Лекція 19. КОРЕЛЯЦІЯ. ВИБІРКОВИЙ КОЕФІЦІЄНТ КОРЕЛЯЦІЇ. Ранг

19.1. Поняття кореляції, коефіцієнта кореляції

Кореляційною називають таку залежність між ознаками X та Y, коли при зміні однієї з ознак змінюється середнє значення іншої.

Кореляційний аналіз використовують для визначення сили статистичних зав'язків між двома наборами даних.

Форму зв'язку між змінними X і Y можна встановити, застосовуючи *кореляційні поля*.

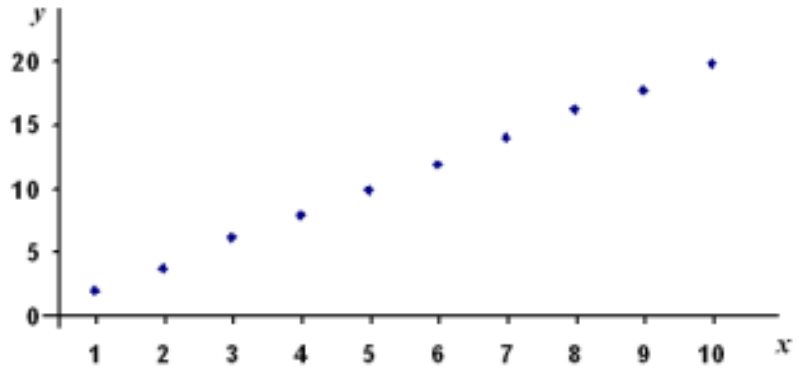
На основі аналізу кореляційного поля можна вирішити питання про наявність чи відсутність залежності між ознаками, прослідкувати характер залежності (лінійна, нелінійна, функціональна чи статистична) та її тенденцію (додатну чи від'ємну).

Приклад 19.1. Задано двовимірну вибірку

x_i	1	2	3	4	5	6	7	8	9	10
y_i	2	3,7	6,2	7,9	9,9	12	14,1	16,3	17,8	19,9

Побудувати кореляційне поле.

Розв'язок. Відкладемо на площині xOy точки з координатами (1;2), (2;3,7), (3;6,2) та ін. Отримаємо кореляційне поле для значень ознак X та Y, на якому чітко видно лінійну залежність Y від X.



Основною задачею кореляційного аналізу є виявлення зв'язку між випадковими величинами та оцінка його тісноти.

Для оцінки тісноти кореляційної залежності між ознаками служить *коефіцієнт кореляції*.

Коефіцієнт кореляції Пірсона між двома змінними дорівнює коваріації двох змінних, або сумі добутків відхилень, поділеній на добуток їх стандартних відхилень.

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}} \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{\operatorname{cov}(X, Y)}{\sigma_x \cdot \sigma_y}$$

Коефіцієнт кореляції Спірмена визначається як коефіцієнт кореляції Пірсона між ранжованими змінними. Для вибірки обсягу n множини X_i, Y_i перетворюються в ряди x_i, y_i та обчислюється наступним чином.

$$\rho = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

Коефіцієнт кореляції Кендалла є мірою рангової кореляції, тобто подібності упорядкування даних, коли вони упорядкуванні за своєю величиною.

$$\tau = \frac{s_1 - s_2}{0{,}5n(n-1)}$$

де s_1 – кількість узгоджених пар, s_2 – кількість неузгоджених пар.

Коефіцієнт кореляції Пірсона

Вибірковим коефіцієнтом кореляції $\left(r_{xy}\right)$ називається відношення різниці між математичним сподіванням добутку випадкових величин та добутку їх математичних сподівань до добутку середніх квадратичних відхилень цих випадкових величин:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{\operatorname{cov}(X, Y)}{\sigma_x \cdot \sigma_y},$$

де $\operatorname{cov}(X, Y)$ - коваріація.

