Одеський національний університет імені І. І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра оптимального керування і економічної кібернетики

Кваліфікаційна робота

на здобуття ступеня вищої освіти «магістр»

«Застосування методів статистичного аналізу в економіці»

«Applying statistical analysis methods in economics»

Виконав: здобувач денної форми навчання

спеціальності 113 Прикладна математика
Освітня програма «Прикладна математика»

Шеванов Владислав Олександрович

Керівник
канд. фіз.-мат. наук, доцент Васильєв О.Б.
(науковий ступінь, вчене звання, прізвище та ініціали, підпис)
Рецензент
канд. фіз.-мат. наук, доцент Вербіцький В.В.
(науковий ступінь, вчене звання, прізвище та ініціали)

Рекомендовано до захисту:
Протокол засідання кафедри
№ ___ від _______________ р.

Завідувач кафедри

_________ _____________________
(підпис) (прізвище, ім'я)

Захищено на засіданні ЕК №_______
протокол № ____ від _____________ р.
Оцінка______________/___ ___/_____
(за національною шкалою, шкалою ECTS, бали)
Голова ЕК

____________ _______________________
(підпис) (прізвище, ім'я)

Одеса – 2023

ЗМІСТ

ВСТУП

Методи багатовимірного статистичного аналізу (МСА), сьогодні називають інтелектуальним інструментарієм дослідника. Вони становлять невіддільну частину фундаментальних курсів університетської освіти й активно використовуються в аналітичній практиці в країнах із передовою економікою.

Постійно зростаючий інтерес до МСА пояснюється насамперед його широкими можливостями у відображенні та моделюванні реальних явищ і процесів, які від самого початку мають, як відомо, багатоозначну природу. Крім того, без базових знань з опрацювання багатовимірних даних просто не можуть розвиватися сучасні математика і статистика. Усі новітні розробки, присвячені проблемам застосування нечітких множин, моделювання катастроф, розпізнавання образів, сценарного прогнозування тощо, припускають багатовимірне представлення спостережуваних об'єктів.

МСА в теоретичному плані являє собою подальший розвиток традиційної одновимірної статистки, його відрізняють трудомісткі алгоритми реалізації обчислювальних процедур, практично завжди розраховані на залучення технічних засобів, і складна інтерпретованість аналітичних результатів. Це вимагає від користувача досить глибокої підготовки як у галузі математичної статистики, так і в галузі, в якій проводяться конкретні дослідження економіки, медицини тощо.

РОЗДІЛ 1 МЕТОДИ ФАКТОРНОГО АНАЛІЗУ

1.1 Загальний алгоритм методів факторного аналізу

Факторний аналіз – це процес групування записаних вами змінних у категорії, які називаються факторами. Це може допомогти вам зменшити загальну кількість змінних, залучених до статистичного дослідження, що може спростити оцінку даних. Під час проведення факторного аналізу ви можете визначити подібності та відмінності між набором змінних, щоб встановити кореляції між ними. Наприклад, якщо ви спостерігаєте набір змінних, пов'язаних із мотивацією співробітників, ви можете використати факторний аналіз, щоб визначити, які мотивації корелюють, а потім згрупувати подібні у фактори.

Ви можете використовувати факторний аналіз, коли проводите дослідження з використанням великої кількості змінних. Цей тип аналізу корисний для того, щоб зробити дані більш керованими та для визначення більш точних висновків зі статистичних даних. Вивчення факторного аналізу може допомогти вам використовувати його у ваших власних статистичних дослідженнях [9].

Існують різні типи методів, які використовуються для вилучення фактора з набору даних:

1. Аналіз головних компонентів: це найпоширеніший метод, який використовують дослідники. PCA починає витягувати максимальну дисперсію та поміщає їх у перший фактор. Після цього він видаляє цю дисперсію, пояснену першими факторами, а потім починає витягувати максимальну дисперсію для другого фактора. Цей процес йде до останнього фактора.

2. Аналіз загальних факторів: другий за популярністю метод дослідників, він виділяє загальну дисперсію та об'єднує їх у фактори. Цей метод не включає унікальну дисперсію всіх змінних. Цей метод використовується в SEM.

3. Факторинг зображення: Цей метод заснований на кореляційній матриці. Метод OLS Regression використовується для прогнозування фактора у факторингу зображення.

4. Метод максимальної правдоподібності: цей метод також працює з кореляційною метрикою, але він використовує метод максимальної правдоподібності для факторів.

5. Ваговий квадрат — ще один метод, заснований на регресії, який використовується для факторизації [6].

Методи факторного аналізу, при їх різноманітті, мають загальний алгоритм розв'язання, представлений на рис. 1.1. Починаючись побудовою матриці вихідних даних X, цей алгоритм завершується отриманням матриць факторного відображення і значень чинників A і F. З урахуванням прийнятих позначень, де n - кількість спостережень, m - кількість аналітичних ознак X, r - кількість значущих узагальнених ознак (латентних факторів), на схемі показано розмірність матриць даних для кожного алгоритмічного кроку.

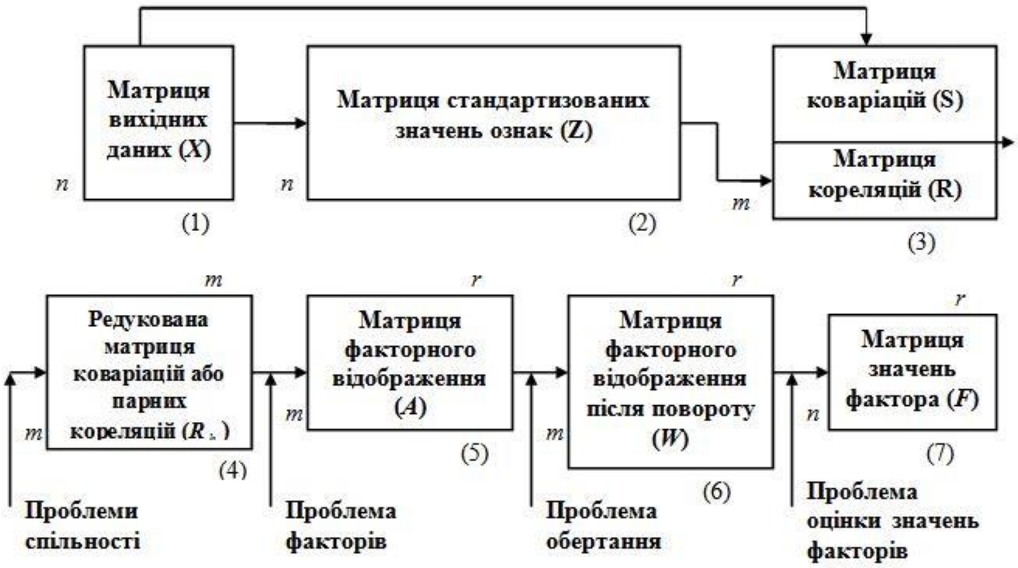

Рис.1.1. Алгоритмічна схема реалізації методів факторного аналізу

Перші кроки алгоритму 1-3 не викликають жодних труднощів. Перехід від матриці вихідних даних X до матриці стандартизованих даних Z здійснюється після перерахунку усіх елементів x_{ij} за формулою:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}. \tag{1.1}$$

Як відомо, для стандартизованих значень z_{ij} математичне очікування дорівнює нулю ($E_{(z)} = 0$) і дисперсія $D_{(z)} = 1$.

На наступному кроці простим перемноженням скаляра — $\frac{1}{n}$ і матриць Z' і Z отримуємо матрицю парних кореляцій:

$$R = \frac{1}{n} Z'Z. \tag{1.2}$$

Крок 2 може бути пропущений і тоді наступне факторне рішення знаходять не за матрицею R, а за матрицею коваріацій: в останньому випадку бажано, щоб аналізовані ознаки X мали ті самі одиниці виміру.

Виконання четвертого кроку алгоритму зумовлюється розв'язанням першої проблеми - побудови редукованої матриці кореляцій (коваріацій). Проблема актуальна саме для методів факторного аналізу, оскільки в методі головних компонент приймають, що всю варіацію вихідних ознак повністю пояснюють латентні чинники, при цьому матриця парних кореляцій R розмірності $m \times m$ залишається і як редукована R_h, у якій усі спільності $\sum h_j^2 = 1$:

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1m} \\ r_{21} & 1 & r_{23} & \cdots & r_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & r_{m3} & \cdots & 1 \end{pmatrix}$$

У факторному аналізі матриця кореляцій R перетворюється на R_h:

$$R_h = \begin{pmatrix} h_1^2 & r_{12} & r_{12} & \cdots & r_{1m} \\ r_{21} & h_2^2 & r_{23} & \cdots & r_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & r_{m3} & \cdots & h_m^2 \end{pmatrix}$$

з $h_j^2 < 1$, тобто варіація ознак $(X_j, j = \overline{1,m})$ може бути пояснена не на 100%, а дещо менше, з урахуванням існування їхньої не розкритої характерності.

Існують досить прості методи пошуку спільностей h_j^2 :

- **метод найбільшої кореляції**. На головній діагоналі з позитивним знаком записується найбільший за величиною коефіцієнт кореляції.
- **метод Барта**. За кожним стовпчиком матриці R спочатку знаходять середнє значення коефіцієнтів кореляції $\bar{r}_j$, потім, якщо $\bar{r}_j$ порівняно велике, за спільність приймається значення, що дещо вище за найбільше в стовпчику коефіцієнта кореляції, і, якщо $\bar{r}_j$ - порівняно мале значення, спільність буде дещо менша за найбільше в стовпчику коефіцієнта кореляції;
- **метод тріад**. Спільності для кожного $j -$ го стовпця R обчислюють за формулою:
-

$$n_j^2 = \frac{r_{1k} r_{1l}}{r_{kl}}, \tag{1.3}$$

де r_{1k} і r_{1l}, - коефіцієнти кореляції, найбільші в стовпці;

- **метод малого центроїда**. Для кожної змінної j будують кореляційну матрицю розмірності 4×4. Включаючи саму змінну в цю матрицю, записують оцінки кореляції трьох інших змінних, особливо тісно пов'язаних із першою. За даними малої матриці кореляцій і розраховують спільності:

$$h_j^2 = \frac{(\sum_l r_{l1})}{\sum_{l,j} r_{lj}}, \tag{1.4}$$

де $\sum_l r_{l1}$- сума елементів першого стовпчика; $\sum_{l,j} r_{lj}$ - сума всіх елементів матриці 4×4.

