Dragomirescu L. Lucrari practice de biostatistica. Editia a III-a revazuta si adaugita, 263 pp.
Editura "Agronomica", Bucuresti, 2003.

Lp 11 Rezumat

Capitolele următoare, 8 şi 9 conţin metode neparametrice.

Capitolul 8. STATISTICĂ BIVARIATĂ. VARIABILE TIP RANG

81. Sinteza numerică bivariată

8.1.1. Coeficientul de corelaţie a rangurilor al lui Spearman

1° Notaţii: R_S, S (pentru populaţii statistice în general), ρ_S (pentru populaţii statistice teoretice), r_S (pentru eşantioane).

2° Definiţie

Sn **coeficientul de corelaţie (a rangurilor) a lui Spearman**

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

unde $d_i = X_i - Y_i$, X_i şi Y_i fiind ranguri.

R_S este R (*coeficientul de corelaţie liniară Bravais-Pearson*) aplicat rangurilor.

3° Proprietăţi

$-1 \le R_S \le 1$ deoarece R_S este un coeficient de corelaţie liniară.

$R_S > 0 \Leftrightarrow$ norul de puncte are forma graficului unei funcţii strict crescătoare.

$R_S < 0 \Leftrightarrow$ norul de puncte are forma graficului unei funcţii strict descrescătoare.

$R_S = 1 \Leftrightarrow$ rangurile sunt identice $\Leftrightarrow$ toate punctele sunt aşezate pe graficul unei funcţii strict crescătoare. În acest caz, vorbim despre o **corelaţie perfectă (directă)**.

R_S = *-1* ⇔ rangurile sunt exact în ordinea inversă ⇔ toate punctele sunt aşezate pe graficul unei funcții strict descrescătoare. În acest caz, vorbim despre o **corelație perfectă inversă.**

8.1.2. Test de semnificație pentru coeficientul de corelație a rangurilor al lui Spearman, r_S

H_0: $\rho_S = 0$ (nu există corelație a rangurilor)

H_A: $\rho_S \neq 0$ (există corelație a rangurilor)

Statistica testului: $t = r_S \sqrt{\dfrac{n-2}{1-r_S^{\,2}}} \in t_{n-2}$, *n* fiind volumul eşantionului aleator simplu.

Decizia statistică: se respinge ipoteza nulă dacă $|t| > t_{n-2;\,\alpha/2}$.

Capitolul 9. STATISTICĂ BIVARIATĂ. VARIABILE CALITATIVE

Dependența între două variabile calitative se numeşte **asociere.** Sn **tabelă de contingență** rezultatul "ventilării unei populații după variantele a două caracteristici calitative" (Benzécri).

9.1. Sinteza grafică

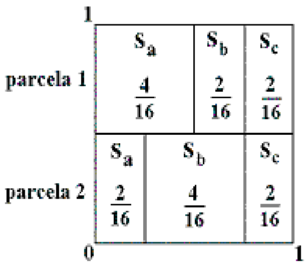
Se face într-o diagramă în **areale în dreptunghiuri** (în cadrul unui pătrat).

Seria bivariată:

Parcela:	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
Specia:	a	a	a	a	b	b	c	c	a	a	b	b	b	b	c	c

Tabela cu dublă intrare:

	specia a	specia b	specia c
parcela 1	4	2	2
parcela 2	2	4	2



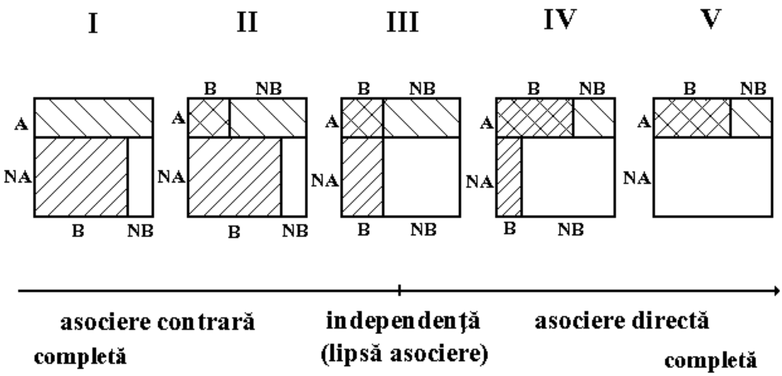
Dragomirescu L. Lucrari practice de biostatistica. Editia a III-a revazuta si adaugita, 263 pp. Editura "Agronomica", Bucuresti, 2003.

