CZU: 372.8004:371.388

ABORDĂRI DIDACTICE PRIVIND APLICAREA METODELOR NEPARAMETRICE LA PRELUCRAREA DATELOR EXPERIMENTALE

Liubomir CHIRIAC, Aureliu DANILOV

llchiriac@gmail.com, aureliu.danilov@gmail.com

Universitatea de Stat din Tiraspol

Abstract. *În acest articol sunt studiate diverse aspecte didactice privind aplicarea metodelor neparametrice la prelucrarea datelor experimentale. La fel, sunt examinate unele modalități de utilizarea a criteriului φ*-Fischer.*

Cuvinte cheie: *metode neparametrice, criteriului φ*-Fischer, efect, experiment.*

Summary. *In this article we study various didactic aspects regarding the application of non-parametric methods to the processing of experimental data. Also, some ways to use the φ*-Fischer criterion are examined.*

Keywords: *non-parametric methods, φ * -Fischer criterion, effect, experiment.*

1. Aspecte teoretice privind prelucrarea statistică a datelor experimentale

Metodele statistice sunt clasificate după informația care există referitor la populația examinată. În contextul dat pentru prelucrarea statistică a datelor, de cele mai multe ori, sunt folosite:

1) Metodele parametrice;
2) Metodele neparametrice.

Mai jos vom examina mai detailat unele din metodele neparametrice de prelucrare a datelor statistice. Să examinăm succint esența metodelor neparametrice.

Metodele neparametrice sunt tehnici statistice pentru care nu este necesar de făcut nici o presupunere de parametri pentru populația pe care o cercetăm. Într-adevăr, metodele respective nu au nici o dependență de populația examinată. Setul de parametri nu mai este fixat și nici distribuția pe care o folosim nu se cunoaște, ca în cazul metodelor parametrice. Din aceste considerente, metodele neparametrice sunt denumite și metode fără distribuție. Multe dintre aceste metode neparametrice sunt ușor de aplicat și de înțeles.

În situația când dimensiunile eșantionului sunt mici este dificil de spus dacă populațiile se subordonează distribuției normale. Anume în astfel de situații este potrivit de aplicat teste neparametrice. Mai jos vom examina, testul neparametric φ^* **- criteriul de transformare unghiulară Fischer, ori succint criteriul** φ^* **- Fischer** [1-4].

2. φ^* - criteriul de transformare unghiulară al lui Fischer

Metoda φ^* - Fischer este utilizată pe scară largă și sub denumirea de „transformarea unghiulară a lui Fischer” (vezi Gubler E.V., 1978; Sidorenko E.V., 2000). φ^* **- criteriul de transformare unghiulară al lui Fischer se consideră un criteriu statistic multifuncțional**.

Criteriul respectiv a fost conceput pentru a compara două eșantioane în funcție de frecvența de apariție a efectului de interes pentru cercetător. Criteriul evaluează fiabilitatea diferențelor dintre procentele a două eșantioane și nu are restricții privind numărul de probe. În ce constă esența metodei neparametrice φ^* - Fischer?

Esența metodei neparametrice φ^* - Fischer constă în convertirea procentelor (acțiunilor) într-o valoare φ, a cărei distribuție este apropiată de normal. Astfel se determină:

- Formula de transformare unghiulară: $\varphi = arcsin\sqrt{p}$, unde p este un procent exprimat în fracții.
- Formula de evaluare a semnificației diferențelor de acțiuni (procente):

$$\varphi^*_{emp} = (\varphi_1 - \varphi_2) * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}}$$

unde n_1 și n_2 sunt dimensiunile eșantionului.

- Procentele p_1 și p_2. Este necesar ca $p_1 \neq 0$ și $p_2 \neq 0$.

În situația când se examinează două eșantioane se recomandă respectarea următoarelor condiții:

1) dacă există 2 observații într-un eșantion ar trebui să fie cel puțin 30 în al doilea: n_1=2→n_2≥30;
2) dacă există 3 observații într-un eșantion ar trebui să fie cel puțin 7 în al doilea:

n_1=3→n_2≥7;

3) dacă există 4 observații într-un eșantion ar trebui să fie cel puțin 5 în al doilea:

n_1=4→n_2≥5;

4) dacă există 5 observații într-un eșantion ar trebui să fie cel puțin 5 în al doilea. Deci, pentru n_1, n_2≥5, orice comparație este posibilă.

