УДК 33.013: 004.852: 519.25) **DOI 10.5281/zenodo.3564704**

Джунь Йосип Володимирович д.фіз-мат.н., професор, завідувач, josif-june@rambler.ru, **Лотюк Юрій Георгійович,** к.пед.н., доцент, доцент кафедри математичного моделювання (Міжнародний економіко-гуманітарний університет імені академіка Степана Дем'янчука, м. Рівне), lotyuk@ukr.net

КРИТЕРІЙ РАНГОВОЇ КОРЕЛЯЦІЇ СПІРМЕНА І ОСОБЛИВОСТІ ЙОГО ЗАСТОСУВАННЯ У ПЕДАГОГІЧНИХ ДОСЛІДЖЕННЯХ

Анотація. *У статті розглянуто застосування непараметричного критерію Спірмена у педагогічних дослідженнях. Наведено формули для визначення тісноти зв'язку між ознаками, коли характер розподілу досліджуваної сукупності невідомий. Показано, що тісноту зв'язку в цьому випадку можна обчислити за допомогою непараметричних критеріїв. Визначено, що особливістю цих критеріїв є те, що тіснота зв'язку між ознаками визначається не за кількісними значеннями варіантів, а за допомогою порівняння їх рангів. Розкрито, що коефіцієнт кореляції рангів – це один з найпростіших показників тісноти зв'язку, його називають ранговим коефіцієнтом кореляції Спірмена. Обґрунтовано, що його суть полягає в тому, що парні спостереження двох взаємопов'язаних ознак (результативної і факторної) ранжируються а тіснота зв'язку визначається на основі близькості рангів.*

Ключові слова*: критерій Спірмена, статистичні закономірності, статистична вибірка, аналіз емпіричних даних в педагогіці і психології.*

Dzhun Yosyp Volodymyrovych, Doctor of Physical and Mathematical Sciences, Professor, Head of the Department, josif-june@rambler.ru **Lotiuk Yurii Heorhiiovych,** Ph.D in Pedagogic Sciences, Associate Professor, Associate Professor of the Department of Mathematical Modeling (Academican Stepan Demianchuk International University of Economics and Humanities, Rivne), lotyuk@ukr.net

CRITERIA OF SPEARMAN'S RANK CORRELATION AND PECULIARITIES OF ITS USE IN PEDAGOGICAL RESEARCH

Abstract.

Introduction. Introduction. *Application of nonparametric Spearman criterion in pedagogical studies is considered in the article. Formulas for determining the closeness of the relationship between the traits, when the nature of the distribution of the studied population is unknown, the connection tightness can be calculated using nonparametric criteria. The peculiarity of these criteria is that the close relationship between the traits is determined not by the* *quantitative values of the variants, but by the comparison of their ranks. The smaller the differences between the ranks, the closer the relationship between the traits. The rank correlation coefficient is one of the simplest indicators of communication closeness, called the Spearman's rank correlation coefficient. Its essence is that paired observations of two interrelated features (productive and factorial) are ranked. Tightness is determined by the closeness of the ranks.*

Purpose. *To introduce the teachers and psychologists with the modern method of discovering statistical laws and the results of pedagogical and psychological surveys and research.*

Methods. *Non-parametric Spearman's criterion is used to establish statistical laws, which, based on empirical data, reflects the close relationship between the traits.*

Results. *The authors have developed a method of establishing a close relationship between the productive and factor traits based on empirical data used in pedagogy and psychology.*

Most of these laws are expressed in the form of statistical distributions. However, finding a statistical distribution that reflects the studied pattern is not known by many teachers, because the methodology developed by Spearman is intended for mathematics professionals, and few teachers are familiar with it. Therefore, the statistical law algorithm is set out in detail, which allows to apply this method in pedagogical practice mathematically and at a modern level.

The authors have developed recommendations that provide prompt, mathematically sound, computer-based solutions to the problem of establishing a close relationship between traits, subject to the studied pedagogical phenomenon based on empirical data.

Originality. *The use of non-parametric Spearman's criterion in pedagogical and psychological research at the appropriate mathematical level is first highlighted, and a computer method to determine the relationship between empirical evidence is proposed. It allows to search at a modern, mathematical level with the utmost automation of establishing the closeness of connection between the features using the recommended computer programs in MathCad mathematical package.*

Conclusion. *The described method of determining the closeness of the relationship between features when the nature of the distribution of the studied population is unknown, is a new and useful theoretical and practical tool that greatly simplifies and facilitates statistical research of teachers and psychologists. This significantly facilitates and accelerates the study of statistical samples necessary for pedagogy and psychology, based on the application of the latest mathematical approaches and their software.*

Key words: *Spearman's criterion, establishment of statistical regularities, statistical sampling, analysis of empirical data in pedagogy and psychology.*

Непараметричний критерій Спірмена має широке застосування в педагогічних дослідженнях внаслідок своєї простоти і зручності в користуванні.