Приклад 19.2. Задано двовимірну вибірку об'ємом $n=11$.

x_i	-4	-3	-2	-1	0	1	2	3	4	5	6
y_i	25	16	9	4	1	0	1	4	9	16	25

Обчислити коваріацію та вибірковий коефіцієнт кореляції. Перевірити, чи існує залежність між ознаками *X* та *Y*.

Розв'язок. Знайдемо значення $\overline{x}$ та $\overline{y}$.

$$\overline{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{-4+(-3)+(-2)+(-1)+0+1+2+3+4+5+6}{11}=1;$$

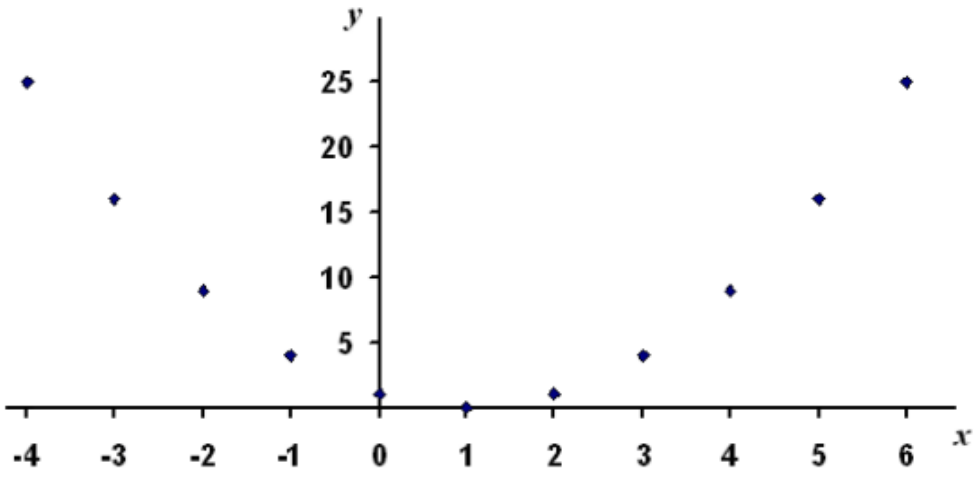
$$\overline{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{25+16+9+4+1+0+1+4+9+16+25}{11}=10.$$

Обчислимо коваріацію за формулою для незгрупованих даних

$$\operatorname{cov}(X,Y)=\frac{\sum_{i=1}^{n}x_iy_i}{n}-\overline{x}\cdot\overline{y}=\frac{-4\cdot25+(-3)\cdot16+(-2)\cdot9+(-1)\cdot4+0\cdot1}{11}+$$
$$+\frac{1\cdot9+2\cdot1+3\cdot4+4\cdot9+5\cdot16+6\cdot25}{11}-1\cdot10=0.$$

Вибірковий коефіцієнт кореляції

$$r_{xy}=\frac{\operatorname{cov}(X,Y)}{\sigma_x\cdot\sigma_y}=\frac{0}{\sigma_x\cdot\sigma_y}=0.$$



На практиці найчастіше використовують іншу модифікацію формули вибіркового коефіцієнта кореляції, за якою r_{xy} знаходять без округлення даних, пов'язаних з розрахунком середніх і відхилень від них:

$$r_{xy} = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i \right)^2}}$$

Властивості вибіркового коефіцієнта кореляції

1. r_{xy} за абсолютною величиною ≤ 1.

2. Якщо r_{xy}=0, то X та Y не пов'язані лінійною кореляційною залежністю, але інша залежність при цьому може мати місце.

3. Якщо $|r_{xy}| = 1$, то X та Y пов'язані лінійною кореляційною залежністю.

4. Якщо при зростанні однієї випадкової величини інша має тенденцію до зростання, то маємо позитивну кореляцію $r_{xy} > 0$; якщо при зростанні однієї випадкової величини інше спадає, то маємо обернену кореляцію $r_{xy} < 0$. Залежність тим ближча до лінійної, чим $|r_{xy}|$ ближчий до одиниці.