În criteriu dat este important să se determine dacă „este efectul” ori „nu este efectul” privind fenomenul studiat. În articolul dat ne vom referi doar la „**efectele**” care au în special caracteristici calitative (de exemplu: exprimarea acordului cu propunerea examinată, alegerea unei opțiuni etc.).

Ipotezele care se verifică:

H_0: Proporția populației care prezintă efectul studiat nu este mai mare în eșantionul 1 comparativ cu eșantionul 2.

H_1: Proporția populației care prezintă efectul studiat este mai mare în eșantionul 1 comparativ cu eșantionul 2.

Valoarea empirică φ^* se calculează folosind formula:

$$\varphi^*_{emp} = (\varphi_1 - \varphi_2) * \sqrt{\frac{n_1 * n_2}{n_1 + n_2}}$$

unde: φ_1 - unghiul corespunzător procentului mai mare; φ_2 - unghiul corespunzător unui procent mai mic; n_1 - numărul de observații din eșantionul 1; n_2 - numărul de observații din eșantionul 2. Valoarea critică se determină conform tabelelor XII și XIII indicate în lucrarea [2, pag. 330]. Astfel, $\varphi^*_{0,01} = 2{,}31$ pentru nivelul de semnificație p=0,01 și $\varphi^*_{0,05} = 1{,}64$ pentru p = 0,05. Ulterior se compară valoarea obținută a lui φ* cu valorile critice: $\varphi^*_{cr} \le 1{,}64$ (p<0,05) și $\varphi^*_{cr} \le 2{,}31$ (p<0,01).

- Dacă $\varphi^*_{exp} < \varphi^*_{cr}$, atunci cu o probabilitate 1 - p se acceptă ipoteza H_0 privind nesemnificația statistică a diferențelor din loturile studiate. Adică proporțiile subiecților care au obținut „efectul” în ambele eșantioane nu diferă statistic. Cota parte a subiecților care au realizat „efectul” în grupa experimentală (GE) **nu este semnificativ mai mare statistic** decât în grupa de control (CG).
- Dacă $\varphi^*_{exp} > \varphi^*_{cr}$, atunci ipoteza H_0 este respinsă cu probabilitatea de 1 - p, iar ipoteza H_1 este acceptată. Cota parte a subiecților care au realizat „efectul” în grupa experimentală (GE) **este semnificativ mai mare statistic** decât în grupa de control (CG).

Pentru experimentele psihopedagogice se consideră o probabilitate de *eroare admisibilă* **p** = 0,05 și distribuția datelor pe eșantion normală. Deoarece **p** = 0,05 atunci *intervalul de încredere*, probabilitatea unei decizii corecte, este de cel puțin 1- **p** sau 95%.

Fie date două eșantioane independente între ele și cu valori independente $\mathbf{X}_1$, $\mathbf{X}_2$, unde $\mathbf{X}_1$ și $\mathbf{X}_2$ – reprezintă nivelul de rezolvare a sarcinilor evaluării pentru două grupe de studenți. Considerăm $\mathbf{X}_1$ – grupa experimentală (GE) și $\mathbf{X}_2$ – grupa de control (GC).

De oarece grupele de studenți sunt mici după număr, nu este simplu de demonstrat statistic semnificația diferențelor $\mathbf{X}_1$ și $\mathbf{X}_2$ din perspectiva efectului.

3. Metode de rezolvare a unor exemple practice

În cazul când eșantioanele sunt mici, **criteriul** φ^* **- Fischer** [2] este destul de eficient. În continuare vom analiza următoarele exemple.

Exemplu 1. Fie date două eșantioane cu valori pe segmentul [0, 100]. Fie $\mathbf{X}_1$={100, 100, 100, 95, 29}, cu un număr de valori $\mathbf{n}_1$=5 și $\mathbf{X}_2$={ 37, 67, 4, 4, 45, 83, 71, 94, 100}, cu un număr de valori $\mathbf{n}_2$=9; unde $\mathbf{X}_1$ și $\mathbf{X}_2$ – reprezintă nivelul de rezolvare a sarcinilor evaluării pentru două grupe de studenți, $\mathbf{X}_1$ – grupa experimentală (GE) și $\mathbf{X}_2$ – grupa de control (GC).