9.2. Sinteza numerică. Coeficientul de contingență al lui Ciuprov

Aşa cum pentru perechile de variabile tip măsurătoare, dependența statistică apare plasată între dependența funcțională şi independență, tot aşa asocierea este plasată între *asocierea completă* şi independență.

9.2.1. Tabele 2 × 2

În cazul variabilelor binare vorbim despre *asociere directă*, respectiv *contrară*.



Două variabile calitative binare *a* (cu variantele *A* şi *NA*) şi *b* (cu variantele *B* şi *NB*) sn **complet asociate în mod direct** dacă *B* are loc doar în prezența lui *A* sau invers şi se numesc **complet asociate în mod contrar** dacă *B* are loc doar în absența lui *A* sau invers.

Dacă proporția lui *B* din *A* este egală cu proporția lui *B* din *NA* (ceea ce implică şi egalitatea cu proporția lui *B*), *a* şi *b* sn **independente**. În caz contrar, cele două variabile sn **asociate**.

9.2.2. χ^2 pentru independență versus asociere, în tabele de contingență p × q

Formulă teoretică şi $\boxed{\text{formulă de calcul:}}$

$$\chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{(o(i,j)-t(i,j))^2}{t(i,j)} = \boxed{\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{o(i,j)^2}{t(i,j)} - n}$$

în care $t(i, j)$ sunt **frecvențe teoretice în ipoteza de independență**: $t(i,j) = l(i) \cdot c(j) / n$, $l(i)$ fiind sumele pe linii, $c(j)$ sumele pe coloane şi n totalul general.

χ^2 măsoară intensitatea absolută a asocierii prin depărtarea de la independență.

1° Coeficientul de contingență al lui Ciuprov

Notat T, s-a introdus pentru a nu depinde de p, q şi n:

$$T = \sqrt{\frac{\chi^2}{n \cdot \sqrt{(p-1) \cdot (q-1)}}}$$

Ti este o măsură relativă a asocierii şi variază între 0 şi 1 dacă $p = q$.

2° χ^2 - Tipuri de aplicații biomedicale

(a) studiul asocierii caracterelor biologice sau de mediu,
(b) studiul factorilor de risc în epidemiologia bolilor cronice,
(c) testarea eficacității unui tratament şi
(d) compararea eficacității a două tratamente.

9.3. Testarea independenței

H_0: "Cele două caractere sunt independente."
H_A: "Cele două caractere nu sunt independente, sunt asociate."
Statistica testului este χ^2, în condițiile specificate în continuare.
Numărul de grade de libertate: $\nu = (p - 1) \cdot (q - 1)$.
Decizia statistică: Se respinge H_0 dacă $\chi^2 > \chi^2_{\nu;\, \alpha}$

χ^2 - Condiții de aplicare

1. Se aplică la tabele de contingență, adică tabelele trebuie să conțină valori absolute, nu procente sau proporții.
2. Observațiile trebuie să fie independente.
3. Alte exigențe: $n > 20$ şi frecvențele teoretice absolute $t(i, j) \geq 5$ sau chiar ≥ 10.
4. Pentru tabele 2×2 se aplică lui χ^2 **corecția de continuitate a lui Yates**:

$$\sum\sum \frac{(|\,o(i,j) - t(i,j)\,| - 0{,}5)^2}{t(i,j)}.$$

Notând frecvențele observate $o(i, j)$ cu a, b, c, d, astfel:

$o(1, 1) = \mathbf{a}$	$o(1, 2) = \mathbf{b}$
$o(2, 1) = \mathbf{c}$	$o(2, 2) = \mathbf{d}$

se obține următoarea formulă de **calcul rapid și exact al lui χ^2 cu corecția lui Yates**:

$$\chi^2 = \frac{n \cdot (|a \cdot d - b \cdot c| - n/2)^2}{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)}$$ în care $n = a + b + c + d$.

Exprimarea în limbaj statistic în cazul celor 4 tipuri de aplicații biomedicale

Dacă se acceptă H_0:

(a) Cele două caractere nu sunt asociate în mod semnificativ.
(b) Factorul F nu este în mod semnificativ un factor de risc în patologia B.
(c) Tratamentul nu este eficace în mod semnificativ.
(d) Cele două tratamente nu diferă în mod semnificativ ca eficacitate.

