

Теоретичні та практичні проблеми застосування математичних методів в аналізі соціально-економічних даних, виміряних в неметричних шкалах

- 1. Теоретичні проблеми застосування математичних методів в аналізі соціально-економічних даних, виміряних в неметричних шкалах**
- 2. Практичні проблеми застосування математичних методів в аналізі соціально-економічних даних, виміряних в неметричних шкалах**

1. Теоретичні проблеми застосування математичних методів в аналізі соціально-економічних даних, виміряних в неметричних шкалах

Методи рангової кореляції є єдиним шляхом узагальнення експертних оцінок, коли ознаки об'єкта в економіці не вдається описати метричною величиною. Ранг вказує, яке місце займає об'єкт в ряді об'єктів, упорядкованих за певною ознакою.

Традиційно в якості основних характеристик парного статистичного зв'язку між упорядкуваннями використовуються коефіцієнти рангової кореляції Спірмена або Кендела. Як відомо, першим запропонував розв'язання задачі перевірки гіпотези про незалежність порядкових ознак психолог Ч. Спірмен у 1900 році. Другий за популярністю коефіцієнт рангової кореляції – коефіцієнт М. Кендела – в якості міри схожості між двома ранжуваннями використовує мінімальну кількість перестановок сусідніх об'єктів, яку потрібно зробити, щоб одне впорядкування об'єктів перетворити в інше.

Нехай p_k, q_k – ранги величин ознак X і Y . Спостереження завжди можна відсортувати в порядку зростання рангів однієї ознаки $p_k = 1, 2, 3, \dots, n$. Ранги q_k – це ті ж самі числа, але в іншому порядку.

Мірою тісноти зв'язку між ознаками X і Y може бути сума квадратів різниць рангів:

$$S = \sum_{k=1}^n (p_k - q_k)^2.$$

Якщо ранги двох ознак співпадають $p_k = q_k$, то $S = 0$, і це відповідає найтіснішому додатному зв'язку.

Коефіцієнт рангової кореляції Спірмена

$$\rho = 1 - \frac{6 \sum (p_k - q_k)^2}{(n-1)n(n+1)}.$$

Коефіцієнт рангової кореляції Спірмена співпадає зі звичайним коефіцієнтом парної кореляції Пірсона, обчисленим за рангами:

$$\rho = r_{pq} = \frac{S_{pq}}{S_p \cdot S_q}.$$

Отже, коефіцієнт рангової кореляції Спірмена дорівнює звичайному коефіцієнту парної кореляції Пірсона, якщо за змінні прийняти ранги p_k, q_k . Таким чином, для розв'язання практичних задач з порядковими ознаками можна використовувати програми, що написані для обчислення коефіцієнтів парної кореляції Пірсона.

Коефіцієнт рангової кореляції Кендела:

$$S = S_{\max} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn}(p_i - p_j) = \frac{n(n-1)}{2};$$

$$\tau = \frac{S}{S_{\max}} = \frac{2S}{n(n-1)}.$$

Значення коефіцієнта Кендела змінюється в межах $-1 \leq \tau \leq 1$.

Міру узгодженості оцінок двох експертів можна оцінити за допомогою коефіцієнта рангової кореляції Спірмена ρ_s . Таких порівнянь буде $\frac{m(m-1)}{2}$.

Але потрібно знайти загальну міру узгодженості всієї групи експертів. Така міра (коефіцієнт конкордації) була запропонована Кенделом у вигляді

$$W = \frac{12 \cdot S}{m^2(n^3 - n)},$$

$$\text{де } S = \sum \left(U_i - m \cdot \frac{n+1}{2} \right)^2.$$

Коефіцієнт конкордації співпадає з індексом детермінації, обчисленим за таблицею рангів y_{ij} .

Аналіз системи елементарних порядкових ознак СЕС слід продовжити дисперсійним аналізом.