Folosind testul Fisher cu transformare unghiulară φ*, să se stabilească dacă diferența dintre șirurile $\mathbf{X}_1$ și $\mathbf{X}_2$ este semnificativă în punctul de efect egal cu $\boldsymbol{f_{efect}} = 55$.

Soluție. Pentru determinarea semnificației „efectului” după testul φ*-Fischer urmăm pașii:

1. Stabilim ipotezele:

H_0: Cota parte a studenților care au acumulat 55 de puncte în GE nu este mai mare decât în GC.

H_1: Cota parte a studenților care au acumulat 55 de puncte în GE este mai mare decât în GC.

P_2. În eșantionul $\mathbf{X}_1$ se enumeră datele mai mari și egale ca $\boldsymbol{f_{efect}}$, obținem $N_{X_1} = 4$. Calculăm procentul $P_{X_1} = \frac{N_{X_1}}{n_1} = \frac{4}{5} = 0{,}80$.

P_3. În eșantionul $\mathbf{X}_2$ se enumeră datele mai mari și egale ca $\boldsymbol{f_{efect}}$, obținem $N_{X_2} = 5$. Calculăm procentul $P_{X_2} = \frac{N_{X_2}}{n_2} = \frac{5}{9} = 0{,}56$.

P_4. În eșantionul $\mathbf{X}_1$ se enumeră datele mai mici decât $\boldsymbol{f_{efect}}$, obținem $N'_{X_1} = 1$. Calculăm procentul $P'_{X_1} = \frac{N'_{X_1}}{n_1} = \frac{1}{5} = 0{,}20$.

P_5. În eșantionul $\mathbf{X}_2$ se enumeră datele mai mici decât $\boldsymbol{f_{efect}}$, obținem $N'_{X_2} = 4$. Calculăm procentul $P'_{X_2} = \frac{N'_{X_2}}{n_2} = \frac{4}{9} = 0{,}44$.

Rezultatele obținute le includem în tabelul 1.

Tabelul 1. Distribuirea „efectului” asupra eșantioanelor

Grupa	Caracteristica studiată				Numărul de observații din eșantion
	Există efect, SCOR ≥ 55		Nu există efect, SCOR < 55		
	Numărul de observații din eșantion	În procente	Numărul de observații din eșantion	În procente	
X_1	4	80%	1	20%	5
X_2	5	56%	4	44%	9
Total	9		5		14

P_6. Se calculează $\varphi_1 = 2 \cdot \arcsin\left(\sqrt{P_{X_1}}\right) = 2 \cdot \arcsin\left(\sqrt{0{,}80}\right) = 2{,}214$.

P_7. Se calculează $\varphi_2 = 2 \cdot \arcsin\left(\sqrt{P_{X_2}}\right) = 2 \cdot \arcsin\left(\sqrt{0{,}56}\right) = 1{,}691$.

P_8. Se calculează $\varphi^*_{exp} = (\varphi_2 - \varphi_1) \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = (2{,}214 - 1{,}691) \cdot \sqrt{\frac{5 \cdot 9}{5+9}} = 0{,}94$.

Construim graficul testului φ*-Fischer pentru condițiile examinate în **Figura 1**.

Luând în considerare intervalele critice $\varphi^*_{cr} = 1{,}64$ pentru p=0,05 și $\varphi^*_{cr} = 2{,}31$ pentru p=0,01, obținem $\varphi^*_{exp} < \varphi^*_{cr}$. Prin urmare este adevărată ipoteza H_0.

Pentru o analiză mai detaliată a experimentului se calculează, în punctele “efect” de la 1 la 100, testul φ*-Fischer. Rezultatele obținute sunt prezentate grafic în Fig. 1.