Dacă se respinge H_0 la nivelul $p < \alpha$:

(a) Cele două caractere sunt asociate în mod #.
(b) Factorul F este în mod # un factor de risc în patologia B.
(c) Tratamentul este în mod # eficace.
(d) Cele două tratamente diferă în mod # ca eficacitate.

Semnul # se înlocuiește prin:

semnificativ (*$p < 0,05$)	dacă $\alpha = 0,05$
foarte semnificativ (**$p < 0,01$)	dacă $\alpha = 0,01$
înalt semnificativ (***$p < 0,001$)	dacă $\alpha = 0,001$.

Observații: (1) Deși este un test unilateral, în practică se utilizează pragurile de test bilateral.

(2) Verificarea eficacității unui tratament (medicament) se efectuează neparametric în doi pași: (I) se verifică efectivitatea printr-un test χ^2 de independență aplicat unei tabele de contingență și, în cazul acceptării efectivității (adică al respingerii independenței), (II) se calculează riscul relativ (RR) al celor tratați față de cei netratați, risc care trebuie să fie supraunitar. Vezi, în continuare, problemele 4 și 11.

Capitolul 10. Elemente de teoria probabilităților

10.1. Istoric

Teoria probabilităților a fost fondată în secolul al XVII-lea de către Pascal și Fermat.

10.2. Conceptul de probabilitate

Două accepţiuni ale noţiunii de probabilitate: 1) **obiectivă** şi 2) **subiectivă**.

10.3. Accepţiunea de probabilitate obiectivă

Probabilitatea obiectivă are două definiţii:

10.3.1. Definiţia clasică

Probabilitatea unui eveniment *A* = numărul de cazuri (evenimente elementare, probe) favorabile lui *A* divizat prin numărul total (posibil) de cazuri $p(A) = n_A / n$. Este o probabilitate **a priori** efectuării experimentului.

10.3.2. Definiţia empiristă (von Mieses)

Probabilitatea unui eveniment *A* = numărul către care tinde să se stabilizeze frecvenţa relativă a evenimentului respectiv pe măsură ce experimentul se repetă de un număr cât mai mare de ori. Este o probabilitate **a posteriori** experimentului.

10.4. Accepţiunea de probabilitate subiectivă

Probabilitatea subiectivă reprezintă o codificare (subiectivă) de informaţii efectuată de o persoană interesată în a o evalua = "traducerea bunului simţ în cifre" [5].

10.5. Definiţia axiomatică a probabilităţii (Kolmogorov)

10.5.1. Corespondenţe între teoria mulţimilor şi cea a probabilităţilor

Vezi [3].

10.5.2. Câmp finit de evenimente

Fie *E* o mulţime finită de evenimente elementare. Sn **câmp (finit) de evenimente** o mulţime de evenimente din *E*, notată K ($\subset P(E)$) care (1) conţine evenimentul sigur; (2) este închisă* la considerarea evenimentului opus unui eveniment dat şi (3) este închisă la disjuncţia de două evenimente.

---

* Expresia "este închisă la o anume operaţie" înseamnă că aplicând operaţia elementelor din mulţimea respectivă, nu putem ieşi din cadrul acelei mulţimi, adică este închisă (la operaţia specificată).

10.5.3. Probabilitate (în sens axiomatic) şi câmp finit de probabilitate

Fie K un câmp finit de evenimente peste E. Sn **probabilitate** o funcţie de mulţime $p: K \to \mathbb{R}_+$ (1) subunitară; (2) cu probabilitatea evenimentului sigur egală cu 1 şi (3) aditivă pentru evenimente mutual incompatibile.
Tripletul $\{E, K, p\}$ sn **câmp (finit) de probabilitate**.

10.5.4. Sistem complet de evenimente

Fie (E, K) un câmp finit de evenimente şi $C = \{A_i\}_{i=1,2,\ldots,n}$ o familie de evenimente din K. C sn **sistem complet de evenimente** dacă (1) orice A_i din C este diferit de evenimentul imposibil; (2) orice două evenimente diferite sunt incompatibile, adică C este o mulţime de evenimente mutual incompatibile şi (3) disjuncţia evenimentelor din C este evenimentul sigur, E.
Exemplu: C = mulţimea fenotipurilor sanguine, E fiind mulţimea genotipurilor sanguine.

10.5.6. Probabilitate condiţionată şi evenimente independente

1° Probabilitate condiţionată

Pentru două evenimente A şi B compatibile, sn **probabilitatea lui *A* condiţionată de *B*** şi se notează ***p (A /B)***, raportul $p\,(A \cap B)\,/\,p\,(B)$.