Зазвичай «дисперсійний аналіз передбачає, що експеримент організований відповідно до плану, який дозволяє за порівняно малої кількості вимірювань незалежно оцінити вплив кожного з факторів на вимірюваний кількісний показник. Застосування дисперсійного аналізу у ході пасивних спостережень, наприклад, в економіці, передбачає належний відбір груп даних з більш значної сукупності даних». У класичному дисперсійному аналізі вважається, що вихідні дані вимірюються за метричними шкалами та є кількісними. Це – найпоширеніший і добре відпрацьований апарат дисперсійного аналізу.

Проте, незважаючи на широке практичне застосування в економіці, існує істотний недолік класичного дисперсійного аналізу – це обтяжливе припущення про нормальний розподіл значень кількісного показника в кожній групі (вибірці) спостережень. Рангові методи значною мірою позбавлені цих недоліків, вони добре пристосовані для роботи з малими вибірками, розподілів яких ми не знаємо. Зрозуміло, що перехід від самих спостережень (x) до їх рангів (u) супроводжується певною втратою інформації, але відразу слід зазначити, що ця втрата невелика (цей оптимістичний висновок впливає з порівняння рангових правил з точними, коли відомий закон розподілу

похибок). Після ранжування (переходу до рангів) статистичні висновки стають більш загальними та надійними.

Розглянемо однофакторну схему дисперсійного аналізу Краскала–Уолліса, яка є узагальненням звичайного дисперсійного аналізу на рангових змінних. У дисперсійному аналізі порівнюються k вибірок спостережень та перевіряється «нуль-гіпотеза» про те, що всі дані були відібрані з однієї сукупності. Вибірки сформовані так, щоб вони відповідали різним станам певної якісної ознаки. Якщо «нуль-гіпотеза» відхиляється (між вибірками є істотні відмінності), то далі слід встановити, між якими вибірками є найістотніші відмінності, тобто потрібно виявити причину значущих відмінностей.

Відомо, що сума квадратів стандартизованих нормально розподілених величин w_j має розподіл χ^2 :

$$\chi^2 = \sum_{j=1}^k \left[\frac{w_j - M(w_j)}{\sqrt{D(w_j)}} \right]^2.$$

Проте середні ранги за групами $\bar{z}_j$ мають дещо інший розподіл. Краскал і Уолліс довели, що аналогічна статистика H_i (з поправковими коефіцієнтами $1 - \frac{n_j}{N}$)

$$H = \sum_{j=1}^k \left(1 - \frac{n_j}{N} \right) \left[\frac{\bar{z}_j - M(\bar{z}_j)}{\sqrt{D(\bar{z}_j)}} \right]^2$$

має розподіл $\chi^2(k-1)$.

Число ступенів свободи (ЧСС) дорівнює $k - 1$, оскільки склад однієї з вибірок завжди відомий, у сумі ранги елементів усіх вибірок складають множину номерів 1, 2, ..., N .

Якщо виявиться, що обчислене значення H буде більшими від табличного $H > \chi_{0,01}^2(k - 1)$, то «нуль-гіпотеза» відхиляється і потрібно виконати попарне порівняння всіх вибірок для визначення причини значущості. Усього потрібно виконати $\frac{k(k - 1)}{2}$ попарних порівнянь середніх рангів за групами.

Стандартизовану величину різниці середніх рангів двох груп, обчислену за формулою:

$$t = \frac{|\bar{z}_j - \bar{z}_i|}{\sqrt{\frac{N(N + 1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}},$$

потрібно порівняти з квантилями нормального розподілу $t_{0,05} = 1,96 \approx 2$ та $t_{0,01} = 2,58$. Різниці вважають значущими, якщо буде $t > t_{0,05}(\text{ЧСС } n_i + n_j - 2)$.