Prin urmare, punctul de efect 55 nu este semnificativ. Pentru valorile:

a) pe segmentul [5, 29], obținem $\varphi^*_{exp} = 1.75$ cu o probabilitate de ρ=0.04;

b) pe segmentul [72, 83], obținem $\varphi^*_{exp} = 1.77$ cu o probabilitate de ρ=0.038;

c) pe segmentul [84, 94], obținem $\varphi^*_{exp} = 2{,}22$ cu o probabilitate de ρ=0.013.

Pentru punctul de efect 95, obținem $\varphi^*_{exp} = 2{,}76$ cu o probabilitate de ρ=0.001. Iar pe segmentul [96, 100], obținem $\varphi^*_{exp} = 1{,}96$ cu probabilitatea de ρ=0.025.

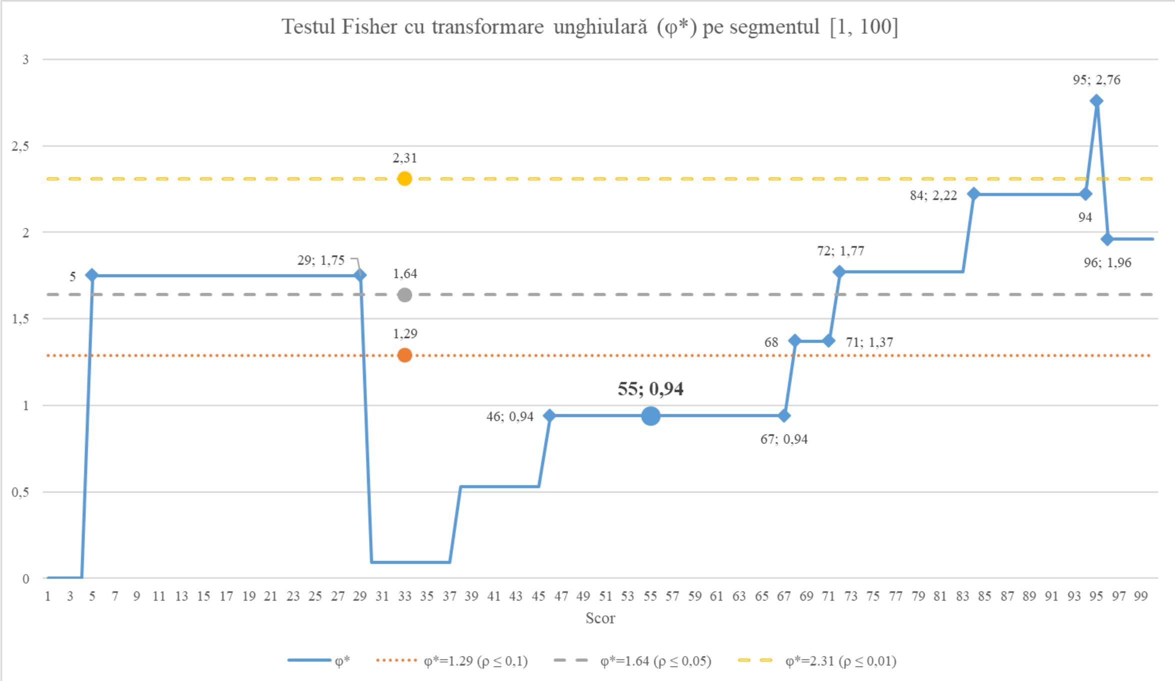

Fig. 1. Testul Fisher cu transformare unghiulară (φ*) pe segmentul [1, 100]

Exemplu 2. Să rezolvăm exemplul 1 luând în considerare eșantioanele: $\mathbf{X}_1$={ 100, 100, 100, 95, 29, 80, 90, 80, 70, 80, 80, 80}, cu un număr de valori $\mathbf{n}_1$=12 și $\mathbf{X}_2$={ 37, 67, 4, 4, 45, 83, 71, 94, 100}, cu un număr de valori $\mathbf{n}_2$=9.

Soluție. Pentru determinarea semnificației „efectului" (testul Fisher φ*) urmăm pașii:

1. Ipotezele rămân aceleași:

H$_0$: Cota parte a studenților care au acumulat 55 de puncte în GE nu este mai mare decât în GC.

H$_1$: Cota parte a studenților care au acumulat 55 de puncte în GE este mai mare decât în GC.