2° Aplicaţii ale probabilităţii condiţionate în epidemiologie

Riscul absolut de contractare a unei anumite boli *b* în prezenţa unui anumit factor *f*, este probabilitatea de a contracta boala b condiţionată de prezenţa factorului f, probabilitate notată $p\,(B / F)$.
Riscul relativ de contractare a unei anumite boli *b* al celor care prezintă factorul *f* (faţă de cei care nu prezintă factorul *f*) este $RR = \dfrac{p(B|F)}{p(B|nonF)}$.

Studiile prospective presupun urmărirea în timp îndelungat a două loturi, unul fiind supus factorului cercetat f, iar altul fiind în situaţia contrară. Sunt fie foarte costisitoare, fie chiar imposibile.
Studiile retrospective compară proporţia de indivizi care au fost supuşi factorului de risc din cadrul unui lot de bolnavi de boala respectivă b, cu aceeaşi proporţie din cadrul întregii populaţii biologice din care provine lotul de bolnavi. Are cost cu mult mai mic şi permite calculul riscului relativ chiar dacă nu poate furniza riscul absolut.

3° Formula probabilității totale

$p(X) = \sum_{j=1}^{n} p(X|A_j) \cdot p(A_j)$, în care $X \in K$ (câmp finit de evenimente din E) şi

$C = \{A_i\}_{i = 1, 2, \ldots, n}$ este un sistem complet de evenimente din K.

5° Evenimente independente versus dependente

Două evenimente A şi B compatibile sn **independente** dacă: $p(B \cap A) = p(B) \cdot p(A)$ sau $p(B) = p(B \mid A)$ sau $p(A) = p(A \mid B)$ sau, altfel spus, dacă probabilitatea realizării unui eveniment rămâne neschimbată dacă se realizează sau nu celălalt eveniment. În caz contrar, se numesc **dependente**.

6° Formula lui Bayes

$p(A_k \mid X) = \dfrac{p(X \mid A_k) \cdot p(A_k)}{\sum_{i=1}^{n} p(X \mid A_i) \cdot p(A_i)}$, în care $X \in K$ (câmp finit de evenimente din E) şi $C = \{A_i\}_{i=1,2,\ldots,n}$ este un sistem complet de evenimente din K.

Recalcularea (prin această formulă) a probabilitățile evenimentelor de interes după producerea a diverse evenimente, produce estimări ale probabilităților din ce în ce mai bune.

Lp 11 Teste, exerciții şi probleme

TG11. Durata 200'' pe calculator.

Dependenta in plan fenomenologic intre doua aspecte exprimate prin variabile tip rang se numeste:
1. asociere.
2. corelatie.
3. contingenta.

Dependenta in plan fenomenologic intre doua aspecte exprimate prin variabile calitative se numeste:
1. corelatie.
2. contingenta.
3. asociere.

Tabela statistica specifica statisticii bivariate se numeste:
1. tabela statistica simpla.
2. tabela de contingenta.
3. tabela de corelatie.
4. tabela cu dubla intrare.

O serie bivariata formata din variabile calitative se grupeaza intr-o tabela denumita:
1. tabela de corelatie.
2. tabela de contingenta.
3. tabela cu dubla intrare.
4. tabela statistica simpla.

Alegeti afirmatia ERONATA:Coeficientul de corelatie a rangurilor al lui Spearman, Rs, are valoarea 1 daca si numai daca:
1. rangurile sunt identice.
2. corelatia este perfecta.
3. corelatia este absoluta.

Alegeti afirmatia ERONATA:
Coeficientul de corelatie a rangurilor al lui Spearman, Rs, are valoarea - 1 daca si numai daca:
1. rangurile nu sunt identice.
2. rangurile sunt exact in ordine inversa.
3. corelatia este perfecta si inversa.

R fiind coeficientul de corelatie liniara, iar Rs coeficientul de corelatie a rangurilor, alegeti afirmatia adevarata pentru seria bidimensionala vizualizata prin:

```
^         .
|        .
|       .
|      .
|     .
|    .
|   .
-------------------------------->
```

1. R = 1 si Rs < 1.
2. R = 1 si Rs = 0.
3. R = Rs = 1.

R fiind coeficientul de corelatie liniara, iar Rs coeficientul de corelatie a rangurilor, alegeti afirmatia adevarata pentru seria bidimensionala vizualizata prin:

```
^      ....
|     . .   ...
|    ..      ..
|   ..        ...
|  ..          ...
| ..             ..
| .               ..
---------------------------------->
```

1. R = 0 si Rs = 1.
2. R = Rs = 0.
3. R = 1 si Rs = 0.

Pentru masurarea legaturii intre doua aspecte dintr-un fenomen exprimate prin variabile calitative, se utilizeaza:
1. coeficientul de contingenta al lui Ciuprov.
2. coeficientul de corelatie al lui Pearson.
3. coeficientul de corelatie al lui Spearman.