Проте широке застосування формула для обчислення коефіцієнту Спірмена має лише у випадку наявності неоднакових рангів. Метод застосування критерію Спірмена у випадку існування однакових рангів у вибірках, маловідомий, але настільки ж ефективний [1].

Метою нашого дослідження є розглянути усі наявні версії застосування непараметричного критерію Спірмена.

Коефіцієнт рангової кореляції Спірмена

(перша версія критерію).

У випадку дослідження залежності між вибірками, які мають істотно не Гаусів розподіл, обчислюють коефіцієнт рангової кореляції Спірмена, розроблений ним у 1904 році [2] за такою формулою:

$$\rho = 1 - \frac{6 \cdot \sum d_i^2}{n^3 - n}, \tag{1}$$

де $d = R_x - R_y$ – ранги елементів 1-ї і 2-ї вибірок; n – число пар R_x, R_y.

Коефіцієнти кореляції рангів, як і лінійний коефіцієнт кореляції набуває значень в межах:

$$-1 \le \rho \le 1$$

Якщо 2 рядки рангів R_x i R_y збігаються, то $\sum d^2 = 0$, тоді $\rho = 1$. При зворотньому зв'язку ранги двох рядків розташовані у зворотному порядку $\rho = -1$.

Приклад 1: У табл. 1 наведені результати IQ тестування 10 магістрів спеціальності 014 Середня освіта (інформатика) та їхні оцінки на випускному державному іспиті. Застосувавши коефіцієнт Спірмена, визначимо у якій мірі корельовані ці результати. Для цього в колонці 4 виставимомо рейтинги R_x для IQ тестувань, а в колонку 5 записуємо рейтинги відповідей.

В колонці 6 запишемо різниці рейтингів, а в колонці 7 їх квадрати. Підставляємо знайдену $\sum d^2$ в формулу (1) і обчислюємо коефіцієнт ρ рангової кореляції Спірмена (рис. 1):

$$\rho = 1 - \frac{6 \cdot 10}{1000 - 10} = 0,939, \tag{2}$$

Отриманий результат свідчить про сильний зв'язок рейтингів R_x i R_y. Перевіримо істотність зв'язку. Це робиться на основі нерівності:

$$|\rho| > \rho_{kp,\alpha,n} \tag{3}$$

Таблиця 1

Допоміжні розрахунки для обчислення коефіцієнта рангової кореляції Спірмена

Код студента	Результати IQ тесту	Результати державного іспиту	R_x	R_y	d_i	d_i^2
1	*2*	*3*	*4*	*5*	*6*	*7*
А	91	80	7	5	2	4
Б	125	88	3	3	0	0
В	111	82	4	4	0	0
Г	139	92	1	2	-1	1
Д	80	70	8	9	-1	1
Е	130	100	2	1	1	1
Ж	100	75	6	7	1	1
З	60	63	10	10	0	0
І	107	78	5	6	-1	1
К	75	73	9	8	1	1
						$\sum d_i^2 = 10$

ORIGIN := 1 $\quad n := 10 \quad i := 1..n$

$$\mathrm{Tabl1R} := \begin{pmatrix} 7 & 3 & 4 & 1 & 8 & 2 & 6 & 10 & 5 & 9 \\ 5 & 3 & 4 & 2 & 9 & 1 & 7 & 10 & 6 & 8 \end{pmatrix} \qquad \mathrm{Tabl1d}_i := \mathrm{Tabl1R}_{1,i} - \mathrm{Tabl1R}_{2,i}$$

Tabl1d =

	1
1	2
2	0
3	0
4	-1
5	-1
6	1
7	-1
8	0
9	-1
10	1

$$\rho := 1 - \frac{6 \cdot \left[\sum_i \left(\mathrm{Tabl1d}_i\right)^2\right]}{n^3 - n} \qquad \rho = 0.939$$

Рис. 1. Робочий аркуш MathCad визначення зв’язку рейтингів R_x і R_y

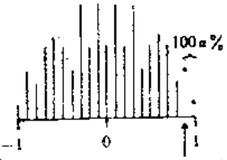
У відповідності з табл. 2 25 % двостороння критична область є

$$|\rho| > \rho_{kp,\alpha,n} = \rho_{kp,2,5\%,n} = 0,648 \qquad (4)$$

Як бачимо з формули (4), отримане за формулою (2) значення коефіцієнта Спірмена є значимим на 5 % рівні. Отже, ми відхиляємо нульову гіпотезу $H_0 : \rho = 0$ і робимо висновок, що $H_0 : \rho \neq 0$.