Друга проблема виникає на етапі побудови матриці відображення A і полягає у виборі оптимального методу для пошуку вагових коефіцієнтів a_{1j} елементів матриці A. Нагадаємо, що найкращі рішення зазвичай знаходять за допомогою сучасних методів факторного аналізу: головних факторів, максимальної правдоподібності тощо. У загальному випадку виділені фактори не обов'язково ортогональні ($R = ACA'$) і тоді вектори (стовпці) матриці A будуть лінійно-залежними.

Виконання кроку 6 алгоритму і розв'язання проблеми обертання простору спільних чинників не обов'язкове. Потреба в цьому виникає, коли просторове розташування спільних чинників F_r, нелогічне або важко піддається інтерпретації.

Факторний аналіз має кілька припущень. До них належать:

1. У даних немає викидів.
2. Розмір вибірки має бути більшим за фактор.
3. Це метод взаємозалежності, тому між змінними не повинно бути ідеальної мультиколінеарності.
4. Факторний аналіз є лінійною функцією, тому він не вимагає гомоскедастичності між змінними.
5. Він також базується на припущенні лінійності. Отже, ми також можемо використовувати нелінійні змінні. Однак після перенесення вони перетворюються на лінійну змінну.
6. Крім того, він передбачає інтервальні дані.

Факторний аналіз включає наступну ключову концепцію:

дослідницький факторний аналіз – передбачає, що будь-яка змінна або показник може бути пов’язаний з будь-яким фактором.

Крім того, це найпоширеніший метод, який використовують дослідники; він не базується на жодній попередній теорії.

Підтверджувальний факторний аналіз – використовується для визначення факторного навантаження та факторів вимірюваних змінних, а також для підтвердження того, що він очікує на основі попередньо встановлених припущень. Крім того, він використовує два підходи:

- Традиційний метод;
- Підхід SEM [6].

1.2 Метод головних компонент

Із числа методів, які дозволяють узагальнювати значення елементарних ознак, метод головних компонент вирізняється простою логічною конструкцією і в той же час на його прикладі стають зрозумілими загальна ідея і цільові установки багаточисленних методів факторного аналізу.

Метод головних компонент дає можливість по m - числу вихідних ознак виділити r головних компонент, або узагальнених ознак. Простір головних компонент ортогональний.

Математична модель головних компонент базується на логічному допущенні, що значення множини взаємопов'язаних ознак породжують деякий загальний результат.

Припустивши лінійну форму зв'язку ознак X_j запишемо в матричній формі рівняння залежності результату F від X: $F = XB$, де B - вектор параметричних значень лінійного рівняння зв'язку. Умовою виконання такої рівності є відповідність дисперсій, тобто $D(X) = D(XB)$. Так як X - багатомірна випадкова величина, її дисперсійна оцінка - це коваріаційна матриця S. Постійна величина B виноситься за знак дисперсії і підноситься в квадрат, отримуємо: $D(F) = B'SB$.

Пошук головних компонент зводиться до задачі послідовного виділення головної компоненти F_1, яка має найбільшу дисперсію, другої головної компоненти F_2 , яка має другу по величині дисперсію і так далі. Подібна задача має рішення при умові введення обмежень. Нехай $B'B = b_1^2 + b_2^2 + \cdots + b_m^2 = 1$.

При $B'B = 1$ максималізуємо $B'SB$, використовуючи метод множників Лагранжа:

$$r = B'SB - \lambda(B'B - 1) \tag{1.5}$$

і

$$\frac{\partial r}{\partial B} = 2SB - 2\lambda B = 0, \tag{1.6}$$

звідки $SB - \lambda B = 0$.

Отже, отримаємо $|S - \lambda E|B = 0$ і характеристичне рівняння для пошуку λ_j буде: $|S - \lambda E| = 0$.

Із множини значень характеристичних чисел λ_j відносно першого, найбільшого λ_1 знаходимо вектор B_1 значень для першої головної компоненти F_1, для другого по величині характеристичного числа λ_2 - вектор значень другої компоненти B_2 и так далі до L_m і B_m для F_m при m - вихідному числі ознак, що аналізуються. Тут B - вектори величин, які представляють координати головних компонент F_r в просторі ознак R^X, вони ж - характеристики сили зв'язку r-ї головної компоненти і j-ї ознаки X_j.

Якщо вихідну матрицю даних X попередньо стандартизувати, то матриця коваріацій S перейде в матрицю парних кореляцій R, і вектор B буде власним вектором по стандартизованим даним U. Вирішальне рівняння в матричній формі приймає вид: $(R - \lambda E)U = 0$.

Результати застосування метода головних компонент видаються даними матриці відображень A. Можливий підсумковий запис залежних значень вихідних ознак від значень головних компонент:

$Z = AF$ або

$$z_{ij} = a_{j1}f_{1i} + a_{j2}f_{2i} + \cdots + a_{jr}f_{ri} \tag{1.7}$$

або залежність значень головних компонент від значень елементарних ознак

$F = A^{-1}Z'$, або

$$f_n = \frac{1}{\lambda_r}(a_{1r}z_{i1} + a_{2r}z_{i2} + \cdots + a_{mr}z_{ir}). \tag{1.8}$$

В рівняннях 1.7 і 1.8 прийняті означення:

z_{ij} - значення j стандартизованої змінної по i-му об'єкту спостереження;

f_{ri} - r-та головна компонента F_r по i-му об'єкту спостереження;

a_{jr} - ваговий коефіцієнт r-ї головної компоненти для j-ї змінної, оцінка приватного коефіцієнта кореляції для F_r і X_j (елементи j-го рядка матриці A);

a_{mr} - вагові коефіцієнти (характеристики сили зв'язку) m елементарних ознак ($j = \overline{1, m}$) для r-ї головної компоненти.

Рівняння 1.8 відносно F (головних компонент) є похідним від 1.7. Покажемо це.

Відомо, що алгебраїчно $A = V\Lambda^{1/2}$: помножимо обидві частини матричного рівняння зліва на $(V\Lambda^{1/2})$ і легко впевнимось, що $A'A = \Lambda$. Далі, маємо $Z = AF$, помножимо обидві частини цього рівняння на A', потім на $(A'A)^{-1}$ і отримаємо $F = (A'A)^{-1}A'Z$, або $F = \Lambda^{-1}A'Z$, тобто:

$$F_r = \frac{1}{\lambda_r}a_{1r}z_{i1} + a_{2r}z_{i2} + \cdots + a_{mr}z_{ir}, \tag{1.9}$$

де $a_{1r}, a_{2r}, \ldots, a_{mr}$ - елементи m-го стовпця для r-ї головної компоненти матриці відображення A.

Обчислювальні процедури метода головних компонент.

Рішення задачі методом головних компонент зводиться до поетапного перетворення матриці вихідних даних X:

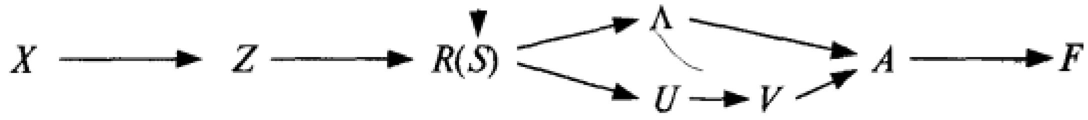

Рис.1.2 Етапи перетворення матриці

де X - матриця вихідних даних, розмірністю $n \times m$;

n - число об'єктів спостереження;

m - число елементарних аналітичних ознак;

Z - матриця стандартизованих значень ознак, елементи матриці обчислюються по формулі:

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j}; \tag{1.10}$$

R - матриця парних кореляцій:

$$R = \frac{1}{n} Z'Z. \tag{1.11}$$

Якщо попередня стандартизація даних не проводилась, то на даному кроці отримують матрицю

$$R = \frac{1}{n} X'X, \tag{1.12}$$

елементи матриці X для розрахунку S будуть центрованими величинами: $x_{ij} - \overline{x_j}$;

Λ - діагональна матриця власних (характеристичних) чисел:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_m \end{pmatrix}$$

Множину значень λ_j знаходять рішенням характеристичного рівняння - це характеристики варіацій, точніше, показники дисперсії кожної головної компоненти.

Сумарне значення дорівнює сумі дисперсій елементарних ознак. При умові стандартизації вихідних даних воно дорівнює числу елементарних ознак m.

Рішення характеристичного рівняння, коли число ознак m достатньо велике і матриця R великої розмірності, викликає труднощі при розрахунку детермінанту $|R|$. Вони успішно долаються з застосуванням різноманітних математичних методів матричної алгебри. Якщо A - деяка симетрична матриця розмірністю $m \times m$, то її детермінант знаходиться по сліду матриць, похідних з A.

Власні значення знаходять після того, як характеристичний багаточлен прирівнюють нулю, отримують характеристичне рівняння і вирішують його відносно характеристичних коренів.

V - матриця нормованих власних (характеристичних) векторів. Число векторів V_j спочатку дорівнює m. Отримують V_j перетворенням ненормованих власних векторів U.

Необхідність повторного, після отримання матриці Z, нормування простору тепер вже узагальнених ознак R пояснюється механічною помилкою в ході розрахунків, яка спотворює нормування простору.

В свою чергу власні вектори U_j знаходять із матричного рівняння. Реально це означає рішення m систем лінійних рівнянь. Кожна система об'єднує однорідні лінійні рівняння, і так як число її рівнянь дорівнює числу невідомих U, має нескінченну множину рішень. Конкретні значення власних векторів при цьому можна знайти, задаючи довільно принаймні величину однієї компоненти кожного вектора і звичайно, для того, щоб не ускладнювати розрахунки, її прирівнюють до одиниці.