Для оцінки *тісноти нелінійного кореляційного зв'язку* використовують такі характеристики, як *вибіркове кореляційне відношення* Y по X $\left(\eta_{yx}\right)$ та вибіркове кореляційне відношення X по Y $\left(\eta_{xy}\right)$. При цьому згадаємо правило додавання дисперсій: $D_{заг.} = D_{сер.груп.} + D_{між.груп.}$ (загальна дисперсія змінної дорівнює сумі середньої групових дисперсій (залишкової) та міжгрупової дисперсії).

19.2. Перевірка гіпотез про значущість коефіцієнтів кореляції

Перевірка гіпотези про значущість вибіркового коефіцієнта кореляції.

Хочемо перевірити: чи впливає кількісний фактор на кількісний відгук.

H_0: фактор не впливає на відгук.

Математичний зміст критерію: оцінка величини коефіцієнта кореляції – міри лінійного зв'язку між фактором та відгуком.

Для того, щоб при рівні значущості α перевірити нульову гіпотезу $H_0 : r_Г = 0$ про рівність нулю генерального коефіцієнта кореляції при конкуруючій гіпотезі $H_1 : r_Г \neq 0$, необхідно обчислити спостережене значення критерію

$$T_{спост} = r_в \sqrt{\frac{n-2}{1-r_в^2}},$$

де $r_в$ - вибірковий коефіцієнт кореляції, n - об'єм вибірки.

Із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості α і числу ступенів вільності $k = n-2$ знаходять критичну точку $t_{кр}\left(\frac{\alpha}{2}; k\right)$.

Якщо $|T_{спост}| < t_{кр}$ - нульову гіпотези приймають. Якщо $|T_{спост}| > t_{кр}$ - нульову гіпотезу відкидають.

Якщо нульова гіпотеза правильна, то між досліджуваними ознаками *X* та *Y* лінійний зв'язок відсутній. Якщо нульова гіпотеза відхиляється, то між досліджуваними ознаками *X* та *Y* існує статистично значущий лінійний зв'язок. Сила і напрям лінійного зв'язку визначаються величиною вибіркового коефіцієнта кореляції.

Приклад 19.4. Провели дослідження залежності середнього часу обслуговування клієнтів банку від стажу роботи працівника.

Знайдений вибірковий коефіцієнт кореляції дорівнює $r_в = 0,7$. Об'єм вибірки складає $n = 10$.

При рівні значущості $\alpha = 0,1$ перевірити нульову гіпотезу $H_0 : r_Г = 0$ про рівність нулю генерального коефіцієнта кореляції. За альтернативну прийняти гіпотезу $H_1 : r_Г \neq 0$.

Розв'язання: Знайдемо спостережене значення критерію

$$T_{спост} = r_в \sqrt{\frac{n-2}{1-r_в^2}} = 0,7 \cdot \sqrt{\frac{10-2}{1-0,7^2}} = 2,77.$$

Із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості $\alpha = 0,1$ і числу ступенів вільності $k = 8$ знаходимо критичну точку $t_{кр}\left(\frac{\alpha}{2};k\right) = t_{кр}\left(\frac{0,1}{2};k\right) = t_{кр}\left(0,05;8\right) = 1,86$.

$|T_{спост}| > t_{кр}$ – нульову гіпотезу відкидаємо. Між середнім часом обслуговування клієнтів банку і стажем роботи працівника існує прямий лінійний зв'язок.