TC11. Durata 5'.

1. R_S este de fapt ____ aplicat rangurilor.
2. Când R_S = 1 înseamnă că rangurile sunt ____________, iar când R_S = -1 rezultă că rangurile sunt ___________________________.
3. În statistica bivariată pentru variabile calitative, χ^2 măsoară depărtarea de la ________________, respectiv intensitatea absolută a ________________.

4. Coeficientul de contingență al lui Ciuprov, notat _____, este o măsură ___________ a asocierii.
5. În cazul unui test statistic pentru problema independență versus asociere, cuplul de ipoteze este:

H_0: ______________________________

H_A: __________________________________.

6. Pentru a putea aplica χ^2 în testarea independenței, tabelele de contingență trebuie să conțină valori _____________, observațiile trebuie să fie __________________, iar pentru tabele 2 × 2 se aplică lui χ^2 o _____________________________.
7. Probabilitatea obiectivă definită clasic este o probabilitate _____________, iar definită empirist este o probabilitate ________________ experimentului.
8. Studiile ______________ presupun urmărirea în timp îndelungat a două loturi, unul fiind supus unui factor cercetat, iar celălalt nu.
9. Evenimentule A şi B se numesc ___________________ dacă probabilitatea a priori a unuia este egală cu probabilitatea sa a posteriori producerii celuilalt, adică $p\ (B) =$ __________.
10. Se pot face predicții din ce în ce mai bune pe măsură ce repetăm experimentul, dacă se aplică _____________________________.

Exerciții sau probleme rezolvate

EGA 1.

Se da urmatorul sir bidimensional reprezentand un esantion aleator de 6 perechi de valori x,y:

x	y
0	0
1	2
2	4
3	7
4	10
5	10

Stiind ca in fenomen exista o legatura intre cele doua variabile:
1. Sa se reprezinte sirul intr-o diagrama de imprastiere.
2. Sa se calculeze rS, coeficientul de corelatie a rangurilor al lui Spearman si
3. statistica tniu* necesara pentru testarea semnificatiei lui rS.
4. Consultand tabela corespunzatoare, testati semnificatia lui rS.
(Rezultatele finale vor fi rotunjite la 2 zecimale.)

* tniu este forma sub care afişează programul de calculator t_ν.

Rezolvare:

LEGENDA: **$**=0,5 puncte. **$$** din oficiu. Numarul de **$** se imparte la 20.

1. Desenam norul de puncte pentru a vedea daca forma corelatiei este o functie strict monotona:

```
$$$  ^
     |                         *        *
     |
     |
     |                    *
     |
     |
     |               *
     |
     |          *
     * ------------------------------------------->
```

Norul de puncte are forma unei curbe strict crescătoare (deci $r_S > 0$, corelație a rangurilor directă), dar punctele nu pot fi aşezate pe graficul unei funcții strict crescătoare, ci doar crescătoare (de aceea $r_S < 1$).

2. Calculul coeficientului de corelatie a rangurilor al lui Spearman.

Alcatuim tabela:

X	Y	rangX **$**	rangXcorectat **$**	rangY **$**	rangYcorectat **$**	d **$**	d^2 **$**
0	0	1	1	1	1	0	0
1	2	2	2	2	2	0	0
2	4	3	3	3	3	0	0
3	7	4	4	4	4	0	0
4	10	5	5	5	5.5	-.5	.25
5	10	6	6	6	5.5	.5	.25

Σ d^2 =.5 **$**

$$ rS = 1 - 6 * Σ d^2 / (n * (n^2 - 1) = 1 – 6 *.5 / (6 * (36 - 1) = 1 - 3 / 210 =
=(210 - 3) / 210 = 207 / 210 =.9857143 ≈.99

3. Calculul lui tniu:

$$ tniu = rS * RAD ((n - 2) / (1 - rS * rS)) =
=.99 * RAD (4 / (1 - (.99) * (.99))) =
=.99 * RAD (4 / (1 -.9801)) =
=.99 * RAD (4 / 1.989998E-02) =
=.99 * RAD (201.0052) =
=.99 * 14.17763 = 14.03585 ≈ 14.04

4. Semnificatia lui rS:

$ Numar de grade de libertate, niu = n - 2 = 6 - 2 = 4

$$$ Deoarece | tniu | = 14.04 > t bilateral tabelat (pt. alfa = 0,001 si niu = 4)= 8.61
exista o corelatie A RANGURILOR inalt semnificativ (***) diferita de 0.
Altfel scris, rS =.99 <> 0 *** ori RS <> 0 *** (p < 0.001).