P$_2$. În eșantionul $\mathbf{X}_1$ se enumeră datele mai mari și egale ca $\boldsymbol{f_{efect}}$, obținem $N_{X_1} = 11$. Calculăm procentul $P_{X_1} = \frac{N_{X_1}}{n_1} = \frac{11}{12} = 0{,}92$.

P$_3$. În eșantionul $\mathbf{X}_2$ se enumeră datele mai mari și egale ca $\boldsymbol{f_{efect}}$, obținem $N_{X_2} = 5$. Calculăm procentul $P_{X_2} = \frac{N_{X_2}}{n_2} = \frac{5}{9} = 0{,}56$.

P$_4$. În eșantionul $\mathbf{X}_1$ se enumeră datele mai nici decât $\boldsymbol{f_{efect}}$, obținem $N'_{X_1} = 1$. Calculăm procentul $P'_{X_1} = \frac{N'_{X_1}}{n_1} = \frac{1}{12} = 0{,}08$.

P$_5$. În eșantionul $\mathbf{X}_2$ se enumeră datele mai nici decât $\boldsymbol{f_{efect}}$, obținem $N'_{X_2} = 4$. Calculăm procentul $P'_{X_2} = \frac{N'_{X_2}}{n_2} = \frac{4}{9} = 0{,}44$.

Rezultatele sunt organizate în tabelul 2.

Tabelul 2. Distribuirea „efectului” asupra eşantioanelor

Grupa	Caracteristica studiată				Numărul de observaţii din eşantion
	Există efect, SCOR ≥ 55		Nu există efect, SCOR < 55		
	Numărul de observaţii din eşantion	În procente	Numărul de observaţii din eşantion	În procente	
X_1	11	92%	1	8%	12
X_2	5	56%	4	44%	9
Total	16		5		21

P6. Se calculează $\varphi_1 = 2 \cdot \arcsin\left(\sqrt{P_{X_1}}\right) = 2 \cdot \arcsin\left(\sqrt{0{,}92}\right) = 2{,}568.$

P7. Se calculează $\varphi_2 = 2 \cdot \arcsin\left(\sqrt{P_{X_2}}\right) = 2 \cdot \arcsin\left(\sqrt{0{,}56}\right) = 1{,}691.$

P8. Se calculează $\varphi^*_{exp} = (\varphi_2 - \varphi_1) \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = (2{,}568 - 1{,}691) \cdot \sqrt{\frac{12 \cdot 9}{12+9}} = 1{,}99.$

În consecinţă relaţia $\varphi^*_{exp} < \varphi^*_{cr}$ este falsă, deoarece $\varphi^*_{exp} = 1{,}99$ cu o probabilitate ρ=0.023. Prin urmare este adevărată ipoteza H_1.

Să calculăm, în punctele considerate “efect” de la 1 la 100, testul φ*-Fisher. Rezultatele sunt prezentate grafic în Fig. 2.