2. Практичні проблеми застосування математичних методів в аналізі соціально-економічних даних, виміряних в неметричних шкалах

Діяльність підприємства визначалась неметричними (порядковими) ознаками поданими у формі балів анкетного опитування експертів. Експертами були керівники відділу маркетингу чи економічного відділу підприємств, яким постачається продукція підприємства, всього 50 осіб. У табл. 1 містяться 14 питань, що ставилися експертам для визначення якісних ознак діяльності підприємства. Оцінки виставлялися експертами за п'ятибальною шкалою, тобто значення відповідей (порядкові ознаки) виражені ординальними числами. Практика свідчить, що в разі наявності неповних оцінок відповідей такі дані необхідно вилучити з матриці. У даному випадку кінцева матриця оцінок має

дещо менші розміри 24×14 (тобто є оцінки 14 якісних ознак діяльності виробника з точки зору 24 споживачів виробленої продукції).

Таблиця 1

Перелік питань анкетування оцінки якості виробничо-господарської діяльності підприємства

Питання	Позначення	Експерти				
		Y_1	Y_2	Y_3	...	Y_{50}
Оцінювання діяльності підприємства						
1. Якість продукції	x_1					
2. Асортимент продукції	x_2					
3. Освоєння нових типів продукції	x_3					
4. Своєчасність постачання	x_4					
5. Ціна на продукцію	x_5					
Оцінювання рівня обслуговування						
6. Виконання спеціальних замовлень	x_6					
7. Технічний супровід	x_7					
8. Надання інформації про можливості та хід виконання замовлення	x_8					
9. Оперативність вирішення питань	x_9					
10. Взаємодія протягом вирішення проблем	x_{10}					
Рекламна підтримка						
11. Забезпечення рекламно-інформаційними матеріалами	x_{11}					
12. Проведення семінарів, конференцій, зустрічей	x_{12}					
13. Оцінка web-сайта	x_{13}					
14. Оцінка авторитету підприємства	x_{14}					

Для обчислення скоректованих коефіцієнтів Кендела, Спірмена і конкордації спочатку були обчислені ранги значень оцінок, оскільки саме ранги значень оцінок дозволяють далі виконувати їх обчислення як з метричними величинами кількісних ознак. У табл. 2 і табл. 3 наведені матриці обчислених значень скоректованих коефіцієнтів рангової кореляції Спірмена (ρ_s), Кендела

(τ_b) і конкордації (W_*), математичні формули яких наведені вище. Обчислення виконані для двох зрізів анкетного опитування експертів. Згідно з оцінками експертів спостерігалася тісна статистична залежність оперативності вирішення питань із взаємодією в процесі їх вирішення (коефіцієнт Спірмена $\rho_{x_9x_{10}} = 0,917$), а також залежність виконання спеціальних замовлень від технічного супроводу продукції ($\rho_{x_6x_7} = 0,748$) (табл. 2). Взаємодія працівників протягом вирішення проблем надзвичайно залежить від надання інформації про можливості та хід виконання замовлення (коефіцієнт Кендела $\tau_{x_8x_{10}} = 0,917$ табл. 3), а ціна на продукцію взаємопов'язана з технічним супроводом ($\tau_{x_6x_7} = 0,715$). Слід відзначити низьку узгодженість питань опитування в анкеті, про що свідчить невелике значення коефіцієнта конкордації $W = 0,310068$. Якщо дане значення розглядати як значення коефіцієнта детермінації, як це було вже доведено, то сукупність питань, наведених в анкеті, мало відображають якісну характеристику підприємства, яка оцінюється за допомогою даної анкети.

Для аналізу відповідей експертів слід транспонувати матрицю балів та обчислити ті ж самі коефіцієнти, що й для порядкових ознак (додаток 3). Згідно зі значеннями коефіцієнта Спірмена табл. 4 відзначено тісний кореляційний зв'язок оцінок експертів y_{37} з y_{10} ($\rho_{y_{37}y_{10}} = 0,8987$), y_{49} з y_{24} ($\rho_{y_{49}y_{24}} = 0,8594$); y_{49} з y_{26} ($\rho_{y_{49}y_{26}} = 0,8411$); y_{49} з y_1 ($\rho_{y_{49}y_1} = 0,8339$). Матриця коефіцієнтів Кендела τ_b наведена в табл. 3.5 додатку 3. Значуща узгодженість величин оцінок експертів за структурною логікою значна для y_1 з y_{24} ($\tau_{y_1y_{24}} = 0,88$), а також y_{10} з y_{37} ($\tau_{y_{10}y_{37}} = 0,876$).