EGA 2.

Să se aplice enunțul anterior şirului bidimensional din EGA 1 din Lp 10.

Rezolvare:

LEGENDA: **$**=0,5 puncte. **$$** din oficiu. Numarul de **$** se imparte la 20.

1. Desenam norul de puncte pentru a vedea daca forma corelatiei este o functie strict monotona:

A se vedea desenul din rezolvarea problemei EGA 1 din Lp 10.

Având formă liniară şi linia dreaptă fiind o funcție strict monotonă, putem considera că forma corelației este cea a unei funcții strict monotone.

2. Calculul coeficientului de corelatie a rangurilor al lui Spearman.

Alcatuim tabela:

X	Y	rangX **$**	rangXcorectat **$**	rangY **$**	rangYcorectat **$**	d **$**	d^2 **$**
1	1	1	1	1	1	0	0
3	3	2	2	2	2	0	0
5	5	3	3.5	3	3	.5	.25
5	6	4	3.5	4	4	-.5	.25
							Σ d^2=.5 **$**

$$ rS = 1 - 6 * Σ d^2 / (n * (n^2 - 1) = 1 – 6 *.5 / (4 * (16 - 1) = 1- 3 / 60 =
=(60 - 3) / 60 = 57 / 60 =.95 ≈.95

3. Calculul lui tniu:

$$ tniu = rS * RAD ((n - 2) / (1 - rS * rS)) =
=.95 * RAD (2 / (1 - (.95) * (.95))) =
=.95 * RAD (2 / (1 -.9025)) =
=.95 * RAD (2 / 9.750003E-02) =
=.95 * RAD (20.51282) =
=.95 * 4.529108 = 4.302652 ≈ 4.3

4. Semnificatia lui rS:

$ Numar de grade de libertate, niu = n - 2 = 4 - 2 = 2

$$$ Deoarece | tniu | = 4.3 < t bilateral tabelat (pt. alfa = 0,05 si niu = 2) = 4.303
NU exista o corelatie A RANGURILOR semnificativ diferita de 0.
Altfel scris, rS =.95 “=” 0 ori RS = 0 (p ≥ 0.05).

3.

În EGA 1 din Lp 10 s-a obținut o corelație liniară semnificativă, iar în exercițiul anterior, EGA 2, s-a constat lipsa unei corelații a rangurilor, adică o corelație după o funcție strict monotonă. Explicați de ce este posibil acest rezultat aparent contradictoriu (căci dacă există o corelație liniară ar trebui să existe şi o corelație după o funcție monotonă, deoarece o funcție liniară este o funcție strict monotonă).

Rezolvare:
Testarea semnificației coeficientului de corelație liniară, r, se face printr-un test parametric, iar semnificația coeficientului de corelația rangurilor, r_S, se verifică printr-un test neparametric. Deoarece testele parametrice sunt mai puternice înseamnă că, pe acelaşi set de date, un test parametric poate să respingă ipoteza nulă în timp ce unul neparametric poate să o accepte. Pentru a obține respingere şi prin testul neparametric, va trebui mărit volumul.

4.

90 de hipertensivi au fost repartizați prin randomizare în două loturi. Cei 50 de pacienți care au ajuns aleator în primul lot au primit substanța activă a unui medicament conceput ca antihipertensiv, iar cei din al doilea lot au primit placebo. Nici pacienții, nici personalul medical care a administrat tabletele respective nu au cunoscut care este lotul tratat şi care este aşa numitul lot martor format din cei netratați, primind doar placebo (studiu dublu orb). În finalul experimentului clinic s-au constatat următoarele rezultate:

	Ameliorați	*Staționari*
Tratați (lot tratat)	30	20
Netratați (lot martor)	10	30

Să se testeze eficacitatea tratamentului respectiv.