A - матриця факторного відображення, її елементи - вагові коефіцієнти. Спочатку A має розмірність $m \times m$- по числу елементарних ознак X_j, потім в аналізі залишається r найбільш значимих компонент, $r \leq m.$ Обчислюють матрицю A по відомим даним матриці власних чисел Λ і нормованих власних векторів V.

F - матриця значень головних компонент розмірністю $r \times m.$

1.3 Метод головних факторів

Метод головних факторів можна розглядати як розвиток методу головних компонент. Основна відмінність полягає у використанні редукованої кореляційної матриці R_h, на головній діагоналі якої розташовані вже не одиниці, а характеристики спільності h_j^2.

Згідно з класичною моделлю факторного аналізу, рівняння для визначення коефіцієнтів при загальних факторах F_r записується у вигляді:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \cdots + a_{jm}F_n + a_jD_i \tag{1.13}$$

або в матричній формі:

$$Z_j = a_{j1} + a_jD_j \tag{1.14}$$

де D_j - характерний фактор.

Розв'язання рівняння за умови максимізації сум:

$\sum_{j=1}^{m} a_{j1}^2 = \lambda_1$— перший максимум в частині описаної дисперсії елементарних ознак $(D(Z_j))$;

$\sum_{j=1}^{m} a_{j1}^2 = \lambda_2$— другий максимум щодо решти, яка залишилась після λ_1 дисперсії і т.д., зводиться до визначення власних значень і власних векторів симетричної матриці R з рівності $(R - \lambda E)U = 0$.

У дослідницькій практиці існують різні прийоми, способи знаходження параметрів моделі головних чинників λ, U_j, а умовно їх легко розділити на дві великі групи.

Перша група орієнтована на алгоритм методу головних компонент і, по суті, повторює його. єдина відмінність у тому, що обчислення проводять за даними редукованої кореляційної матриці, а не звичайної матриці парних кореляцій. У цьому разі одразу отримують усі m значень власних чисел λ_j і m власних векторів. У факторному аналізі, однак, такий підхід використовують рідко, його вважають менш адекватним традиційним цілям факторного аналізу і менш економічним.

У другій групі об'єднуються прийоми, які дають змогу послідовно, починаючи з першого, встановлювати значення власних чисел і власних векторів. Наступні кроки виконуються після попередньої перевірки на достатність інформативності вже виділених головних чинників. Такий підхід можна вважати класичним у факторному аналізі. Найбільш представницьким тут є метод, розроблений Г. Готслінгом. Він дає змогу порівняно швидко виокремити невелику кількість загальних чинників, що враховують майже всю сумарну спільність. Метод передбачає ітеративне рішення. На першому етапі редуковану матрицю парних кореляцій багаторазово зводять до квадрата, чим домагаються збіжності до першого власного значення λ_1, потім обчислюють відповідні значення власного вектора U_{1}, і факторні навантаження a_{j1}. На завершення етапу знаходять добутки векторів $A_1 \times A_1'' = R^+$, де R^+ - відтворена кореляційна матриця. Залишкова матриця парних кореляцій буде $R - R^+$ Якщо при перевірці різниця $R - R^+$ істотна, переходять до другого етапу, і описана вище ітерація повторюється, але відносно другого власного числа - λ_2, обчислюваного за даними матриці залишкових коефіцієнтів кореляції, і т.д.

Ітерації повторюють доти, доки різниця $R - R^+$ не стає достатньо малою і тоді алгоритм завершується.

1.4 Перевірка надійності рішень, отриманих методами факторного аналізу

Коректне розв'язання завдань за допомогою методів факторного аналізу передбачає підтвердження значущості вихідної матриці парних кореляцій (коваріацій) і достатності числа узагальнених факторних ознак в аналізі.

Значимість кореляційної матриці піддається перевірці, якщо взяти до уваги, що незначущі кореляційні (коваріаційні) зв'язки елементарних ознак не дають підстав взагалі для пошуку узагальнених ознак. У цьому разі всі обчислювальні власні числа $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ будуть близькі до одиниці, і узагальнені фактори за складом стають дуже схожими на елементарні ознаки. Проведення факторного аналізу втрачає сенс.

Перевірка значимості матриці парних кореляцій здійснюється за допомогою критерія Уілкса $-x^2$, його спостережуване значення оцінюється за допомогою формули:

$$x_H^2 = -\left(n - \frac{1}{6}(2m + 5)\right)\ln|R|, \tag{1.15}$$

де $R -$ матриця парних кореляцій; $n, m -$ відповідно число спостережуваних об’єктів і число елементарних ознак в аналізі.

Розрахунок x^2 $-$критерія не представляє труднощів, так як вжу на початковій стадії дослідження, після отримання власних чисел λ_j визначник $|R|$ легко знаходиться з добутку

$$\prod_{j=1}^{m} \lambda_j,$$

тобто

$$|R| = \lambda_1 * \lambda_2 * \ldots * \lambda_m. \tag{1.16}$$

Спостережуване значення критерія Уилкса при заданому рівні параметра α і числі ступенів свободи $v = 1/2\, m(m-1)$. Значимість кореляційної матриці підтверджується при $x_H^2 > x_{\alpha,v}^2$.

Критерій x^2використовується з трішки іншою структурою і в оцінці достатності числа виділених загальних ознак(факторів). Для метода головних компонентів розрахунок $x_H^2 -$ критерія здійснюється по формулі Бартлетта:

$$x_H^2 = -\left(n - \frac{1}{6}(2m+5) - \frac{2}{3}r\right)\ln R_{m-r}\,, \tag{1.17}$$

де $r -$ число залишених в аналізі головних компонент F_r;

$$R_{m-r} = \frac{|R|}{\lambda_1 * \lambda_2 * \dots * \lambda_m \left(\frac{m - \lambda_1 - \lambda_2 - \cdots - \lambda_m}{m-r}\right)^{m-r}} \tag{1.18}$$

Число ступенів свободи для x^2 буде $v = 1/2\left((m-r)\text{^}2 - m - r - 1\right)$, спостережуване значення зрівнюється з табличними $x_{\alpha,v}^2$ і, якщо $x_H^2 < x_{\alpha,v}^2$, приймається припущення про те, що виділення r головних компонент достатньо повністю представляє дисперсію m елементарних ознак і інші головні компоненти: $F_{r+1}, F_{r+2}, \dots, F_m$ в аналізі можуть не розглядатися через незначний рівень їх інформативності. Якщо $x_H^2 > x_{\alpha,v}^2$, то в аналізі для формулювання висновків повинні бути введені додаткові інші головні компоненти.

У факторному аналізі для перевірки гіпотези про достатню кількість узагальнених ознак (чинників) використовують x^2 - критерій Лоулі, що має аналогічне, як і в попередньому випадку, смислове навантаження:

$$x_H^2 = (n-1)\ln\frac{|R^+|}{|R|} \tag{1.19}$$

де $|R^+|$ і $|R|$ є визначниками матриць парних кореляцій: відтвореної ($R^+ = A, A'$) і вихідної.

Критичне значення для x^2 −критерію знаходять за таблицями заданого рівня значущості, α і числа ступенів свободи $= 1/2\,((m-r)^2 - m - r - 1)$. Припущення про достатню кількість спільних чинників підтверджується, коли $x_H^2 < x_{\alpha,v}^2$.

РОЗДІЛ 2 ПОРІВНЯННЯ ЕФЕКТИВНОСТІ ВИРОБНИЧО-ГОСПОДАРСЬКОЇ ДІЯЛЬНОСТІ ПІДПРИЄМСТВ ЗА ЗНАЧЕННЯМИ УЗАГАЛЬНЕНИХ ФАКТОРІВ

Розглянемо детально традиційний підхід до оцінки ефективності виробничої діяльності підприємства [3]. Показники першого рівня, які оцінюють ефективність використання ресурсів підприємства є:

- рентабельність активів;
- рентабельність власного капіталу;
- рентабельність інвестованого капіталу;
- рентабельність продукції;
- рентабельність оперативної діяльності;

Показники другого рівня характеризують ефективність використання окремих видів ресурсів і дають змогу менеджерам зосередити увагу на використанні окремого ресурсу. Доцільно аналізувати їх динаміку, а не абсолютний рівень, оскільки абсолютне значення, наприклад рентабельності основних чи оборотних засобів, не можна порівнювати з альтернативною дохідністю, наприклад за банківськими депозитами, оскільки отриманий підприємством прибуток є результатом функціонування не лише одного виду активів підприємства, а всієї їх сукупності. Вивчення динаміки зазначених показників дозволяє оцінити, з ефективністю використання яких активів у підприємства виникає найбільше проблем.

Показниками ефективності використання основних засобів є:

- віддача основних засобів;
- рентабельність основних засобів;

Показники ефективності використання трудових ресурсів визначаються як відношення обсягу реалізованої продукції чи товарообороту, чи ефекту у вигляді прибутку, до середньої чисельності працівників чи суми затрат на оплату праці всіх працівників. Найважливішими показниками ефективності використання трудових ресурсів є:

- виробіток;

- трудомісткість;
- продуктивність.

Показники стану та ефективності використання основних фондів включають в себе аналіз основних фондів – активів підприємства, які мають вартісну оцінку і не втрачають матеріально-речову форму в процесі експлуатації. Відіграють значну роль у процесі праці, так як вони у своїй сукупності утворюють виробничо-технічну базу і визначають виробничу потужність підприємства. Оцінюючи ефективність виробничої діяльності підприємства доцільно приділити особливу увагу таким показникам, як «фондовіддача», «фондомісткість» та «фондоозброєність», що характеризують ефективність задіяних основних фондів.