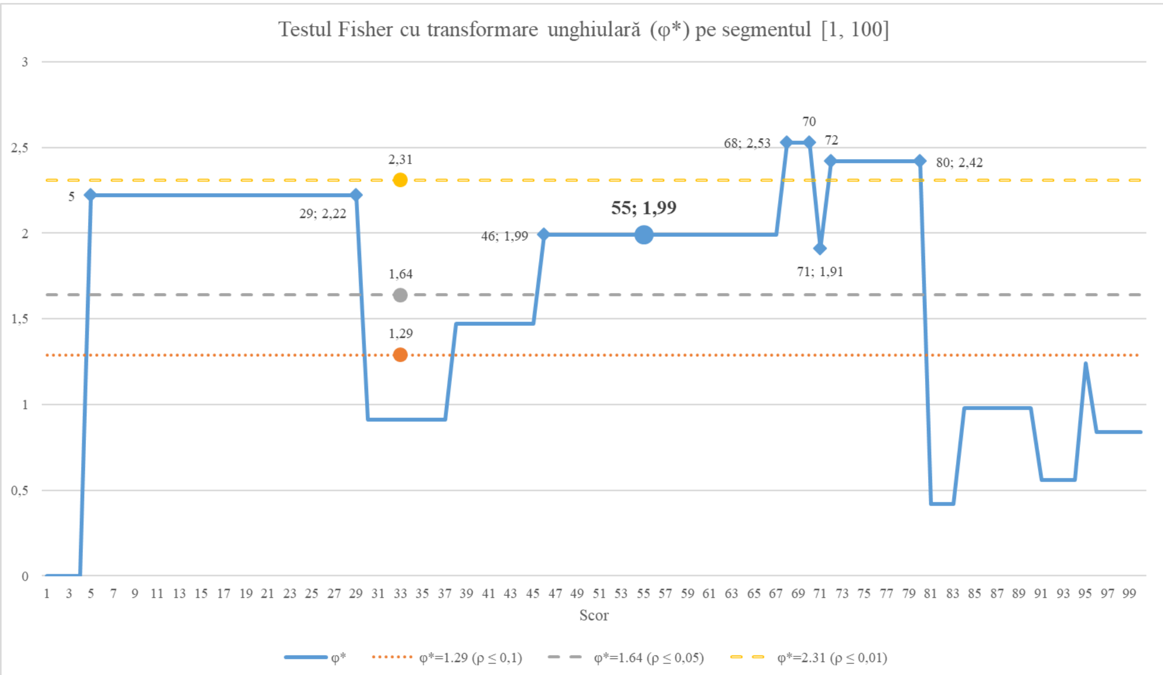

Fig. 2. Testul Fisher cu transformare unghiulară (φ*) pe segmentul [1, 100]

Prin urmare pentru valorile:

a) pe segmentul [5, 29], obţinem $\varphi^*_{exp} = 2.22$ cu o probabilitate de ρ=0.013;

b) pe segmentul [46, 67], obţinem $\varphi^*_{exp} = 1.99$ cu o probabilitate de ρ=0.023;

c) Pe segmentul [68, 70], obţinem $\varphi^*_{exp} = 2.53$ cu o probabilitate de ρ=0.004.

Pentru punctul de efect 71, obținem $\varphi^*_{exp} = 1{,}91$ cu o probabilitate de ρ=0.028. Iar pe segmentul [72, 80], obținem $\varphi^*_{exp} = 2.42$ cu probabilitatea de ρ=0.006.

Astfel, în exemplul 1, experimentul N1 a influențat semnificativ scorurile mici de pe segmentul [5, 29], și scorurile mari pe segmentul [72, 100]. Cu alte cuvinte experimentul N1 influențează semnificativ scorurile mici și mari. În exemplul 2, experimentul N2 a influențat semnificativ scorurile mici pe segmentul [5, 29], și scorurile medii, segmentul [46, 80]. Altfel spus experimentul N2 influențează semnificativ scorurile mici și medii.

Utilizarea criteriilor multifuncționale permite să fie determinată proporția populației în situația când observațiile din eșantionul examinat se caracterizează prin efectul de interes. Astfel, metodele neparametrice cresc în popularitate și influență din mai multe motive. Principalul motiv este că nu suntem constrânși la fel de mult ca atunci când folosim o metodă parametrică. Nu este nevoie să facem atâtea ipoteze cu privire la populația cu care lucrăm, cât și ceea ce trebuie să facem cu o metodă parametrică.

Articol elaborat în cadrul proiectului de cercetări științifice „Metodologia implementării TIC în procesul de studiere a științelor reale în sistemul de educație din Republica Moldova din perspectiva inter/transdisciplinarității (concept STEAM)", inclus în „Program de Stat" (2020-2023), Prioritatea IV: Provocări societale, cifrul 20.80009.0807.20.

Bibliografie

1. БОЛЬШЕВ, Л.Н.; СМИРНОВ, Н.В. Таблицы математической статистики. Москва, 1983. 416 с.
2. СИДОРЕНКО, Е.В. Методы математической обработки в психологии. С.-Пб.: «Речь», 2003. 350 с.
3. LABĂR, A. V. SPSS pentru științele educației. Metodologia analizei datelor în cercetarea pedagogică. POLIROM, 2008. 347 p.
4. ГУБЛЕР, Е. В. Вычислительные методы анализа и распознавания патологических последствий. Ленинград: Медицина, 1978. 295 с.