За величиною коефіцієнта W можна зробити висновок про узгодженість думок експертів в оцінці задоволеності діяльністю підприємства як виробника. Значення $W = 0,412$ свідчить про достатню узгодженість думок експертів в оцінці даної діяльності підприємства.

Дослідження якості діяльності підприємства з точки зору споживачів були продовжені і наступного року. Характерним залишився взаємозв'язок оперативності вирішення питань із взаємодією в процесі їх вирішення (коефіцієнт Спірмена $\rho_{x_9x_{10}} = 0,8445$), яка, у свою чергу, залежить від проведення семінарів, конференцій, зустрічей, що надзвичайно природно ($\rho_{x_{10}x_{12}} = 0,7061$). Максимальне значення коефіцієнта Кендела для ознак по підприємству дорівнювала $\tau_{x_8x_{10}} = 0,8183$, що характеризує таку ж саму структуру послідовності величин ознак, що і для підприємства, $W = 0,300873$.

Співпадали думки щодо діяльності підприємства експертів y_{37} та y_{45} ($\rho_{y_{37}y_{45}} = 0,899$), а також y_{44} та y_{46} ($\rho_{y_{44}y_{46}} = 0,835$). Для названих експертів y_{37} та y_{45} співпадає і структурна послідовність значень балів, про свідчить значення коефіцієнта Кендела ($\tau_{y_{37}y_{45}} = 0,876$), для y_{44} та y_{46} ($\tau_{y_{44}y_{46}} = 0,791$). У цілому ж думка експертів в якісній характеристиці виробника їх продукції була погано узгодженою $W = 0,161138$, тобто наявні найрізноманітніші думки з даного приводу.

Найчастіше перевірка узгодженості оцінок експертів здійснюється за допомогою обчислення коефіцієнта конкордації W . Якщо виявиться, що сукупність оцінок різних експертів не узгоджена, до слід перевірити $\frac{m(m-1)}{2}$ пар експертів і знайти групи експертів, що мають різні точки зору. Для цього слід застосувати процедуру – тест Краскала–Уолліса. Як приклад, в табл. 3 наведені обчислення для x_1 (перший рік дослідження діяльності підприємства з позиції споживача).

Тест Краскала–Уоліса

Групи	Кількість об'єктів у групі	Середній ранг
1	12	15,6667
2	12	9,33333

Обчислена значення $H = 6,13333$, для якого p -значення дорівнює $0,0132634 < 0,05$. Таким чином, «нуль-гіпотеза» відхиляється (між вибірками є істотні відмінності) і як уже зазначалося, потрібно виявити причину значущих відмінностей.

Описані вище процедури аналізу були виконані для всіх порядкових ознак діяльності підприємства. За результатами дисперсійного аналізу порядкових ознак маємо однорідність суджень експертів за якістю продукції (x_1), асортиментом продукції (x_2), освоєння нових типів продукції (x_3), ціни на продукції (x_5), виконання спеціальних замовлень (x_6). На дві групи розділилися експерти за своїми позиціями щодо своєчасності постачання (x_4), технічного супроводження (x_7), надання інформації про можливості та хід виконання замовлення (x_8), оперативності вирішення питань (x_9), взаємодії протягом вирішення проблем (x_{10}), забезпечення рекламно-інформаційними матеріалами (x_{11}), проведення семінарів, конференцій, зустрічей (x_{12}), оцінки web-сайта (x_{13}), оцінки авторитету підприємства (x_{14}). Чотири ознаки, за якими сукупність експертів однорідна, є визначальними та задовільними в оцінці діяльності підприємства; за рештою ознак думки експертів не співпадають. Це потрібно врахувати в процесі формування маркетингової стратегії підприємства.

Аналіз елементарних порядкових ознак СЕС, їх взаємозв'язку є першим етапом в моделюванні складних порядкових ознак систем.