Ефективність виробничої діяльності підприємства є узагальнюючим показником ефективності використання трудових та матеріальних ресурсів. Зобразимо схематично модель оцінки ефективності виробничої діяльності підприємства на рис. 2.1 [3]:

Рис. 2.1 Складові оцінки ефективності виробничої діяльності

Приклад 1

Нехай є матриця Х вихідних даних, що представляє сукупність чотирьох промислових підприємств, оцінених за трьома ознаками: X_1 - рівень виробітку на одного середньорічного працівника; X_2- рівень рентабельності продукції, %; X_3- рівень фондовіддачі основних фондів, грошова одиниця. Треба порівняти ефективність виробничої діяльності даних чотирьох підприємств, по можливості зменшивши кількість показників. Для вирішення задачі будемо використовувати метод головних компонент.

Розв”язання.

Звернемо увагу, що у факторному аналізі вихідна матриця (X) будується як транспонована.

$$X^T = \begin{pmatrix} 9{,}26 & 9{,}38 & 12{,}11 & 10{,}81 \\ 13{,}26 & 10{,}16 & 13{,}72 & 12{,}85 \\ 1{,}45 & 1{,}30 & 1{,}37 & 1{,}65 \end{pmatrix}$$

Кореляційна матриця:

$$R = \frac{1}{n} Z Z^T \tag{2.1}$$

Характеристичне рівняння для знаходження власних значень:

$$|R - \lambda E| = 0 \tag{2.2}$$

Власні вектори:

$$(R - \lambda E)U = 0 \tag{2.3}$$

Нормалізація:

$$V_j = \frac{U_j}{|U_J|} \tag{2.4}$$

Матриця навантажень:

$$A = V\Lambda^{\frac{1}{2}} \quad (2.5)$$

В приведених формулах:

$$Z = \left\| z_{ij} \right\| \quad (2.6)$$

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j} \quad (2.7)$$

$$X^T - \quad (2.8)$$

транспонована матриця вихідних даних розмірності $(n \times m)$.

Схематично алгоритм метода головних компонент має такий вигляд:

1. Стандартизуємо значення досліджуваних ознак:

$z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j}$, отримаємо $Z' = \begin{pmatrix} -0,971 & -0,868 & 1,478 & 0,361 \\ 0,549 & -1,684 & 0,882 & 0,253 \\ 0,076 & -1,069 & -0,534 & 1,527 \end{pmatrix}$.

2. Знайдемо матрицю парних кореляцій $R = \frac{1}{n} Z^T Z$

$$R = \begin{pmatrix} 1 & 0,581 & 0,154 \\ 0,581 & 1 & 0,439 \\ 0,154 & 0,439 & 1 \end{pmatrix}.$$

3. Пошук власних чисел означає рішення характеристичного рівняння $|R - \lambda E| = 0$.

Існує безліч різних підходів для обчислення елементів матриці λ, зупинимося на одному з них, який широко застосовують на практиці та використовують рекурентні співвідношення Фаддєєва: нехай A - деяка симетрична матриця, розмірністю $m \times m$, тоді її визначник знаходимо за слідом матриць, похідних від A:

$A_1 = A$	$P_1 = trA_1$	$B_1 = A_1 - P_1E$
$A_2 = AB_1$	$P_2 = 1/2trA_2$	$B_2 = A_2 - P_2E$
…	…	…
$A_{m-1} = AB_{m-2}$	$P_{m-1} = 1/(m-1)trA_{m-1}$	$B_{m-1} = A_{m-1} - P_{m-1}E$
$A_m = AB_{m-1}$	$P_m = 1/m \cdot trA_m$	$B_m = A_m - P_mE$

На заключному етапі розрахунків $P_m = |A|$, можемо записати характеристичний многочлен

$$P_m(\lambda) = \lambda^m - P_1\lambda^{m-1} - P_2\lambda^{m-2} - \cdots - P_m. \qquad (2.9)$$

Прирівнюючи характеристичний многочлен до нуля, можна отримати його корені, тобто множину значень λ.

У нашому випадку $R = A, A = A_1, P_1 = trA_1 = 1 + 1 + 1 = 3$,

$$B_1 = A_1 - P_1E = \begin{pmatrix} -2 & 0{,}581 & 0{,}154 \\ 0{,}581 & 2 & 0{,}439 \\ 0{,}154 & 0{,}439 & 2 \end{pmatrix};$$

$$A_2 = AB_1 = \begin{pmatrix} 1 & 0{,}581 & 0{,}154 \\ 0{,}581 & 1 & 0{,}439 \\ 0{,}154 & 0{,}439 & 1 \end{pmatrix} \begin{pmatrix} -2 & 0{,}581 & 0{,}154 \\ 0{,}581 & -2 & 0{,}439 \\ 0{,}154 & 0{,}439 & -2 \end{pmatrix} =$$

$$= \begin{pmatrix} -1{,}638 & -0{,}513 & 0{,}102 \\ -0{,}513 & -1{,}469 & -0{,}350 \\ 0{,}101 & -0{,}350 & -1{,}783 \end{pmatrix};$$

$$P_2 = \frac{1}{2trA_2} = \frac{1}{2[(-1{,}638) + (-1{,}469) + (-1{,}783)]} = -2{,}445;$$

$$B_2 = A_2 - P_2E = \begin{pmatrix} 0{,}807 & -0{,}513 & 0{,}101 \\ -0{,}513 & 0{,}976 & -0{,}350 \\ 0{,}101 & -0{,}350 & 0{,}662 \end{pmatrix} =$$

$$= \begin{pmatrix} 0{,}524 & 0 & 0 \\ 0 & 0{,}524 & 0 \\ 0 & 0 & 0{,}524 \end{pmatrix};$$

$$P_3 = \frac{1}{3(3 \times 0{,}524)} = 0{,}524;\ B_3 = A_3 - P_3E = 0$$

Таким чином, $|R| = 0{,}524$ і характеристичний багаточлен буде $\lambda^3 - 3\lambda^2 + 2{,}44\lambda - 0б524 = 0$, звідки $\lambda_1 = 1{,}798, \lambda_2 = 0{,}875, \lambda_3 = 0{,}327$.

$$\lambda = \begin{pmatrix} 1{,}798 & 0 & 0 \\ 0 & 0{,}875 & 0 \\ 0 & 0 & 0{,}327 \end{pmatrix}.$$

4. Матрицю власних векторів визначимо, розв'язуючи системи лінійних рівнянь для кожного із власних чисел (λ_j) відповідно. Оскільки подібні системи рівнянь мають нескінченну множину розв'язків, щоразу одній із невідомих ознак задаватимемо довільне значення, наприклад одиницю:

для λ_1= 1,798 маємо

$$\left.\begin{array}{r} (1-1{,}798)u_{11} + 0{,}581u_{22} + 0{,}154u_{33} = 0 \\ 0{,}581u_{11} + (1-1{,}798)u_{21} + 0{,}439u_{31} = 0 \\ 0{,}154u_{11} + 0{,}439u_{21} + (1-1{,}798)u_{31} = 0 \end{array}\right| \begin{array}{l} u_{11} = 1{,}262 \\ u_{21} = 1{,}469 \\ u_{31} = 1{,}000 \end{array}$$

для λ_2= 0,875 маємо

$$\left.\begin{array}{r} (1-0{,}875)u_{12} + 0{,}581u_{22} + 0{,}154u_{32} = 0 \\ 0{,}581u_{12} + (1-0{,}875)u_{22} + 0{,}439u_{32} = 0 \\ 0{,}154u_{12} + 0{,}439u_{22} + (1-0{,}875)u_{32} = 0 \end{array}\right| \begin{array}{l} u_{31} = 1{,}307 \\ u_{22} = 1{,}799 \\ u_{32} = 1{,}000 \end{array}$$

для $\lambda_3 = 0{,}327$

$$\left.\begin{array}{r} (1-0{,}327)u_{13} + 0{,}581u_{23} + 0{,}154u_{33} = 0 \\ 0{,}581u_{13} + (1-0{,}327)u_{23} + 0{,}439u_{33} = 0 \\ 0{,}154u_{13} + 0{,}39u_{23} + (1-0{,}327)u_{33} = 0 \end{array}\right| \begin{array}{l} u_{13} = 1{,}307 \\ u_{23} = 1{,}779 \\ u_{33} = 1{,}000 \end{array}$$

Матриця власних векторів приймає вид:

$$U = \begin{pmatrix} 1{,}262 & -0{,}144 & 1{,}307 \\ 1{,}469 & -0{,}234 & -1{,}779 \\ 1{,}000 & 1{,}000 & 1{,}000 \end{pmatrix}.$$

5. Оскільки під час розрахунків власні вектори набувають різних шкал вимірювання, слід привести їх до нормованого вигляду $V_j = \frac{U_j}{|U_j|}$:

$$V = \begin{pmatrix} 0{,}579 & -0{,}139 & 0{,}539 \\ 0{,}674 & -0{,}225 & -0{,}734 \\ 0{,}459 & 0{,}964 & 0{,}413 \end{pmatrix}.$$

6. Тепер можемо побудувати матрицю факторного відображення

$$A = V\lambda^{1/2}$$

$$A = \begin{pmatrix} 0{,}579 & -0{,}139 & 0{,}539 \\ 0{,}674 & -0{,}225 & -0{,}734 \\ 0{,}459 & 0{,}964 & 0{,}413 \end{pmatrix} \begin{pmatrix} \sqrt{1{,}798} & 0 & 0 \\ 0 & \sqrt{0{,}875} & 0 \\ 0 & 0 & \sqrt{0{,}327} \end{pmatrix} =$$

	F_1	F_2	F_3
X_1	*0,776*	*-0,130*	*0,308*
X_2	*0,904*	*-0,210*	*-0,420*
X_3	*0,616*	*0,902*	*0,236*

Елементи матриці A – це коефіцієнти частинної кореляції, що характеризують щільність зв'язків X_j – елементарних ознак з F_r – латентними факторами (головними компонентами).