Таблиця 2

Верхні 100α %-ні точки розподілу коефіцієнта рангової кореляції Спірмена. Критична область включає табличне значення критерію [3, с. 212]



Табличне значення критерію

Обсяг вибірки n	$\alpha = 0,05$	0,025	0,01
5	0,900	1,000	1,000
6	0,828	0,885	0,942
7	0,714	0,785	0,892
8	0,643	0,738	0,833
9	0,600	0,683	0,783
10	0,563	0,648	0,745
11	0,527	0,609	0,700
12	0,497	0,580	0,671
13	0,478	0,555	0,643
14	0,459	0,534	0,622
15	0,443	0,518	0,600
16	0,427	0,500	0,582
17	0,412	0,485	0,564
18	0,399	0,472	0,548
19	0,390	0,458	0,533
20	0,379	0,445	0,520
21	0,369	0,435	0,508
22	0,360	0,424	0,496
23	0,352	0,415	0,485
24	0,344	0,406	0,475
25	0,336	0,398	0,465
26	0,330	0,389	0,456
27	0,324	0,382	0,448
28	0,318	0,375	0,440
29	0,311	0,369	0,432
30	0,306	0,362	0,425

Застосування коефіцієнту Спірмена у випадку наявності однакових рангів (*друга версія критерію)*

Для оцінки сили зв'язку між x і y у випадку, коли між x та y існує нелінійний зв'язок, або вибірки не розподілені за нормальним законом і при наявності однакових рангів, коефіцієнт рангової кореляції Спірмена розраховують за формулою [4, с.73–75]:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2 + T_x + T_y}{n^3 - n} , \tag{5}$$

де n – кількість пар вибіркових даних; $d_i = R_x - R_y$; T_x, T_y – поправки, що пов'язані з однаковими рангами:

$$T_x = \frac{\sum_{i=1}^{L_x} (T_{x_i}^3 - T_{x_i})}{12}; \quad T_y = \frac{\sum_{i=1}^{L_y} (T_{y_i}^3 - T_{y_i})}{12}, \tag{6}$$

де L_x та L_y – кількість груп однакових рангів; T_{x_i}, T_{y_i} – розміри i -тих зв'язок (кількість елементів в них).

Приклад 2: Вивчимо залежність між середнім балом успішності учня за рік по атестату і його емоційним відношенням до навчання в школі (табл. 3).

Потрібно оцінити силу зв'язку між досліджуваними факторами за коефіцієнтом кореляції Спірмена і перевірити його статистичну значущість.

Знаходимо поправки, що пов'язані з однаковими рангами в стовпчику рангів R_x є дві групи однакових рангів – в першій 2 елементи, в другій теж 2 елементи. Отже:

$$L_x = 2 , \; T_{x_1} = 2 , T_{x_2} = 2 , \tag{7}$$

В стовпці R_x є також 2 групи однакових рангів, в 1-й з них 2 елементи, в 2-й – три елементи. Отже:

$$L_y = 2 , \; T_{y_1} = 2 , T_{y_2} = 3 , \tag{8}$$

Підставивши (7) і (8) у формулу (6) і, обчисливши поправки, отримаємо:

Таблиця 3

Розрахунки до обчислення коефіцієнта рангової кореляції Спірмена з врахуванням поправок, пов'язаних з однаковими рангами

Код учня	Середній бал по атестату	Емоційне відношення до навчання, бали	R_x	R_y	d_i	d_i^2
1	2	3	4	5	6	7
А	9,0	16	12	12	0	0
Б	7,9	12	10	9	-1	1
В	6,5	5	4,5	2,5	-2	4
Г	6,0	4	1	1	0	0
Д	9,5	17	13	13	0	0
Е	6,3	6	3	4	1	1
Ж	7,3	15	8,5	11	2,5	6,25
З	6,5	7	4,5	5	0,5	0,25
І	10,1	13	14	10	-4	16
К	11,5	20	15	15	0	0
Л	8,4	10	11	7	-4	16
М	7,1	10	7	7	0	0
Н	6,9	10	6	7	1	1
О	6,2	5	2	2,5	0,5	0,25
П	7,3	19	8,5	14	5,5	30,25
						$\sum d_i^2 = 76$

$$T_x = \frac{\sum_{i=1}^{L_x}(T_{x_i}^3 - T_{x_i})}{12} = \frac{(2^3-2)+(2^3-2)}{12} = 1 \qquad (9)$$

$$T_y = \frac{\sum_{i=1}^{L_y}(T_{y_i}^3 - T_{y_i})}{12} = \frac{(2^3-2)+(3^3-3)}{12} = 1 \qquad (10)$$