Перевірка гіпотези про значущість вибіркового коефіцієнта рангової кореляції Спірмена

Для того, щоб при рівні значущості α перевірити нульову гіпотезу $H_0 : \rho_Г = 0$ про рівність нулю генерального коефіцієнта рангової кореляції Спірмена при конкуруючій гіпотезі $H_1 : \rho_Г \neq 0$, необхідно обчислити критичну точку

$$T_{кр} = t_{кр}\left(\frac{\alpha}{2};k\right) \cdot \sqrt{\frac{1-\rho_в^2}{n-2}},$$

де $\rho_в$ - вибірковий ранговий коефіцієнт кореляції Спірмена, n - об'єм вибірки, $t_{кр}\left(\frac{\alpha}{2};k\right)$ - критична точка, яку знаходять із таблиці критичних точок

розподілу Стьюдента (додаток 6), за заданим рівнем значущості α і числу ступенів вільності $k = n - 2$.

Якщо $|\rho_в| < T_{кр}$ - нульову гіпотези приймають. Ранговий кореляційний зв'язок між досліджуваними ознаками статистично незначущий. Якщо $|\rho_в| > T_{кр}$ - нульову гіпотезу відкидають. Між досліджуваними ознаками існує значущий ранговий кореляційний зв'язок.

Приклад 19.5. Експертна комісія оцінила 10 виробів за двома ознаками *A* та *B*.

Знайдений коефіцієнт рангової кореляції Спірмена дорівнює $\rho_в = 0,5$.

При рівні значущості $\alpha = 0,1$ перевірити нульову гіпотезу $H_0 : \rho_Г = 0$ про рівність нулю генерального коефіцієнта рангової кореляції Спірмена. За альтернативну прийняти гіпотезу $H_1 : \rho_Г \neq 0$.

Розв'язок. Із таблиці критичних точок розподілу Стьюдента (додаток 6), за заданим рівнем значущості $\alpha = 0,1$ і числу ступенів вільності $k = 8$ знаходимо критичну точку $t_{кр}\left(\frac{\alpha}{2}; k\right) = t_{кр}(0,05; 8) = 1,86$.

Знайдемо значення критичної точки $T_{кр}$

$$T_{кр} = t_{кр}\left(\frac{\alpha}{2}; k\right) \cdot \sqrt{\frac{1-\rho_в^2}{n-2}} = 1,86 \cdot \sqrt{\frac{1-0,5^2}{10-2}} = 0,57.$$

Якщо $|\rho_в| < T_{кр}$ – нульову гіпотезу приймають. Ранговий кореляційний зв'язок між досліджуваними ознаками статистично незначущий.

Перевірка гіпотези про значущість вибіркового коефіцієнта рангової кореляції Кендалла

Для того, щоб при рівні значущості α перевірити нульову гіпотезу $H_0 : \tau_Г = 0$ про рівність нулю генерального коефіцієнта рангової кореляції

Кендалла при конкуруючій гіпотезі $H_1 : \tau_Г \neq 0$, необхідно обчислити критичну точку

$$T_{кр} = z_{кр} \cdot \sqrt{\frac{2 \cdot (2n+5)}{9n \cdot (n-1)}},$$

n - об'єм вибірки, $z_{кр}$ - критична точка, яку знаходять із таблиці значень функції Лапласа (додаток 2) із рівняння $\Phi(z_{кр}) = \frac{1-\alpha}{2}$ за рівнем значущості α.

Якщо $|\tau_в| < T_{кр}$ - нульову гіпотези приймають. Якщо $|\tau_в| > T_{кр}$ - нульову гіпотезу відкидають. Між досліджуваними ознаками існує значущий ранговий кореляційний зв'язок.

Приклад 19.6. Експертна комісія оцінила роботу 20 підприємств за двома параметрами *A* та *B*. Знайдений коефіцієнт рангової кореляції Кендалла дорівнює $\tau_в = 0,6$.

При рівні значущості перевірити нульову гіпотезу $H_0 : \tau_Г = 0$ про рівність нулю генерального коефіцієнта рангової кореляції Кендалла. За альтернативну прийняти гіпотезу $H_1 : \tau_Г \neq 0$.

