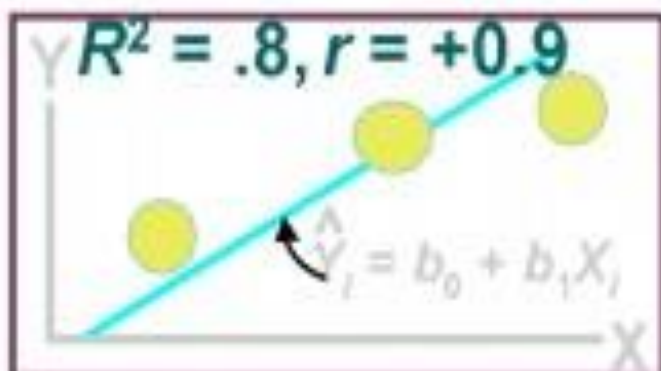
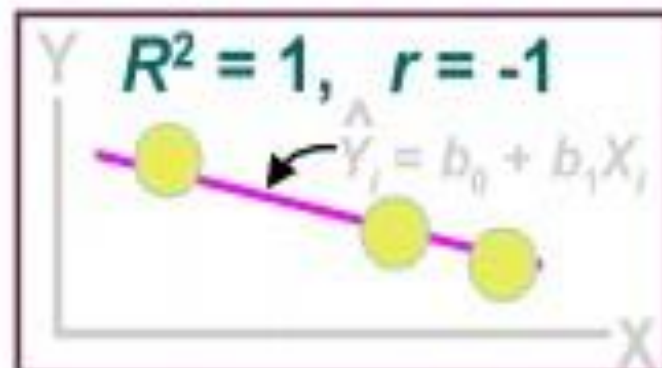
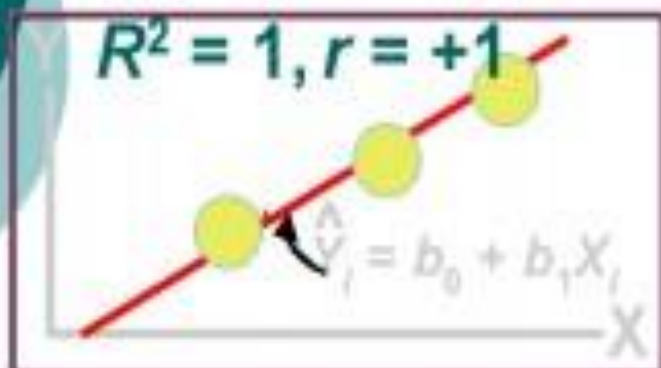
Graficul Scatter-Plot (Corelograma)