Для матриці A виконується умова $A^T A = \lambda$, тобто $\sum_j a_{jr}^2 = \lambda_r$, або $\sum_j a_{j1}^2 = 0{,}776^2 + 0{,}904^2 + 0{,}616^2 - 1{,}798$; $\sum_j a_{j2}^2 = 0{,}875$; $\sum_j a_{j3}^2 = 0{,}327$.

Матриця А, що представляє собою ознакову структуру кожної із головних компонент, дозволяє визначати назви головних компонент:

F_1- пояснює (1,798/3 = 0,599) більше 60% спільної дисперсії ознак – може бути названа "ефективністю виробництва";

F_2- пояснює (0,875/3 = 0,292) біля 30% спільної дисперсії ознак - назвемо її «ефективність використання основних виробничих фондів»;

F_3- пояснює (0,327/3 = 0,109) біля 11% спільної дисперсії. З огляду низької цінності цієї головної компоненти в подальшому аналізі вона може не розглядатись.

Для точнішої оцінки структури кожної із головних компонент може застосовуватись спеціальний коефіцієнт рівня інформативності складових іі елементарних ознак:

$$K_u = \frac{\sum_j a_{jr}^2}{\sum_{j=1}^{m} a_{jr}^2}, \tag{2.10}$$

де $\sum_{j=1}^{m} a_{jr}^2$ – сума квадратів всіх значень навантажень елементрних ознак для головної компоненти F_r; $\sum_j a_{jr}^2$ – сума квадратів навантажень тіх елементарних ознак, які найбільш значимі і в основному формуюьб назву головної компоненти (F_r).

Обчислена матриця А дозволяє записати рівняння зв'язку елементарних ознак з головними компонентами:

$$Z_1 = 0{,}776F_1 - 0{,}130F_2 + 0{,}308F_3;$$
$$Z_2 = 0{,}904F_1 - 0{,}210F_2 + 0{,}420F_3;$$
$$Z_3 = 0{,}616F_1 - 0{,}902F_2 + 0{,}236F_3,$$

і навпаки – головних компонент з елементарними ознаками:

$$F_1 = 1/1{,}798(0{,}776Z_1 + 0{,}904Z_2 + 0{,}616Z_3);$$
$$F_2 = 1/1{,}875(-0{,}130Z_1 - 0{,}210Z_2 + 0{,}236Z_3).$$

7. Обчислення матриці F дає можливість отримати значення головних компонент по кожному об'єкту $F = A^{-1}Z^T$, який спостерігається:

$$F = \begin{pmatrix} 0{,}542 & 0{,}507 & 0{,}196 \\ -0{,}776 & -0{,}010 & 0{,}994 \\ 1{,}554 & -1{,}283 & -0{,}075 \end{pmatrix} \begin{pmatrix} -0{,}971 & -0{,}868 & 1{,}478 & 0{,}361 \\ 0{,}549 & -1{,}684 & 0{,}882 & 0{,}253 \\ 0{,}076 & -1{,}069 & -0{,}534 & 1{,}527 \end{pmatrix} =$$

=

	n_1	n_2	n_3	n_4
F_1	-0,233	-1,533	1,143	0,623
F_2	0,823	-0,372	1,687	1,236
F_3	-2,218	0,892	1,205	0,121

Перевірка: $\sum f_{ri} = 0$.

Після аналізу значень головних компонент F_1 і F_2 вибираємо **третє** підприємство.

Приклад 2

Показники	Номер підприємства									
	1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10	11
1. Вироблено продукції, штук	10000	12000	15000	14000	13500	14500	13200	14100	14500	15000
2. Ціна одиниці продукції, грн.	40	45	35	50	40	37	42	45	37	42
3. Собівартість одиниці продукції,	30	35	30	42	31	29	35	37	30	32
4. Середня вартість основних засобів,	310000	315000	320000	330000	335000	340000	327000	333000	340000	337000
5. Кількість працівників, чол.	20	25	30	35	28	37	40	42	45	38
6. Фонд оплати праці працівників,	100000	137000	140000	155000	147000	185000	220000	230000	240000	235000

Необхідно порівняти ефективність виробничої діяльності 10 підприємств за значеннями наступних 4 показників:

X_1 - Продуктивність праці одного працівника (V)

X_2 - Рівень рентабельності обсягу продукції (РП)

X_3 - Капітоломісткість основних засобів (1/f)

x_4 - Капіталоозброєність праці одного працівника (W).

По можливості, зменшити кількість показників для порівняння. Задачу вирішувати методом головних компонентів.

Розв'язання.

Розрахуємо значення показників:

$$V = \frac{\text{Вартість виробленої продукції, грн}}{\text{Кількість працівників, чол}} \tag{2.11}$$

$$Р\Pi = \frac{\text{Ціна одиниці продукції, грн.} - \text{Собівартість одиниці продукції, грн}}{\text{Ціна одиниці продукції, грн.}} \tag{2.12}$$

$$1/f = \frac{\text{Вартість ОЗ, грн}}{\text{Вартість виробленої продукції, грн}} \tag{2.13}$$

$$W = \frac{\text{Вартість ОЗ, грн}}{\text{Кількість працівників, чол}} \tag{2.14}$$

Отримали початкова матрицю:

$$X = \begin{pmatrix} 20000.000000 & 0.250000 & 0.775000 & 15500.000000 \\ 21600.000000 & 0.222222 & 0.583333 & 12600.000000 \\ 17500.000000 & 0.142857 & 0.609524 & 10666.666667 \\ 20000.000000 & 0.160000 & 0.471429 & 9428.571429 \\ 19285.714286 & 0.225000 & 0.620370 & 11964.285714 \\ 14500.000000 & 0.216216 & 0.633737 & 9189.189189 \\ 13860.000000 & 0.166667 & 0.589827 & 8175.000000 \\ 15107.142857 & 0.177778 & 0.524823 & 7928.571429 \\ 11922.222222 & 0.189189 & 0.633737 & 7555.555556 \\ 16578.947368 & 0.238095 & 0.534921 & 8868.421053 \end{pmatrix}$$

Далі переходимо до стандартизованої матриці:

$$Z = \begin{pmatrix} 0.936056 & 1.407638 & 2.169101 & 2.114227 \\ 1.441248 & 0.643909 & -0.175367 & 0.960080 \\ 0.146694 & -1.538173 & 0.144995 & 0.190649 \\ 0.936056 & -1.066844 & -1.544187 & -0.302090 \\ 0.710524 & 0.720282 & 0.277670 & 0.707078 \\ -0.800540 & 0.478778 & 0.441173 & -0.397360 \\ -1.002617 & -0.883549 & -0.095938 & -0.800988 \\ -0.608838 & -0.578057 & -0.891069 & -0.899062 \\ -1.614460 & -0.264309 & 0.441173 & -1.047516 \\ -0.144123 & 1.080325 & -0.767551 & -0.525019 \end{pmatrix}$$

Знайдемо матрицю кореляцій для стандартизованих даних:

$$R = \begin{pmatrix} 1.000000 & 0.295412 & 0.026074 & 0.782861 \\ 0.295412 & 1.000000 & 0.492226 & 0.542161 \\ 0.026074 & 0.492226 & 1.000000 & 0.639064 \\ 0.782861 & 0.542161 & 0.639064 & 1.000000 \end{pmatrix}$$

Розрахуємо власні значення:

$$\det(R - \lambda E) = 0. \tag{2.15}$$

де R - матриця кореляцій, λ - власне значення, E - одинична матриця. Отримаємо:

[2.43852829 1.03323065 0.526217165 0.00202390088].

Перевіримо значущість отриманої матриці за допомогою критерію Уілкса:

$$|R| = 0.002683;$$

$$x_H^2 = 46.37869643856727;$$

$$v = 6\ - \text{кількість ступенів свободи.}$$

При рівні значущості 0,01:

$$x_{\alpha,v}^2 = 16.812;$$

$$x_H^2 > x_{\alpha,v}^2;$$

Таким чином з надійністю 0,99 підтверджено значущість кореляційної матриці R.

Далі перейдемо до матриці нормованих власних векторів:

$$V = \begin{pmatrix} 0.44159818 & 0.71197184 & 0.01469567 & 0.54577576 \\ 0.47552961 & -0.28579536 & -0.83191533 & 0.01046362 \\ 0.44490348 & -0.62145873 & 0.47331454 & 0.43797635 \\ 0.61719003 & 0.15876446 & 0.28926506 & -0.71428009 \end{pmatrix}.$$

Вибираємо тільки ті власні вектори, які відповідають найбільшим власним значенням.

Перейдемо до пошуку головних компонент. Для першої головної компоненти F_1 пояснювальна дисперсія : $2.4385/4 \approx 0.60963$ (61%). Тоді беремо ще другу головну компоненту F_2: 1.0332/4=0.258 (25.8%). Сумарна пояснювальна дисперсія двох головних компонент дорівнює: 86.8%. Тоді зупиняємося і третю головну компоненту не беремо.

Вибрані 2 головні компоненти:

$$\begin{array}{cc} F_1 & F_2 \end{array}$$

$$\begin{pmatrix} 0.44159818 & 0.71197184 \\ 0.47552961 & -0.28579536 \\ 0.44490348 & -0.62145873 \\ 0.61719003 & 0.15876446 \end{pmatrix}.$$

Отримаємо матрицю навантажень:

$$A = \begin{pmatrix} 0.689590349 & 0.723704805 & 0.0106603613 & 0.0245532431 \\ 0.742576953 & -0.290505137 & -0.603478262 & 0.000470735188 \\ 0.694751826 & -0.631700085 & 0.343346284 & 0.0197035864 \\ 0.963790852 & 0.161380820 & 0.209835265 & -0.0321338797 \end{pmatrix}$$

У першому стовпчику цієї матриці найбільші значення навантажень $A_{1.2}, A_{1.4}$, ці навантаження відповідають вихідним ознакам X_2, X_4.Тому назвемо першу головну компоненту F_1 – "ефективність використання матеріальних ресурсів і виробничих фондів". У другому стовпчику цієї матриці найбільші значення навантаження $A_{2.1}$, це навантаження відповідає вихідній ознаці X_1, тому назвемо другу головну компоненту F_2 – "ефективність використання трудових ресурсів "

Перевіримо коефіцієнт інформативності.