Розв'язок. Знайдемо критичну точку $z_{кр}$

$$\Phi(z_{кр}) = \frac{1-\alpha}{2} = \frac{1-0,05}{2} = 0,475.$$

Із таблиці значень функції Лапласа (додаток 2) знаходимо $z_{кр} = 1,96$.

Знайдемо значення критичної точки

$$T_{кр} = z_{кр} \cdot \sqrt{\frac{2 \cdot (2n+5)}{9n \cdot (n-1)}} = 1,96 \cdot \sqrt{\frac{2 \cdot (2 \cdot 20+5)}{9 \cdot 20 \cdot (20-1)}} = 0,32.$$

Так як $|\tau_в| > T_{кр}$ - нульову гіпотезу відкидають. Між досліджуваними ознаками існує значущий ранговий кореляційний зв'язок.

19.3. Рангова кореляція

Ранжируваний ряд – це ряд значень досліджуваної кількісної чи якісної ознаки, які розташовані у порядку зростання або спадання величини ознаки.

Зв'язка – це група елементів ранжируваного ряду, які мають однакові значення. Об'єм зв'язки – кількість однакових значень у зв'язці.

Ранг – це порядковий номер значення досліджуваної ознаки у ранжируваному ряду. Елементам зв'язки присвоюють однакові ранги, які називають середніми рангами. Середній ранг – це середнє арифметичне порядкових номерів елементів зв'язки у ранжируваному ряду.

Приклад 19.9. На виставці представлено п'ять виробів. Результати оцінювання їх експертною комісією занесені у таблицю

Номер виробу	Кількість балів
1	6
2	7
3	10
4	8
5	7

Провести ранжирування даних за кількістю балів. Записати для кожного виробу його ранг.

Розв'язок. Розмістимо вироби у порядку зменшення отриманих балів

Номер виробу	Кількість балів
3	10
4	8
2	7
5	7
1	6

Вироби з номерами 2 та 5 набрали однакову кількість балів і утворюють у ранжируваному ряду зв'язку, тому для присвоєння їм рангів знаходимо середнє арифметичне їх номерів у ряду $R_{2,5} = \frac{3+4}{2} = 3{,}5$. Отже вироби з номерами 2 та 5 отримують однакові ранги $R_2 = R_5 = 3{,}5$.

Результуюча таблиця матиме вигляд

Номер виробу	Кількість балів	Ранг
3	10	1
4	8	2
2	7	3,5
5	7	3,5
1	6	5

Ранговий коефіцієнт кореляції Спірмена.

Коефіцієнт кореляції Спірмена використовують для перевірки наявності зв'язку між ознаками *A* та *B*, які представлені за ранжированою (порядковою) шкалою.

Нехай результати дослідження занесені у таблицю:

Об'єкт дослідження	Ранг об'єкта за ознакою A, R_{A_i}	Ранг об'єкта за ознакою B, R_{B_i}
1	R_{A_1}	R_{B_1}
2	R_{A_2}	R_{B_2}
…	…	…
n	R_{A_n}	R_{B_n}

Ранговий коефіцієнт кореляції Спірмена обчислюють за формулою

$$\rho = 1 - \frac{6\left[\left(R_{A_1} - R_{B_1}\right)^2 + \left(R_{A_2} - R_{B_2}\right)^2 + \cdots + \left(R_{A_n} - R_{B_n}\right)^2\right]}{n^3 - n} = 1 - \frac{6\sum_{i=1}^{n}\left(R_{A_i} - R_{B_i}\right)^2}{n^3 - n},$$

де R_{A_i} та R_{B_i} - ранги за ознаками A та B об'єкта з номером i, n - кількість досліджуваних об'єктів.