Знаходимо коефіцієнт рангової кореляції Спірмена за формулою (5):

$$\rho = 1 - \frac{6\cdot\sum_{i=1}^{n} d_i^2 + T_x + T_y}{n^3 - n} = 1 - \frac{6\cdot 76 + 1 + 2,5}{15^3 - 15} = 0,863 \qquad (11)$$

За значенням отриманого ρ можна зробити висновок, що між x і y існує сильний прямий зв’язок.

Перевіремо статистичну значимість знайденого ρ . Для цього розрахуємо τ -статистику за формулою (рис. 2):

$$\tau = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} = \frac{0{,}863\cdot\sqrt{15-2}}{\sqrt{1-0{,}863^2}} = 6{,}159 \tag{12}$$

$$n := 15 \qquad\qquad i := 1..n$$

$$Tabl2R := \begin{pmatrix} 12 & 10 & 4.5 & 1 & 13 & 3 & 8.5 & 4.5 & 14 & 15 & 11 & 7 & 6 & 2 & 8.5 \\ 12 & 9 & 2.5 & 1 & 13 & 4 & 11 & 5 & 10 & 15 & 7 & 7 & 7 & 2.5 & 14 \end{pmatrix}$$

$$Tabl2d_i := Tabl2R_{2,i} - Tabl2R_{1,i}$$

Tabl2d =

	1
1	0
2	-1
3	-2
4	0
5	0
6	1
7	2.5
8	0.5
9	-4
10	0
11	-4
12	0
13	1
14	0.5
15	5.5

$$T_x := 1 \qquad T_y := 1$$

$$\rho := 1 - \frac{6\cdot\left[\sum_i \left(Tabl2d_i\right)^2\right] + T_x + T_y}{n^3 - n} \qquad \rho := 0.863$$

$$\tau := \frac{\rho\cdot\sqrt{n-2}}{\sqrt{1-\rho^2}} \qquad \tau = 6.159$$

Рис. 2. Робочий аркуш MathCad перевірки статичтичної значимості рангової кореляції Спірмена

Обираємо рівень значущості $\alpha = 5\%$ оскільки це критерій двосторонній, то по табл. 4 [5, с. 240] для q =2,5 і числа ступенів свободи $n = l - 2 = 15 - 2 = 13$ отримаємо, що $\left| t_{kp,2,5\%,13} \right| = 2{,}160 < t = 6{,}159$, тобто, з імовірністю 95 % обчислений коефіцієнт ρ свідчить про значимий зв'язок між факторами x та y .