Коефіцієнт інформативності для 1-ї головної компоненти:

$$К_{\mathrm{i}} = \frac{0.742576953^2 + 0.963790852^2}{2.43852829} \approx 0.61 < 0.75$$

Коефіцієнт інформативності для 2-ї головної компоненти:

$$К_{\mathrm{i}} = \frac{0.723704805^2 + (-0.631700085)^2}{1.03323065} \approx 0.85 > 0.75$$

Коефіцієнт інформативності для першої головної компоненти не достатньо надійний.

Створимо нову матрицю даних. Перемножимо стандартизовані дані на вибрані власні вектори.

$$\begin{array}{cc} & \begin{array}{cc} F_1 & F_2 \end{array} \\ F^T = & \begin{pmatrix} 5.235434 & -0.760523 \\ 2.275503 & 1.121697 \\ -0.756572 & 0.492184 \\ -1.510696 & 1.914063 \\ 1.899222 & 0.243669 \\ -0.272981 & -1.061258 \\ -2.186136 & -0.537583 \\ -2.334681 & 0.145107 \\ -2.012666 & -1.539347 \\ -0.336428 & -0.018009 \end{pmatrix} \end{array}$$

Виходячи зі значень цієї матриці (головних компонент) знаходимо найбільш ефективні підприємства - це 1,2 та 5. До того ж найбільш надійним виглядає положення другого підприємства.

В Додатку А можна ознайомитися з програмою для вирішення задачі методом головних факторів і її результатами.

Приклад 3

У зв'язку з тим, що в 2 прикладі коефіцієнт інформативності не є надійним для першої компоненти F_1, спробуємо змінити набір показників ефективності. А саме, у якості показника X_3 **замість капіталомісткості** основних засобів, візьмемо **капіталовіддачу** основних засобів:

X_1 - Продуктивність праці одного працівника (V)

X_2 - Рівень рентабельності обсягу продукції (РП)

X_3 – **Капіталовіддача** основних засобів (f)

X_4 - Капіталоозброєність праці одного працівника (W).

Розв'язання.

Розрахуємо значення показників:

$$V = \frac{\text{Вартість виробленої продукції, грн}}{\text{Кількість працівників, чол}}$$

$$Р\text{П} = \frac{\text{Ціна одиниці продукції, грн.} - \text{Собівартість одиниці продукції, грн}}{\text{Ціна одиниці продукції, грн.}}$$

$$f = \frac{\text{Вартість виробленої продукції, грн}}{\text{Вартість ОЗ, грн}}$$

$$W = \frac{\text{Вартість ОЗ, грн}}{\text{Кількість працівників, чол}}$$

Отримали початкову матрицю:

$$X = \begin{pmatrix} 20000.000000 & 0.250000 & 1.290323 & 15500.000000 \\ 21600.000000 & 0.222222 & 1.714286 & 12600.000000 \\ 17500.000000 & 0.142857 & 1.640625 & 10666.666667 \\ 20000.000000 & 0.160000 & 2.121212 & 9428.571429 \\ 19285.714286 & 0.225000 & 1.611940 & 11964.285714 \\ 14500.000000 & 0.216216 & 1.577941 & 9189.189189 \\ 13860.000000 & 0.166667 & 1.695413 & 8175.000000 \\ 15107.142857 & 0.177778 & 1.905405 & 7928.571429 \\ 11922.222222 & 0.189189 & 1.577941 & 7555.555556 \\ 16578.947368 & 0.238095 & 1.869436 & 8868.421053 \end{pmatrix}.$$

Далі переходимо до стандартизованої матриці:

$$Z = \begin{pmatrix} 0.936056 & 1.407638 & -1.819149 & 2.114227 \\ 1.441248 & 0.643909 & 0.061359 & 0.960080 \\ 0.146694 & -1.538173 & -0.265367 & 0.190649 \\ 0.936056 & -1.066844 & 1.866300 & -0.302090 \\ 0.710524 & 0.720282 & -0.392599 & 0.707078 \\ -0.800540 & 0.478778 & -0.543404 & -0.397360 \\ -1.002617 & -0.883549 & -0.022353 & -0.800988 \\ -0.608838 & -0.578057 & 0.909079 & -0.899062 \\ -1.614460 & -0.264309 & -0.543404 & -1.047516 \\ -0.144123 & 1.080325 & 0.749536 & -0.525019 \end{pmatrix}.$$

Знайдемо матрицю кореляцій для стандартизованих даних:

$$R = \begin{pmatrix} 1.0000000 & 0.2954122 & 0.05421261 & 0.78286087 \\ 0.2954122 & 1.00000000 & -0.46659954 & 0.54216077 \\ 0.05421261 & -0.46659954 & 1.00000000 & -0.5652167 \\ 0.78286087 & 0.54216077 & -0.5652167 & 1.0000000 \end{pmatrix}.$$

Розрахуємо власні значення.

$$\det(R-\lambda E)=0,$$

де R - матриця кореляцій, λ - власне значення, E - одинична матриця. Отримаємо:

[2.35806299 1.11116633 0.52267245 0.00809824]

Перевіримо критерій Уілкса:

$$|R| = 0{,}01109059;$$

$$x_H^2 = 35.628339;$$

$$v = 6 \ -\text{кількість ступенів свободи.}$$

При рівні значущості 0,01:

$$x_{\alpha,v}^2 = 16.812;$$

$$x_H^2 > x_{\alpha,v}^2;$$

Таким чином з надійністю 0,99 підтверджено значущість кореляційної матриці R.

Далі перейдемо до матриці нормованих власних векторів:

$$V = \begin{pmatrix} 0.45020525 & 0.68245792 & 0.5726927 & 0.05991246 \\ 0.4883416 & -0.25977258 & 0.01281061 & -0.83299254 \\ -0.40997613 & 0.66681371 & -0.42474601 & -0.45482949 \\ 0.62510582 & 0.14875775 & -0.7010348 & 0.30929606 \end{pmatrix}.$$

Вибираємо тільки ті власні вектори, які відповідають найбільшим власним значенням.

Перейдемо до пошуку головних компонент. Для першої головної компоненти F_1 пояснювальна дисперсія : $2.3581/4 \approx 0.5895$ (58.95%).Тоді розраховуємо пояснювальну дисперсію для другої головної компоненти F_2: $1.1112/4 \approx 0.278$ (27.8%). Сумарна пояснювальна дисперсія двох головних компонент дорівнює: 86.75%. Тоді зупиняємося і третю головну компоненту не беремо.

Вибираємо тільки ті власні вектори, які відповідають найбільшим власним значенням.

Вибрані 2 головні компоненти:

$$\begin{array}{cc} F_1 & F_2 \end{array}$$
$$\begin{pmatrix} 0.45020525 & 0.68245792 \\ 0.4883416 & -0.25977258 \\ -0.40997613 & 0.66681371 \\ 0.62510582 & 0.14875775 \end{pmatrix}$$

Отримаємо матрицю навантажень.

$$A = \begin{pmatrix} 0.69133453 & 0.71939168 & 0.04331437 & 0.05153674 \\ 0.74989666 & -0.27383115 & -0.60222103 & 0.00115283 \\ -0.62955877 & 0.70290084 & -0.32882393 & -0.03822299 \\ 0.95991161 & 0.15680833 & 0.22360895 & -0.06308628 \end{pmatrix}$$

У першому стовпчику цієї матриці найбільші значення мають навантаження $A_{1.1}, A_{1.2}, A_{1.4}$, ці навантаження відповідають вихідним ознакам X_1, $X_2, X_{4.}$ Тому назвемо першу головну компоненту F_1 – “ефективність використання трудових, матеріальних ресурсів і виробничих фондів”. У другому стовпчику цієї матриці найбільші значення мають навантаження $A_{2.1}, A_{2.3}$ ці навантаження відповідають вихідним ознакам X_1, X_3, тому назвемо другу головну компоненту F_2 – “продуктивність праці і фондовіддача”.

Розрахуємо значення коефіцієнту інформативності:

Коефіцієнт інформативності для 1-ї головної компоненти F_1:

$$К_\mathrm{i} = \frac{(0.69133453^2 + 0.74989666^2 + 0.95991161^2)}{2.35806299} \approx 0.83 > 0.75.$$

Коефіцієнт інформативності для 2-ї головної компоненти F_2:

$$К_\mathrm{i} = \frac{0.71939168^2 + 0.70290084^2}{1.11116633} \approx 0.91 > 0.75.$$

Створимо нову матрицю даних. Перемножимо стандартизовані дані на вибрані власні вектори:

$$\begin{array}{cc} & \begin{array}{cc} F_1 & F_2 \end{array} \\ F^T = & \begin{pmatrix} 4.877443 & -0.659217 \\ 2.362213 & 1.054177 \\ -0.701986 & 0.370099 \\ -1.617820 & 2.229979 \\ 1.957243 & 0.148827 \\ -0.233732 & -1.151274 \\ -2.110519 & -0.620644 \\ -2.289733 & 0.218309 \\ -1.977754 & -1.635271 \\ -0.265354 & 0.045014 \end{pmatrix} \end{array}$$

Виходячи зі значень головних компонент цієї матриці знаходимо найбільш ефективні підприємства - це 1,2 та 5. До того ж найбільш надійним виглядає положення **другого** підприємства.

В Додатку Б можна ознайомитися з програмою для вирішення задачі методом головних компонент і її результатами.

ВИСНОВКИ

Методи головних компонент і факторного аналізу базуються на загальній ідеї, що зв'язки елементарних ознак $(X_1, X_2, \dots, X_n)$ - це результат впливу порівняно невеликої кількості неявних, тобто латентних, чинників$(F_1, F_2, \dots, F_r)$. Метод головних компонент гіпотетично дає змогу пояснити всю, на 100%, варіацію $X_{j\,j}$ через латентні чинники. Класичні методи факторного аналізу допускають наявність неприхованої характерності ознак і від самого початку орієнтовані на пояснення не всієї, а тільки певної частини варіації елементарних ознак, що залишається за вирахуванням характерності.