При наявності однакових рангів використовують формулу

$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left(R_{A_i} - R_{B_i}\right)^2 + T_A + T_B}{n^3 - n},$$

де T_A і T_B - правобічні коефіцієнти, які розраховують за формули

$$T_A = \frac{\sum_{i=1}^{L_A}\left(a_i^3 - a_i\right)}{12},\ T_B = \frac{\sum_{j=1}^{L_B}\left(b_j^3 - b_j\right)}{12},$$

де L_A - кількість зв'язок у ранговому ряду A, a_i - об’єм зв’язки з номером i у ранговому ряду A;

де L_B - кількість зв'язок у ранговому ряду B, b_j - об’єм зв’язки з номером j у ранговому ряду B.

Значення коефіцієнта кореляції Спірмена можуть змінюватися в межах від −1 до +1, при цьому −1 відповідає повній протилежності послідовностей рангів, *а*+1 − їх повному збігу.

Приклад 19.10. В олімпіаді, яка складається з теоретичного та практичного турів, прийняло участь 10 студентів. Результати олімпіади занесені у таблицю.

Встановити, чи існує зв’язок між теоретичними знаннями студентів та практичними навичками.

Прізвище та ініціали	Кількість балів за теоретичний тур	Кількість балів за практичний тур
Антонюк В.Р.	18	15
Борисенко К.Н	10	12
Варава Н.Г.	14	11
Грицай О.П.	15	17
Дордюк М.І.	9	12
Каленик О.О.	11	12
Лукаш Б.Б.	15	16
Мірченко Т.Р.	10	13
Тарасюк А.І.	20	19
Федчик Р.О.	17	18

Розв’язок. Позначимо: ознака *А* – кількість балів за теоретичний тур, ознака *В* – кількість балів за практичний тур.

Оскільки закони розподілу ознак *А* та *В* невідомі, то застосовувати для виявлення зв’язку коефіцієнт кореляції Пірсона некоректно. Тому використаємо ранговий коефіцієнт кореляції Спірмена.

Присвоїмо кожному студенту ранги за теоретичний та практичний тури відповідно до кількості набраних балів. Для цього проранжируємо значення ознаки A у порядку зменшення балів і запишемо порядкові номери та ранги для елементів отриманого ряду

Кількість балів	20	18	17	15	15	14	11	10	10	9
Порядковий номер	1	2	3	4	5	6	7	8	9	10
Ранг	1	2	3	4,5	4,5	6	7	8,5	8,5	10

Згідно цієї таблиці студенту, який набрав 20 балів присвоюється ранг 1, студенту, який набрав 18 балів – ранг 2 , студенту, який набрав 17 – ранг 3. Студенти, які набрали по 15 балів утворюють зв'язку і займають 4 та 5 місця у ранжируваному ряду, тому їм присвоюють середній ранг 4,5. Студенти, які набрали по 10 балів також утворюють зв'язку і займають 8 та 9 місця у ранжируваному ряду, тому їм присвоюють середній ранг 8,5.

Тепер проранжируємо значення ознаки B у порядку зменшення балів і запишемо порядкові номери та ранги для елементів цього ряду:

Кількість балів	19	18	17	16	15	13	12	12	12	11
Порядковий номер	1	2	3	4	5	6	7	8	9	10
Ранг	1	2	3	4	5	6	8	8	8	10

Добавимо у вихідну таблицю з результатами олімпіади два стовпця з рангами студентів, а також для спрощення обчислень стовпець із значеннями $R_{A_i} - R_{B_j}$ та стовпець із значеннями $\left(R_{A_i} - R_{B_j}\right)^2$, $i = j$. Отримаємо таблицю

Прізвище та ініціали	Кількість балів за теоретичний тур, ознака A	Ранг R_{A_i}	Кількість балів за практичний тур, ознака B	Ранг R_{B_i}	$R_{A_i} - R_{B_i}$	$\left(R_{A_i} - R_{B_i}\right)^2$
Антонюк В.Р.	18	2	15	5	-3	9
Борисенко К.Н	10	8,5	12	8	0,5	0,25
Варава Н.Г.	14	6	11	10	-4	16

Грицай О.П.	15	4,5	17	3	1,5	2,25
Дордюк М.І.	9	10	12	8	2	4
Каленик О.О.	11	7	12	8	-1	1
Лукаш Б.Б.	15	4,5	16	4	0,5	0,25
Мірченко Т.Р.	10	8,5	13	6	2,5	6,25
Тарасюк А.І.	20	1	19	1	0	0
Федчик Р.О.	17	3	18	2	1	1

Так як рангові ряди містять зв'язки, то необхідно врахувати поправочні коефіцієнти T_A і T_B.