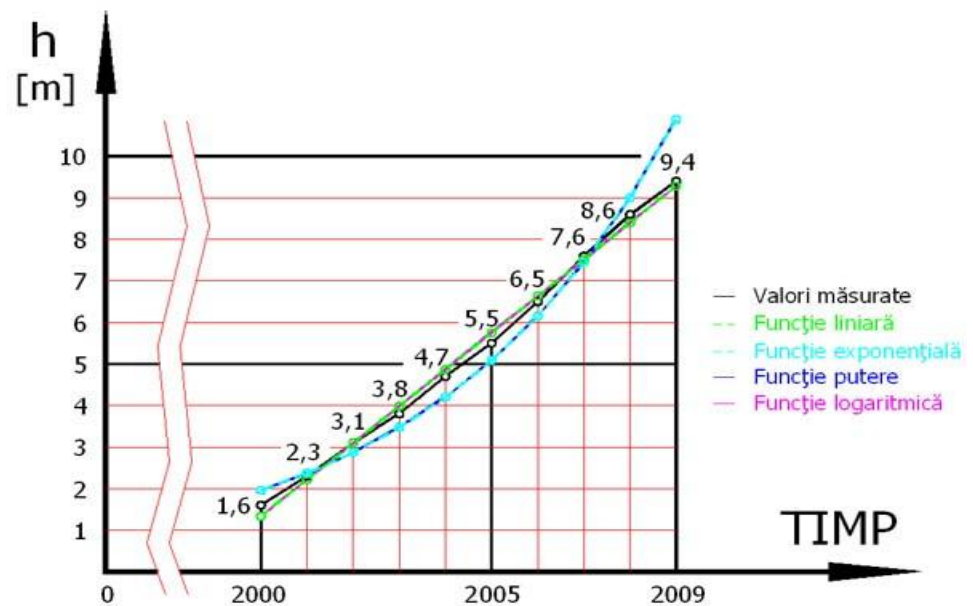
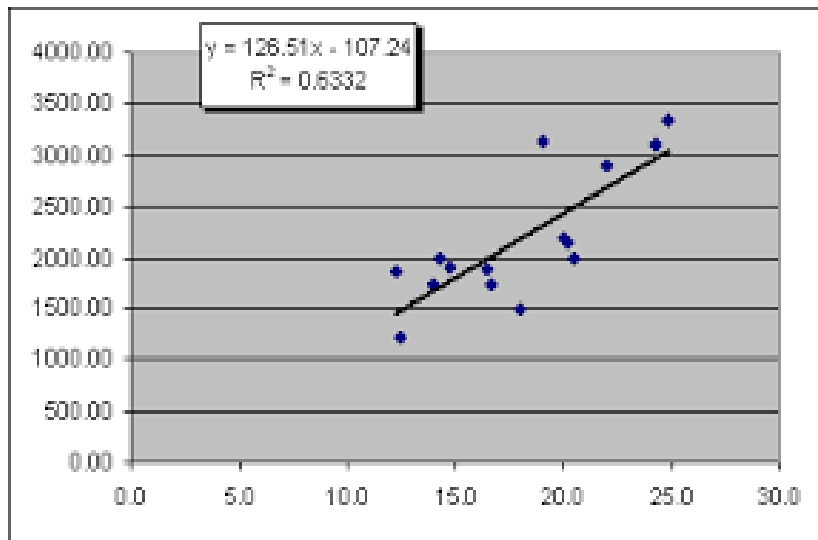
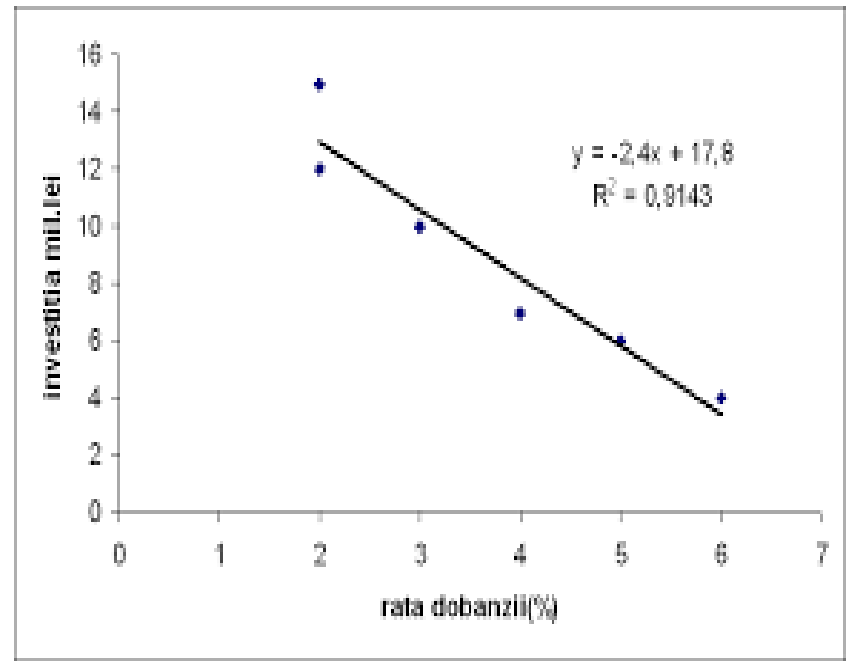
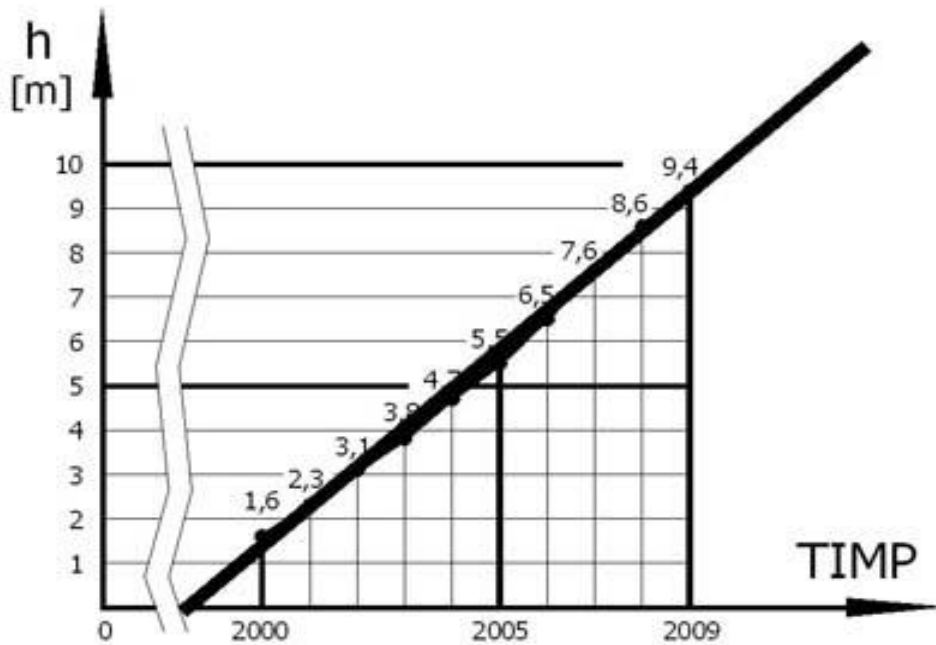
- Acest tip de grafic folosește cele două axe de coordonate, ordonată și abscisă. Pe axa x (abscisă) se înscriu valorile variabilei independente x , iar pe axa y (ordonată) – valorile variabilei dependente y .
- În rețeaua graficului este reprezentată prin câte un punct fiecare frecvență la nivelul valorii variantei corespunzătoare fenomenului x de pe abscisă și la înălțimea valorii variantei corespunzătoare fenomenului y de pe ordonată. Se realizează astfel „norul de puncte”.