Оскільки кількість латентних чинників r зазвичай значно менша за кількість елементарних ознак m, основним завданням факторного аналізу вважають стиснення аналізованого ознакового простору і перехід від масиву вихідних даних розміром $n \times m$ до аналітичних даних у матрицях факторного відображення $(m \times r)$ і значень спільних факторів $(n \times r)$ за умови $r < m$.

Складним і завжди з неоднозначним розв'язанням залишається завдання визначення достатності числа й інтерпретації загальних факторів. Результат приймається суб'єктивно, певною мірою інтерпретація полегшується можливістю обертання факторного простору R^F і одержання простішої структури факторів.

При порівнянні ефективності виробничої діяльності підприємств у прикладах 2 і 3 вдалося зменшити кількість показників із 4 до 2 майже без втрати інформації, тобто, відбулося нібито стиснення інформації про ефективність роботи підприємств. За двома показниками зробити вибір найкращих підприємств набагато легше. В результаті за значеннями двох головних компонент з 10 підприємств в прикладі 2 було обрано перше, друге та п’яте підприємства. Зробивши перевірку на інформативність отримали такі результати: коефіцієнт інформативності для 1-ї головної компоненти: 0.61; коефіцієнт інформативності для 2-ї головної компоненти: 0.85. Як бачимо з результатів, коефіцієнт інформативності для 1-ї головної компоненти менше за 0.75. Тому було розглянуто приклад 3, в якому було змінено третій показник ефективності виробництва: замість капіталомісткості була введена капіталовіддача. Зробивши перевірку на інформативність головних компонент після заміни X_3 отримали такі результати: коефіцієнт інформативності для 1-ї головної компоненти став 0.8319206, коефіцієнт інформативності для 2-ї головної компоненти: 0.9103945. Як бачимо з результатів, дані отримані в прикладі 3 є більш надійними, тому що для обох головних компонент значення коефіцієнтів інформативності більше ніж 0.75. За значеннями головних компонент у прикладі 3 теж було обрано перше, друге та п’яте підприємства. Найбільш привабливим здається друге підприємство.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Яровий А. Т., Страхов Є. М. Багатовимірний статистичний аналіз. Одеса : Одеський націон. університет імені І. І. Мечникова, 2016. 148 с.
2. Костюк В. О., Гайденко С. М. Статистика підприємств міського господарства. Харків : ХНУМГ ім. О.М.Бекетова, 2019. 41 с.
3. Гречко А. В., Гречухін А. С. Оцінка ефективності виробничої діяльності підприємства. Київ : Київ. політехн. ін-т ім. Ігоря Сікорського, 2016. 7 с.
4. Zaki M. J., Jr W. M. Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, 2014. 660 p.
5. Introduction to statistical learning: with applications in R / G. James et al. New York : Springer, 2013. 418 p.
6. Toppr:better learning for better results. *Toppr:Better learning for better results.* URL: https://www.toppr.com.
7. Kaggle: your machine learning and data science community. *Kaggle: Your Machine Learning and Data Science Community*. URL: https://www.kaggle.com.
8. Кафедра програмного забезпечення Дніпровського державного технічного університету. *Кафедра програмного забезпечення Дніпровського державного технічного університету*. URL: http://pzs.dstu.dp.ua.
9. What is factor analysis? (plus 5 methods for conducting it) | indeed.com. *What Is Factor Analysis? (Plus 5 Methods for Conducting It) | Indeed.com*. URL: https://www.indeed.com/career-advice/career-development/factor-analysis.

ДОДАТКИ

ДОДАТОК А

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from scipy.stats import chi2

# Задані дані
data = {
    'Номер підприємства': [i for i in range(1, 11)],
    'Продукція, шт.': [10000, 12000, 15000, 14000, 13500, 14500, 13200, 14100,
14500, 15000],
    'Ціна, грн.': [40, 45, 35, 50, 40, 37, 42, 45, 37, 42],
    'Собівартість, грн.': [30, 35, 30, 42, 31, 29, 35, 37, 30, 32],
    'Вартість ОЗ, грн.': [310000, 315000, 320000, 330000, 335000, 340000,
327000, 333000, 340000, 337000],
    'Працівники, чол.': [20, 25, 30, 35, 28, 37, 40, 42, 45, 38],
    'ФОП, грн.': [100000, 137000, 140000, 155000, 147000, 185000, 220000,
230000, 240000, 235000]
}

df = pd.DataFrame(data)

# Розрахунок показників діяльності
df['V'] = df['Ціна, грн.']*df['Продукція, шт.'] / df['Працівники, чол.']
df['РП'] = (df['Ціна, грн.'] - df['Собівартість, грн.']) / df['Ціна, грн.']
df['f'] = df['Вартість ОЗ, грн.']/(df['Ціна, грн.']*df['Продукція, шт.'])
df['W'] = df['Вартість ОЗ, грн.'] / df['Працівники, чол.']

# Вибираємо колонки для PCA
```

```
pca_data = df[['V', 'РП', 'f', 'W']]

# Порівняння ефективності виробничо-господарської діяльності
efficiency_comparison = df[['Номер підприємства', 'V', 'РП', 'f', 'W']]

# Стандартизуємо дані
df_std = (pca_data - pca_data.mean()) / pca_data.std()

# Розраховуємо коваріаційну матрицю
cov_matrix = np.cov(df_std, rowvar=False)

# Знаходимо власні вектори та власні значення
eig_vals, eig_vecs = np.linalg.eig(cov_matrix)

# Розмірність
n = len(df)
m = 4

# Розрахунок критерію Уілкса
R = np.linalg.det(cov_matrix)
Wilks_lambda = np.prod(eig_vals)
chi2_Wilks_df = 0.5 * m * (m - 1)  # Ступені свободи

# Обчислення спостережуваного значення критерію Уілкса
chi2_Wilks = -(n - (1/6) * (2 * m + 5)) * np.log(Wilks_lambda)

# Табличне значення хі-квадрат для заданого рівня значущості
alpha = 0.01
chi2_Wilks_table = chi2.ppf(1 - alpha, chi2_Wilks_df)
```

```
#Обчислення пояснювальної дисперсії
explained_variance_ratio = eig_vals / np.sum(eig_vals)

explained_variance_ratio_pca1 = explained_variance_ratio[0]
explained_variance_ratio_pca2 = explained_variance_ratio[1]

# Сортуємо власні значення в спадаючому порядку
sorted_indices = np.argsort(eig_vals)[::-1]
eig_vals = eig_vals[sorted_indices]
eig_vecs = eig_vecs[:, sorted_indices]

# Вибираємо кількість головних компонент
num_components = 4
top_eig_vecs = eig_vecs[:, :num_components]

# Нормалізуємо власні вектори
norm_eig_vecs = top_eig_vecs / np.linalg.norm(top_eig_vecs, axis=0)

# Розраховуємо матрицю навантажень
loading_matrix = np.dot(norm_eig_vecs, np.diag(np.sqrt(eig_vals)))

# Виведення результатів
print("\nРезультати порівняння ефективності виробничо-господарської
діяльності:")
print(efficiency_comparison)

# Виведення стандартизованих даних для показників діяльності
print("Стандартизовані дані для показників діяльності:")
print(df_std)
```

```
# Вивід кореляційної матриці
print("Кореляційна матриця:")
print(df[['V', 'РП', 'f', 'W']].corr())

# Вивід власних векторів та значень
print("\nВласні значення:")
print(eig_vals)
print("\nВласні вектори:")
print(eig_vecs)

# Перевірка значущості критерію Уілкса
if chi2_Wilks > chi2_Wilks_table:
    print(f"\nКритерій Уілкса: {chi2_Wilks} > {chi2_Wilks_table}. Значущий
ефект.")
else:
    print(f"\nКритерій Уілкса: {chi2_Wilks} <= {chi2_Wilks_table}. Ефект не
значущий.")

print("\nПояснювальна дисперсія для PCA1:", explained_variance_ratio_pca1)
print("Пояснювальна дисперсія для PCA2:", explained_variance_ratio_pca2)

# Вивід матриці навантажень
print("\nМатриця навантажень:")
print(loading_matrix)

# Квадрати навантажень для 1-ї головної компоненти
info_coef_1 = (loading_matrix[1, 0]**2 + loading_matrix[3, 0]**2) / eig_vals[0]

# Квадрати навантажень для 2-ї головної компоненти
info_coef_2 = (loading_matrix[0, 0]**2 + loading_matrix[2, 1]**2) / eig_vals[1]
```

```
print("\nКоефіцієнт інформативності для 1-ї головної компоненти:", info_coef_1)
print("Коефіцієнт інформативності для 2-ї головної компоненти:", info_coef_2)

# Проекція на новий простір
pca_result = df_std.dot(loading_matrix)

# Додаємо результат PCA до вихідного DataFrame
df['Ефективність використання матеріальних ресурсів і виробничих фондів'] = pca_result.iloc[:, 0]
df['Ефективність використання трудових ресурсів'] = pca_result.iloc[:, 1]

# Видалення непотрібних стовбців
df.drop(['Продукція, шт.', 'Ціна, грн.', 'Собівартість, грн.', 'Вартість ОЗ, грн.', 'Працівники, чол.', 'ФОП, грн.'], axis=1, inplace=True)

# Результат PCA
print("\nРезультат PCA:")
print(df[['Ефективність використання матеріальних ресурсів і виробничих фондів', 'Ефективність використання трудових ресурсів']])
```