У ранговому ряду A є дві зв'язки, кожна об'ємом по 2 елементи: два ранги мають значення 4,5 і два ранги мають значення 8,5. Звідси $L_A = 2$, $a_1 = a_2 = 2$, а коефіцієнт T_A дорівнює

$$T_A = \frac{\sum_{i=1}^{L_A}\left(a_i^3 - a_i\right)}{12} = \frac{\left(2^3 - 2\right) + \left(2^3 - 2\right)}{12} = 1,$$

У ранговому ряду B є одна зв'язка об'ємом 3 елементи. Отже $L_B = 1$, $b_1 = 3$. Коефіцієнт T_B дорівнює

$$T_B = \frac{\sum_{j=1}^{L_B}\left(b_j^3 - b_j\right)}{12} = \frac{3^3 - 3}{12} = 2.$$

Знаходимо ранговий коефіцієнт кореляції Спірмена

$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left(R_{A_i} - R_{B_i}\right)^2 + T_A + T_B}{n^3 - n} =$$

$$= 1 - 6 \cdot \frac{9 + 0,25 + 16 + 2,25 + 4 + 1 + 0,25 + 6,25 + 0 + 1 + 1 + 2}{10^3 - 10} = 0,74.$$

Згідно шкали Чеддока між теоретичними знаннями та практичними навиками студентів існує сильний зв'язок.

Коефіцієнт рангової кореляції Кендалла

Ранговий коефіцієнт кореляції Кендалла (як і ранговий коефіцієнт кореляції Спірмена) використовують для перевірки наявності зв'язку між ознаками *A* та *B*, які представлені за порядковою шкалою.

Нехай провели дослідження n об'єктів сукупності і присвоїли їм відповідні ранги за ознаками *A* та *B*. Упорядкуємо ці об'єкти за зростанням рангів ознаки *A* і занесемо отримані результати у таблицю

Ранг *A*	R_{A_1}	R_{A_2}	…	R_{A_n}
Ранг *B*	R_{B_1}	R_{B_2}	…	R_{B_n}

Ранги ознаки *B* при цьому, як правило, будуть неупорядковані (упорядкованими за зростанням вони будуть тільки у випадку, коли ранги обох ознак співпадають).

Обчислимо кількість перестановок елементів рангового ряду *B* для упорядкування їх за зростанням. Для цього знайдемо кількість рангів ознаки *B*, які менші від R_{B_1} і знаходяться правіше від R_{B_1}. Позначимо отримане число через K_1. Знайдемо кількість рангів ознаки *B*, які менші від R_{B_2} і знаходяться правіше від R_{B_2}. Позначимо отримане число через K_2. Виконаємо цю операцію для усіх рангів ознаки *B* крім останнього. У результаті виконаних дій отримаємо ряд чисел $K_1, K_2, ..., K_{n-1}$. Знайдемо суму цих чисел

$$K = K_1 + K_2 + ... + K_{n-1}.$$

Коефіцієнт рангової кореляції Кендалла обчислюють за формулою

$$\tau = 1 - \frac{4K}{n(n-1)}.$$

У випадку, коли рангові ряди містять зв'язки, використовують формулу

$$\tau = 1 - \frac{2K}{\sqrt{\left(\frac{n(n-1)}{2} - T_A\right) \cdot \left(\frac{n(n-1)}{2} - T_B\right)}}.$$

де T_A та T_B - поправочні коефіцієнти, які розраховуються за формулами:

$$T_A = \frac{\sum_{i=1}^{L_A}\left(a_i^2 - a_i\right)}{2},\ T_B = \frac{\sum_{j=1}^{L_B}\left(b_j^2 - b_j\right)}{2},$$

де L_A - кількість зв'язок у ранговому ряду A, a_i - об’єм зв’язки з номером i у ранговому ряду A;

де L_B - кількість зв'язок у ранговому ряду B, b_j - об’єм зв’язки з номером j у ранговому ряду B.