Graficul Scatter-Plot (Corelograma)

- Dacă norul de puncte se va dispune fuziform, oblic de jos în sus și de la stânga la dreapta, între cele două fenomene există o corelație directă.
- Dacă norul de puncte se va difuza de sus în jos, de la dreapta la stânga, corelația este indirectă.
- Intensitatea legăturii corelative dintre cele două fenomene se apreciază, după unghiul pe care dreapta ce trece prin mijlocul norului de puncte îl realizează cu abscisa. Cu cât acest unghi ascuțit este mai mare, având tendința să se apropie de 45 grade, cu atât corelația dintre fenomene este mai puternică și invers.
- Dacă punctele se dispun pe toată rețeaua grafică, neavând nici o tendință de a se grupa, înseamnă că între fenomene nu există nici o legătură de dependență, fenomenele evoluând independent unul față de celălalt. În cazul acesta, dreapta care trece prin mijlocul punctelor este paralelă fie cu ordonata, fie cu abscisa.

Coeficientul de determinație și coeficientul de corelație liniară





Testul de semnificație pentru coeficientul de corelație:

$$t = \frac{r_{xy}}{\sqrt{1-r_{xy}^2}} \times \sqrt{n-2} = \frac{0,98}{\sqrt{1-0,96}} \times \sqrt{6-2} =$$

$$t = \pm \frac{m}{n}; \quad m = \pm \frac{1-r^2}{\sqrt{n-2}}$$

r_{xy} - coeficientul de corelație

r_{xy}^2 - coeficientul de corelație la pătrat (coef. de determinare)

n - numărul variantelor perechi care se corelează

m - valoarea erorii coeficientului obținut

1- valoarea absolută a coeficientului de corelație

- Valoarea calculată se compară cu cea tabelară, stabilită probabilistic pentru un nivel de semnificație.
- dacă **$t_{\text{calculat}} > t_{\text{tabelar}}$** coeficientul de corelație este semnificativ
- dacă **$t_{\text{calculat}} < t_{\text{tabelar}}$** , legătura este nesemnificativă și trebuie căutat un alt factor esențial cu care să se studieze corelația.

Corelația rangurilor (Spearman)

- Corelația rangului lui Spearman este folosită atunci când una sau ambele variabile relevante sunt ordinale (sau una ordonată și una numerică) și atunci când observațiile numerice sunt înclinate cu valori extreme.

$$\rho = 1 - \frac{6 \Sigma d^2}{n \times (n^2 - 1)}$$

1 = valoarea absolută a coeficientului de corelație; ;

ρ = coeficientul de corelație Spearman;

6 = valoare constantă ;

Σd^2 = suma diferențelor dintre ranguri;

n = numărul de perechi.

