Capitolul 5
TRATAREA SIMULTANĂ A DOUĂ VARIABILE ORDINALE

Acest capitol şi următorul cuprind statistică neparametrică. Cu toate acestea, dependenţa între două variabile ordinale se numeşte, de asemenea, **corelaţie** ca şi în cazul a două dimensiuni, parametri. În schimb, testele de semnificaţie pentru coeficienţii de corelaţie specifici sunt teste neparametrice.

Sinteza grafică pentru aceste perechi de variabile se face ca şi pentru cele cantitative, dar se ţine cont de pierderea proprietăţilor legate de distanţe, adică de ceea ce caracterizează exclusiv dimensiunile. Recomandăm să nu se utilizeze stereohistograme.

§ 5.1. Sinteza numerică bivariată

Postulatul epistemologic este valabil şi în acest cadru.

Pentru măsurarea unei corelaţii presupuse între două variabile odinale, cel mai utilizat indicator este *coeficientul de corelaţie al lui Spearman.*

5.1.1. Coeficientul de corelaţie a rangurilor al lui Spearman

Fie un şir bivariat de ranguri, $(X_i, Y_i)_{i = 1,2,,...,N}$.

1° Notaţii: $\boldsymbol{R_S}$, $\boldsymbol{S}$ (pentru populaţii statistice în general), $\boldsymbol{\rho_S}$ (pentru populaţii statistice teoretice), $\boldsymbol{r_S}$ (pentru eşantioane).

2° Definiţie

Coeficientul de corelaţie (a rangurilor) al lui Spearman şi este dat de formula:

$$R_S = 1 - \frac{6 \cdot \sum_{i=1}^{N} d_i^{\,2}}{N \cdot (N^2 - 1)}, \text{ unde } d_i = X_i - Y_i.$$

Din [19] rezultă că acest coeficient nu este altceva decât coeficientul de corelaţie liniară Bravais-Pearson care aplicat rangurilor se poate rescrie mai simplu prin formula anterioară.

Evident, acest coeficient poate fi aplicat şi oricărui şir bivariat de măsurători (x_i, y_i), introducând în formula de mai sus, în locul măsurătorilor, rangurile acestora notate (X_i, Y_i), adică X_i = *rang* x_i şi Y_i = *rang* y_i. Dacă între valorile x_i, respectiv, y_i apar cazuri de "ex aequo", adică de egalitate, atunci pentru ca şi rangurile să fie egale, se vor corecta prin înlocuirea rangurilor distincte corespunzătoare, cu media lor aritmetică. Prin urmare, diferenţele d_i se vor calcula pe baza rangurilor corectate (X_i', Y_i'), după cum rezultă din următorul exemplu.

Exemplul 5.1.1.

x_i	y_i	rang de x_i = X_i	rang de x_i, corectat = X_i'	rang de y_i = Y_i	rang de y_i, corectat = Y_i'	$d_i = X_i' - Y_i'$	d_i^2
5	6	4	4	1	1	3	9,00
6	7	5	5	2	2	3	9,00
4	**8**	*3*	<u>*2,5*</u>	*3*	<u>*4*</u>	-1,5	2,25
3	**8**	1	1	*4*	<u>*4*</u>	-3	9,00
4	**8**	*2*	<u>*2,5*</u>	*5*	<u>*4*</u>	-1,5	2,25
							$\sum d_i^2 = 31{,}50$

Observăm că atât variabila x_i, cât şi variabila y_i conţin cazuri de egalitate. Astfel, în şirul (x_i) valoarea 4 apare de două ori, respectiv pe locurile 2 şi 3 în şirul ordonat ascendent. Deci rangurile 2 şi 3 se vor corecta înlocuindu-se cu rangul mediu 2,5 [= (2 + 3) / 2]. In mod analog, în şirul (y_i) valoarea 8 apare de 3 ori pe locurile 3, 4 şi 5, în şirul ordonat ascendent. Rangurile 3, 4 şi 5 se vor corecta, deci, aducându-se la valoarea rangului mediu 4 [=(3+4+5)/ 3].

Calculând R_S, după formula de mai sus, obţinem:

$$R_S = 1 - \frac{6 \cdot \sum_{i=1}^{N} d_i^2}{N \cdot (N^2 - 1)} = 1 - \frac{6 \cdot 31{,}5}{5 \cdot (25 - 1)} = -0{,}575 .$$

3° Proprietăţi

- $-1 \le R_S \le 1$, deoarece R_S este un coeficient de corelaţie liniară.
- $R_S > 0 \Leftrightarrow$ norul de puncte (x_i, y_i) are forma graficului unei funcţii strict crescătoare.
- $R_S < 0 \Leftrightarrow$ norul de puncte (x_i, y_i) are forma graficului unei funcţii strict descrescătoare.
- $R_S = 1 \Leftrightarrow$ rangurile X_i, respectiv, Y_i sunt identice $\Leftrightarrow$ toate punctele (x_i, y_i) sunt aşezate pe graficul unei funcţii strict crescătoare.
- $R_S = -1 \Leftrightarrow$ rangurile X_i, respectiv, Y_i sunt exact în ordinea inversă $\Leftrightarrow$ toate punctele (x_i, y_i) sunt aşezate pe graficul unei funcţii strict descrescătoare.

4° Exemple biomedicale

Literatura de specialitate, evidenţiază exemple de corelaţii ale rangurilor între: vârstă şi valoarea encefalogramei, vârstă şi concentraţia de minerale în ţesuturi, densitatea populaţiei şi rata mortalităţii, starea de sănătate a dinţilor şi concentraţia de fluor din organism, etc.