Rezolvare:
Conform observației din paragraful 9. 3. (vezi rezumatul)
I) Testăm, mai întâi, <u>efectivitatea</u> tratamentului. Pentru aceasta calculăm totalurile marginale şi cel general, n:

	Ameliorați	*Staționari sau agravați*	*Totaluri:*
Tratați (lot tratat)	$a = 30$	$b = 20$	$a+b = 50$
Netratați (lot martor)	$c = 10$	$d = 30$	$c+d = 40$
	$a + c = 40$	$b + d = 50$	$n = a+b+c+d = 90$

Cuplul de ipoteze de verificat este:

H_0: „Medicamentul nu are efect.“ sau „Medicamentul şi placebo au acelaşi efect.“
H_A: „Medicamentul are efect.“ sau „Medicamentul şi placebo au efecte diferite.“

Aplicăm *formula de calcul rapid şi exact al lui χ^2 cu corecția lui Yates*:

$$\chi^2 = \frac{n \cdot (|a \cdot d - b \cdot c| - n/2)^2}{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)} = \frac{90 \cdot (|30 \cdot 30 - 20 \cdot 10| - 90/2)^2}{50 \cdot 40 \cdot 40 \cdot 50} = 9{,}65.$$

Deoarece tabela are $p = 2$ linii şi $q = 2$ coloane, numărul de grade de libertate $\nu = (2-1) \cdot (2-1) = 1$. În tabela χ^2 din Anexa 4, găsim pe linia 1 valorile 6,635, respectiv 10,827, pentru nivelele de semnificație 0,01, respectiv 0,001.

Decizia statistică pentru efectivitate: Deoarece valoarea calculată, 9,65 > 6,635 respingem ipoteza nulă, considerând că există o efectivitate foarte semnificativă (**p < 0,01), dar nu şi înalt semnificativă (valoarea calculată, 9, 65 < 10,827).

Decizia de specialitate: Considerăm cu risc sub 1% că tratamentul este <u>efectiv</u>, adică are efect, dar nu ştim dacă benefic sau malefic. Pentru a stabili acest lucru executăm pasul II.

II) Calculăm riscul relativ (*RR*) pentru a vedea care dintre cele două loturi are un “risc” mai mare de a fi ameliorat şi de câte ori este mai mare: $RR = (a / (a + b)) / (c / (c + d))$ = (30 / 50) / (10 / 40) = (3 / 5) / (1 / 4) = 3 · 4 / 5 = 12 / 5 = 2,4.

Decizia statistică şi de specialitate, finală: Cuplând cele două puncte ale problemei putem conchide, cu **p < 0,01, că “medicamentul are o <u>eficacitate</u> mai mare de 2,4 decât placebo”.

5.

Pentru a se studia selectivitatea în hrănire a unei specii de erbivore, un lot de indivizi a fost lăsat să pască pe o porțiune de păşune de 10 m^2 care conținea două specii ierboase. S-a calculat abundența celor două specii ierboase pe m^2 (exprimată în nr. exemplare / m^2), înainte şi după păscut. Datele obținute au fost următoarele:

Densitatea → *Specia ierboasă* ↓	*înainte de păscut*	*după păscut*
A	11	6
B	9	7

1) Manifestă specia de erbivore considerată preferinţe pentru vreuna dintre cele două specii ierboase (luând $\alpha = 0{,}05$) ?
2) Dacă da, pe care o preferă şi de câte ori o preferă mai mult?

Rezolvare:

Rezolvare greşită:

χ^2 aplicat direct tabelei de mai sus. Dublă greşeală: nu conţine coloanele necesare: număr exemplare păscute, respectiv nepăscute şi datele sunt relative (la 1 m^2), nu absolute (referitoare la întreaga suprafaţă testată).

Rezolvarea corectă:

Calculăm tabela de valori absolute:

Număr exemplare → *Specia ierboasă* ↓	*înainte de păscut*	*după păscut*
A	110	60
B	90	70

apoi tabela cu datele necesare:

Număr exemplare → *Specia ierboasă* ↓	*păscute*	*nepăscute (după păscut)*	*înainte de păscut*
A	$a = 50$	$b = 60$	$a+b = 110$
B	$c = 20$	$d = 70$	$c+d = 90$
	$a + c = 70$	$b + d = 130$	$n = a+b+c+d = 200$

la care am adăugat totalurile. Cuplul de ipoteze de verificat este:

H_0: "Cele două caractere sunt independente."

H_A: "Cele două caractere nu sunt independente, sunt asociate."