Interpretare

- Calculul corelației rangului Spearman, simbolizat ca r_s , implică ordonarea valorilor pentru fiecare dintre caracteristici de la mai mică la cea mai mare;
- Ca coeficient de corelație Pearson, coeficientul de corelare Spearman variază în valoare de la -1 la 1. Valorile lui r_s aproape de extreme indică un grad ridicat de corelație între X și Y; valorile de lângă 0 implică o lipsă de asociere liniară între cele două variabile.

Exemplu: Pentru 6 studenți dintr-o grupă se cunosc: calificativele pentru nivelul de pregătire al studenților la biostatistică, obținute în timpul anului și notele obținute la examenul de biostatistică:

Nr. student	Calificativ anual, x	Nota examen, y	Rang calificativ, x	Rang calificativ, y	Diferența dintre ranguri, d	Diferența ridicată la pătrat, d^2
1	foarte slab	5	1	2	-1	1
2	slab	3	2	1	1	1
3	satisfăcător	6	3	3	0	0
4	bun	8	4	4	0	0
5	foarte bun	9	5	5	0	0
6	exceptional	10	6	6	0	0

$$\rho = \pm 1 - \frac{6 \sum d^2}{n*(n^2-1)} = \pm 1 - \frac{6*2}{6*(6^2-1)} = \pm 1 - \frac{12}{6(36-1)} = \pm 1 - 0,057 = \pm 0,94$$

Interpretare: între calificativul anual și nota la examen există o legătură de corelație puternică.

Testul de semnificație pentru
coeficientul de corelație Spearman:

$$t = \frac{\rho}{\sqrt{\frac{1-\rho^2}{n-2}}}$$

Coeficientul de asociere

- Presupune întocmirea unui tabel de asociere, care prezintă colectivitatea după două caracteristici corelate logic, care au caracter alternativ (prezența fenomenului – lipsa fenomenului).

Tabelul de asociere

- Tabelul de asociere (2x2) a variabilelor (x,y):

x/y	Y1	Y2	Total
X1	a	b	a+b
X2	c	d	c+d
Total	a+c	b+d	a+b+c+d

$$q = \frac{ad - cb}{ad + cb}$$

Exemplu: Vaccinarea contra unei maladii și morbiditatea populației unei localități.

x/y	Bolnavi	Sănătoși	Total
Vaccinați	40	5900	5940
Nevaccinați	250	9620	9870
Total	290	15520	15810

$$q = \frac{ad - cb}{ad + cb} = \frac{(40 * 9620) - (250 * 5900)}{(40 * 9620) + (250 * 5900)} = \frac{-1090200}{1859800} = -0,58$$

Concluzie: între vaccinare și morbiditate este legătură inversă, astfel cu cât mai multe persoane sunt vaccinate, cu atât mai puține cazuri de îmbolnăvire se înregistrează.

Regresia. Noțiuni generale

- Coeficientul de regresie, stabilește cu cât crește sau descrește sub aspect cantitativ, un fenomen, când cel cu care se corelează crește sau descrește cu o unitate de măsură.
- Regresia poate fi simplă și multiplă; liniară și neliniară. Ca și corelația, regresia poate fi directă, când fenomenele evoluează în același sens (crește x , crește y sau scade x scade și y), sau indirectă, când fenomenul evoluează în sens opus (crește x scade y sau scade x crește y).

Formula coeficientului de regresie este:

$$b_{xy} = r_{xy} \frac{\sigma_x}{\sigma_y} \text{ sau } b_{yx} = r_{yx} \frac{\sigma_y}{\sigma_x}$$

- b_{yx} = coeficientul de regresie a lui y în funcție de x . El exprimă, cantitativ, cu cât crește sau scade fenomenul y când x crește sau scade cu o unitate de măsură;
- b_{xy} = coeficientul de regresie a lui x în funcție de y . El exprimă, cantitativ, cu cât crește sau scade fenomenul x când y crește sau scade cu o unitate de măsură;
- r_{xy} = coeficientul de corelație liniară Bravais-Pearson;
- σ_x = deviația standard a fenomenului x ;
- σ_y = deviația standard a fenomenului y .

Interpretarea coeficientului de regresie

- Interpretarea „coeficientului de regresie” este următoarea:
- $b=0$, variabila y nu depinde de variabila x , ele sunt independente
- $b \neq 0$, cele două variabile sunt dependente astfel:
 - $b > 0$, legătura este directă
 - $b < 0$, legătura este inversă.

Multumesc