Таблиця 4

Процентні точки розподілу Стьюдента

n \ Q	40%	25%	10%	5%	2,5%	1%	0,5%	0,25%	0,1%	0,05%
1	0,3249	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	318,3088	636,6192
2	2887	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	22,3271	31,5991
3	2767	7649	6377	3534	3,1824	4,5407	5,8409	7,4533	10,2145	12,9240
4	2707	7407	5332	1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103
5	2672	7267	4759	0150	5706	3649	0321	4,7733	5,8934	6,8688
6	0,2648	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	2632	7111	4149	8946	3646	2,9980	4995	0293	4,7853	4079
8	2619	7064	3968	8595	3060	8965	3554	3,8325	5008	0413
9	2610	7027	3830	8331	2622	8214	2498	6897	2968	4,7809
10	2602	6998	3722	8125	2281	7638	1693	5814	1437	5869
11	0,2596	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	2590	6955	3562	7823	1788	6810	0545	4284	3,9296	3178
13	2586	6938	3502	7709	1604	6503	0123	3725	8520	2208
14	2582	6924	3450	7613	1448	6245	2,9768	3257	7874	1405
15	2579	6912	3406	7530	1314	6025	9467	2860	7328	0728
16	0,2576	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	2573	6892	3334	7396	1098	5669	8982	2224	6458	3,9651
18	2571	6884	3304	7341	1009	5524	8784	1966	6105	9216
19	2569	6876	3277	7291	0930	5395	8609	1737	5794	8834
20	2567	6870	3253	7247	0860	5280	8453	1534	5518	8495
21	0,2566	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193
22	2564	6858	3212	7171	0739	5083	8188	1188	5050	7921
23	2563	6853	3195	7139	0687	4999	8073	1040	4850	7676
24	2562	6848	3178	7109	0639	4922	7969	0905	4668	7454
25	2561	6844	3163	7081	0595	4851	7874	0782	4502	7251
26	0,2560	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066
27	2559	6837	3137	7033	0518	4727	7707	0565	4210	6896
28	2558	6834	3125	7011	0484	4671	7633	0469	4082	6739
29	2557	6830	3114	6991	0452	4620	7564	0380	3962	6594
30	2556	6828	3104	6973	0423	4573	7500	0298	3852	6460
32	0,2555	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,3653	3,6218
34	2553	6818	3070	6909	0322	4411	7284	0020	3479	6007
36	2552	6814	3055	6883	0281	4345	7195	2,9905	3326	5821
38	2551	6810	3042	6860	0244	4286	7116	9803	3190	5657
40	2550	6807	3031	6839	0211	4233	7045	9712	3069	5510
42	0,2550	0,6804	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,2960	3,5377
44	2549	6801	3011	6802	0154	4141	6923	9555	2861	5258
46	2548	6799	3002	6787	0129	4102	6870	9488	2771	5150
48	2548	6796	2994	6772	0106	4066	6822	9426	2689	5051
50	2547	6794	2987	6759	0086	4033	6778	9370	2614	4960
55	0,2546	0,6790	1,2971	1,6730	2,0040	2,3961	2,6682	2,9247	3,2561	3,4764
60	2545	6786	2958	6706	0003	3901	6603	9146	2317	4602
65	2544	6783	2947	6686	1,9971	3851	6536	9060	2204	4466
70	2543	6780	2938	6669	9944	3808	6479	8987	2108	4350
80	0,2542	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163
90	2541	6772	2910	6620	9867	3685	6316	8779	1833	4019
100	2540	6770	2901	6602	9840	3642	6259	8707	1737	3905
120	2539	6765	2886	6577	9799	3578	6174	8599	1595	3735
150	2538	6761	2872	6551	9759	3515	6090	8492	1455	3566
200	2537	6757	2858	6525	9719	3451	6006	8385	1315	3398
250	2536	6755	2849	6510	9695	3414	5956	8322	1232	3299
300	2536	6753	2844	6499	9679	3388	5923	8279	1176	3233
400	2535	6751	2837	6487	9659	3357	5882	8227	1107	3150
500	0,2535	0,6750	1,2832	1,6479	1,9647	2,3338	2,5857	2,8195	3,1066	3,3101

Оскільки значення $t = 6{,}159 >> t_{kp,2,5\%,13} = 2{,}160$, то можна стверджувати, що t більше t_{kp} навіть для крайніх, можливих рівнів значущості, поданих в табл. 4. Наприклад, для $\alpha = 0{,}1\%$ маємо:

$$t = 6{,}159 > \left| t_{kp,0,05\%,13} \right| = 4{,}221,$$

тобто, навіть з імовірністю 99,9% можна стверджувати, що зв'язок між x_i та y_i існує.

Список використаних літературних джерел

1. Джунь Й. В. Застосування математичних методів і інформаційних технологій в педагогічних дослідженнях. Практикум. Рівне, 2019.

2. Spearman C. The proof and measurement of association between two things. Amer J. Psyhol, 1904, 15, 8.

3. Поллард Дж. Справочник по вычислительным методам статистики. Пер. с англ. В. С. Занадворова; Под. ред. и с предисл. Е. И. Четыркина. М.: Финансы и статистика, 1982. 344 с.

4. Сенча І. А. Економіко-математичні моделі та методи проектного менеджменту. Ч. 1. Статистичні моделі та методи. Одеса: ОРІДУ НАДУ, 2009. 101 с.

5. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. Москва: ВЦ АН СССР, 1968. 475 с.

References

1. Dzhun Y. V. (2019). Zastosuvannia matematychnykh metodiv i informatsiinykh tekhnolohii v pedahohichnykh doslidzhenniakh. Praktykum (In Ukrainian).

2. Spearman C. (1904). The proof and measurement of association between two things. – Amer J. Psyhol, 15, 8.

3. Pollard Dzh. (1982). Spravochnik po vychislitelnym metodam statistiki Moskva: Finansy i statistika (In Russian).

4. Sencha I. A. (2009) Ekonomiko-matematychni modeli ta metody proektnoho menedzhmentu. Ch. 1. Statystychni modeli ta metody. Odesa: ORIDU NADU (In Ukrainian).

5. Bolshev L. N., Smirnov N. V. (1968) Tablitsy matematicheskoy statistiki. Moskva: VTs AN SSSR (In Russian).

Рецензент: д.геогр.н., професор Калько А. Д.