Як і ранговий коефіцієнт кореляції Спірмена, коефіцієнт кореляції Кендала може змінюватися в межах від −1 до +1, при цьому −1 відповідає повній протилежності послідовностей рангів, *a*+1 − їх повному збігу.

Рангові коефіцієнти кореляції Спірмена і Кендалла пов’язані рівнянням $\rho \approx 1{,}5 \cdot \tau$ (при умові, що ρ і τ не дуже близькі до 1).

Для вибірок із генеральних сукупностей із нормальними розподілами можуть бути отримані оцінки для звичайного коефіцієнта кореляції

$$r_{xy} = 2\sin\left(\rho\frac{\pi}{6}\right) \text{ та } r_{xy} = \sin\left(\tau\frac{\pi}{2}\right)$$

Приклад 19.11. Експертна комісія оцінила 10 виробів за двома ознаками *A* та *B*. Результати оцінювання наведено у таблиці

Номер виробу	1	2	3	4	5	6	7	8	9	10
Ранг R_{A_i}	2	8	6	4	10	7	5	9	1	3
Ранг R_{B_i}	5	8	10	3	7	9	4	6	1	2

Знайти рангові коефіцієнти кореляції Кендалла та Спірмена.

Розв’язок. Запишемо в окрему таблицю пари рангів, упорядкувавши їх за зростанням рангів ознаки *A*

Ранг R_{A_i}	1	2	3	4	5	6	7	8	9	10
Ранг R_{B_i}	1	5	2	3	4	10	9	8	6	7

Правіше $R_{B_1} = 1$ немає менших рангів, тому $K_1 = 0$. Правіше $R_{B_2} = 5$ є 3 менших ранги, тому $K_2 = 3$. Правіше $R_{B_3} = 2$ немає менших рангів, тому

$K_3 = 0$. Аналогічно знаходимо: $K_4 = 0; K_5 = 0; K_6 = 4; K_7 = 3; K_8 = 2; K_9 = 0$. Сума $K = 12$.

Знайдемо ранговий коефіцієнт кореляції Кендалла

$$\tau = 1 - \frac{4K}{n(n-1)} = 1 - \frac{4 \cdot 12}{10 \cdot 9} = 0,47.$$

Знайдемо ранговий коефіцієнт кореляції Спірмена. Для спрощення обчислень додаємо у таблицю з рангами два рядки із значеннями $R_{A_i} - R_{B_j}$ та $\left(R_{A_i} - R_{B_j}\right)^2, i = j$. Отримаємо таблицю

Ранг R_{A_i}	1	2	3	4	5	6	7	8	9	10
Ранг R_{B_i}	1	5	2	3	4	10	9	8	6	7
$R_{A_i} - R_{B_i}$	0	-3	1	1	1	-4	-2	0	3	3
$\left(R_{A_i} - R_{B_i}\right)^2$	0	9	1	1	1	16	4	0	9	9

Підставимо отримані значення у формулу для коефіцієнта рангової кореляції Спірмена

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n}\left(R_{A_i} - R_{B_i}\right)^2}{n^3 - n} = 1 - 6 \cdot \frac{0+9+1+1+1+16+4+0+9+9}{10^3 - 10} = 0,7.$$

Отримані результати узгоджуються із твердженням, що коефіцієнти рангової кореляції Спірмена і Кендалла пов’язані співвідношенням $\rho \approx 1,5 \cdot \tau$.