Aplicăm *formula de calcul rapid şi exact al lui χ^2 cu corecţia lui Yates*:

$$\chi^2 = \frac{n \cdot (|a \cdot d - b \cdot c| - n/2)^2}{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)} = \frac{200 \cdot (|50 \cdot 70 - 60 \cdot 20| - 200/2)^2}{110 \cdot 90 \cdot 70 \cdot 130} = 10{,}74.$$

Deoarece tabela are $p = 2$ linii şi $q = 2$ coloane, numărul de grade de libertate $\nu = (2-1) \cdot (2-1) = 1$. În tabela χ^2 din Anexa 4, găsim pe linia 1 valorile 6,635, respectiv 10,827, pentru nivelele de semnificaţie 0,01, respectiv 0,001.

Decizia statistică finală: Deoarece valoarea calculată, 10,74 > 6,635 respingem ipoteza nulă, considerând că există o asociere foarte semnificativă (**$p < 0{,}01$), dar nu şi înalt semnificativă (valoarea calculată, 10,74 < 10,827).

Decizia de specialitate: Considerăm cu risc sub 1% că specia de erbivore manifestă o preferinţă între cele două specii ierboase.

2) Calculăm riscul relativ (RR) pentru vedea care specie este preferată şi de câte ori mai mult: $RR = (a / (a + b)) / (c / (c + d))$ = (50 / 110) / (20 / 90) = 0,45 / 0,22 = 2,05.

Specia *A* este preferată speciei *B* de 2,05 ori mai mult.

Observație: Cuplând cele două puncte ale problemei putem conchide că “specia *A* este preferată speciei *B* de 2,05 ori mai mult cu **p < 0,01”.

Exerciții sau probleme propuse

EGA 6.

Se dă şirul bidimensional din EGA 4 (Lp 10).

Ştiind că în fenomen există o legătură între cele două variabile:

1. Să se reprezinte şirul într-o diagramă de împrăştiere.
2. Să se calculeze r_S, coeficientul de corelație a rangurilor al lui Spearman şi
3. statistica t_v necesară pentru testarea semnificației lui r_S.
4. Consultând tabela corespunzătoare, testați semnificația lui r_S.

(Rezultatele finale vor fi rotunjite la 2 zecimale.)

7.

Prelucrând aceeaşi serie statistică în EGA 4 din Lp 10, s-a obținut o corelație liniară semnificativă (*), iar în EGA 6 de mai sus, o corelație a rangurilor înalt semnificativă (***). Explicați acest rezultat.

EGA

Pentru fiecare şir bivariat din următoarele trei:

8.		**9.**		**10.**	
x	y	x	y	x	y
0	9	0	0	0	0
0	8	0	1	1	2
1	8	1	0	2	4
2	4	1	1	3	2
3	2			4	0
4	1				
5	1				

stiind ca in fenomen exista o legatura intre cele doua variabile:

1. Sa se reprezinte sirul intr-o diagrama de imprastiere.
2. Sa se calculeze rS, coeficientul de corelatie a rangurilor al lui Spearman si

3. statistica tniu necesara pentru testarea semnificatiei lui rS.
4. Consultand tabela corespunzatoare, testati semnificatia lui rS.
(Rezultatele finale vor fi rotunjite la 2 zecimale.)

11.

Pentru evaluarea efectului antipiretic al unei substanțe s-a organizat un experiment clinic dublu orb în care 80 de voluntari au primit substanța iar 40 au primit placebo. Pentru fiecare lot s-a notat numărul celor la care s-a observat un efect antipiretic, respectiv numărul celor fără efect evidențiat:

	Cu efect	*Fără efect*
Tratați (lot tratat)	20	60
Netratați (lot martor)	30	10

Să se testeze eficacitatea substanței respective.

12.

În vederea distrugerii unui dăunător al fagului, s-au testat două metode: una chimică şi una biologică. Pentru aceasta, s-au considerat două parcele de pădure, cărora li s-au aplicat cele două tratamente. S-au numărat arborii afectați de dăunător în ambele parcele, înainte şi după tratament. Datele obținute au fost următoarele:

Arbori atacați $\rightarrow$ *Tratament* $\downarrow$	*înainte de tratament*	*după tratament*
Biologic	83	42
Chimic	221	156

1) Diferă ca eficacitate cele două tratamente (considerând $\alpha = 0,05$) ?
2) Dacă da, care este mai eficace şi de câte ori este mai eficace.