```
Результати порівняння ефективності виробничо-господарської діяльності:
   Номер підприємства             V        РП         f            W
0                   1  20000.000000  0.250000  0.775000  15500.000000
1                   2  21600.000000  0.222222  0.583333  12600.000000
2                   3  17500.000000  0.142857  0.609524  10666.666667
3                   4  20000.000000  0.160000  0.471429   9428.571429
4                   5  19285.714286  0.225000  0.620370  11964.285714
5                   6  14500.000000  0.216216  0.633737   9189.189189
6                   7  13860.000000  0.166667  0.589827   8175.000000
7                   8  15107.142857  0.177778  0.524823   7928.571429
8                   9  11922.222222  0.189189  0.633737   7555.555556
9                  10  16578.947368  0.238095  0.534921   8868.421053
Стандартизовані дані для показників діяльності:
          V        РП         f         W
0  0.936056  1.407638  2.169101  2.114227
1  1.441248  0.643909 -0.175367  0.960080
2  0.146694 -1.538173  0.144995  0.190649
3  0.936056 -1.066844 -1.544187 -0.302090
4  0.710524  0.720282  0.277670  0.707078
5 -0.800540  0.478778  0.441173 -0.397360
6 -1.002617 -0.883549 -0.095938 -0.800988
7 -0.608838 -0.578057 -0.891069 -0.899062
8 -1.614460 -0.264309  0.441173 -1.047516
9 -0.144123  1.080325 -0.767551 -0.525019
Кореляційна матриця:
           V        РП         f         W
V   1.000000  0.295412  0.026074  0.782861
РП  0.295412  1.000000  0.492226  0.542161
f   0.026074  0.492226  1.000000  0.639064
W   0.782861  0.542161  0.639064  1.000000
```

```
Власні значення:
[2.43852829e+00 1.03323065e+00 5.26217165e-01 2.02390088e-03]

Власні вектори:
[[ 0.44159818  0.71197184  0.01469567  0.54577576]
 [ 0.47552961 -0.28579536 -0.83191533  0.01046362]
 [ 0.44490348 -0.62145873  0.47331454  0.43797635]
 [ 0.61719003  0.15876446  0.28926506 -0.71428009]]

Критерій Уілкса: 46.37869643856727 > 16.811893829770927. Значущий ефект.

Пояснювальна дисперсія для PCA1: 0.609632071986946
Пояснювальна дисперсія для PCA2: 0.0005059752197859791

Матриця навантажень:
[[ 6.89590349e-01  7.23704805e-01  1.06603613e-02  2.45532431e-02]
 [ 7.42576953e-01 -2.90505137e-01 -6.03478262e-01  4.70735188e-04]
 [ 6.94751826e-01 -6.31700085e-01  3.43346284e-01  1.97035864e-02]
 [ 9.63790852e-01  1.61380820e-01  2.09835265e-01 -3.21338797e-02]]

Коефіцієнт інформативності для 1-ї головної компоненти: 0.607051943941079
Коефіцієнт інформативності для 2-ї головної компоненти: 0.8464517106039812
```

```
Результат PCA:
   Ефективність використання матеріальних ресурсів і виробничих фондів  \
0                                           5.235434
1                                           2.275503
2                                          -0.756572
3                                          -1.510696
4                                           1.899222
5                                          -0.272981
6                                          -2.186136
7                                          -2.334681
8                                          -2.012666
9                                          -0.336428

   Ефективність використання трудових ресурсів
0                                    -0.760523
1                                     1.121697
2                                     0.492184
3                                     1.914063
4                                     0.243669
5                                    -1.061258
6                                    -0.537583
7                                     0.145107
8                                    -1.539347
9                                    -0.018009
```

ДОДАТОК Б

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from scipy.stats import chi2

# Задані дані
data = {
    'Номер підприємства': [i for i in range(1, 11)],
    'Продукція, шт.': [10000, 12000, 15000, 14000, 13500, 14500, 13200, 14100,
14500, 15000],
    'Ціна, грн.': [40, 45, 35, 50, 40, 37, 42, 45, 37, 42],
    'Собівартість, грн.': [30, 35, 30, 42, 31, 29, 35, 37, 30, 32],
    'Вартість ОЗ, грн.': [310000, 315000, 320000, 330000, 335000, 340000,
327000, 333000, 340000, 337000],
    'Працівники, чол.': [20, 25, 30, 35, 28, 37, 40, 42, 45, 38],
    'ФОП, грн.': [100000, 137000, 140000, 155000, 147000, 185000, 220000,
230000, 240000, 235000]
}

df = pd.DataFrame(data)

# Розрахунок показників діяльності
df['V'] = df['Ціна, грн.']*df['Продукція, шт.'] / df['Працівники, чол.']
df['РП'] = (df['Ціна, грн.'] - df['Собівартість, грн.']) / df['Ціна, грн.']
df['f'] = df['Ціна, грн.']*df['Продукція, шт.']/df['Вартість ОЗ, грн.']
df['W'] = df['Вартість ОЗ, грн.'] / df['Працівники, чол.']

# Вибираємо колонки для PCA
```

```
pca_data = df[['V', 'РП', 'f', 'W']]

# Порівняння ефективності виробничо-господарської діяльності
efficiency_comparison = df[['Номер підприємства', 'V', 'РП', 'f', 'W']]

# Стандартизуємо дані
df_std = (pca_data - pca_data.mean()) / pca_data.std()

# Розраховуємо кореляційну матрицю
cov_matrix = np.cov(df_std, rowvar=False)

# Знаходимо власні вектори та власні значення
eig_vals, eig_vecs = np.linalg.eig(cov_matrix)

# Сортуємо власні значення в спадаючому порядку
sorted_indices = np.argsort(eig_vals)[::-1]
eig_vals = eig_vals[sorted_indices]
eig_vecs = eig_vecs[:, sorted_indices]

# Розмірність
n = len(df)
m = 4
# Розрахунок критерію Уілкса
R = np.linalg.det(cov_matrix)
Wilks_lambda = np.prod(eig_vals)
chi2_Wilks_df = 0.5 * m * (m - 1)  # Ступені свободи

# Обчислення спостережуваного значення критерію Уілкса
chi2_Wilks = -(n - (1/6) * (2 * m + 5)) * np.log(Wilks_lambda)
```

```
# Табличне значення хі-квадрат для заданого рівня значущості
alpha = 0.01
chi2_Wilks_table = chi2.ppf(1 - alpha, chi2_Wilks_df)

# Вибираємо кількість головних компонент
num_components = 4
top_eig_vecs = eig_vecs[:, :num_components]

# Нормалізуємо власні вектори
norm_eig_vecs = top_eig_vecs / np.linalg.norm(top_eig_vecs, axis=0)

# Розраховуємо матрицю навантажень
loading_matrix = np.dot(norm_eig_vecs, np.diag(np.sqrt(eig_vals)))

# Виведення результатів
print("\nРезультати порівняння ефективності виробничо-господарської
діяльності:")
print(efficiency_comparison)

# Виведення стандартизованих даних для показників діяльності
print("\nСтандартизовані дані для показників діяльності:")
print(df_std)

# Вивід кореляційної матриці
print("\nКореляційна матриця:")
print(df[['V', 'РП', 'f', 'W']].corr())

# Вивід власних векторів та значень
print("\nВласні значення:")
print(eig_vals)
```

```
print("\nВласні вектори:")
print(eig_vecs)

# Перевірка значущості критерію Уілкса
if chi2_Wilks > chi2_Wilks_table:
    print(f"\nКритерій Уілкса: {chi2_Wilks} > {chi2_Wilks_table}. Значущий ефект.")
else:
    print(f"\nКритерій Уілкса: {chi2_Wilks} <= {chi2_Wilks_table}. Ефект не значущий.")

#Обчислення пояснювальної дисперсії
explained_variance_ratio = eig_vals / np.sum(eig_vals)

explained_variance_ratio_pca1 = explained_variance_ratio[0]
explained_variance_ratio_pca2 = explained_variance_ratio[1]

print("\nПояснювальна дисперсія для PCA1:", explained_variance_ratio_pca1)
print("Пояснювальна дисперсія для PCA2:", explained_variance_ratio_pca2)

# Вивід матриці навантажень
print("\nМатриця навантажень:")
print(loading_matrix)

# Квадрати навантажень для 1-ї головної компоненти
info_coef_1 = (loading_matrix[0, 0]**2 + loading_matrix[1, 0]**2 + loading_matrix[3, 0]**2) / eig_vals[0]

# Квадрати навантажень для 2-ї головної компоненти
info_coef_2 = (loading_matrix[0, 1]**2 + loading_matrix[2, 1]**2) / eig_vals[1]
```

```
print("\nКоефіцієнт інформативності для 1-ї головної компоненти:",
info_coef_1)
print("Коефіцієнт інформативності для 2-ї головної компоненти:",
info_coef_2)

# Проекція на новий простір
pca_result = df_std.dot(loading_matrix)

# Додаємо результат PCA до вихідного DataFrame
df['Ефективність використання трудових,матеріальних ресурсів і виробничих
фондів'] = pca_result.iloc[:, 0]
df['Продуктивність праці і фондовіддача'] = pca_result.iloc[:, 1]

# Видалення непотрібних стовбців
df.drop(['Продукція, шт.', 'Ціна, грн.', 'Собівартість, грн.', 'Вартість ОЗ, грн.',
'Працівники, чол.', 'ФОП, грн.'], axis=1, inplace=True)

# Результат PCA
print("\nРезультат PCA:")
print(df[['Ефективність використання трудових,матеріальних ресурсів і
виробничих фондів', 'Продуктивність праці і фондовіддача']])
```

```
Результати порівняння ефективності виробничо-господарської діяльності:
   Номер підприємства            V        РП         f             W
0                   1  20000.000000  0.250000  1.290323  15500.000000
1                   2  21600.000000  0.222222  1.714286  12600.000000
2                   3  17500.000000  0.142857  1.640625  10666.666667
3                   4  20000.000000  0.160000  2.121212   9428.571429
4                   5  19285.714286  0.225000  1.611940  11964.285714
5                   6  14500.000000  0.216216  1.577941   9189.189189
6                   7  13860.000000  0.166667  1.695413   8175.000000
7                   8  15107.142857  0.177778  1.905405   7928.571429
8                   9  11922.222222  0.189189  1.577941   7555.555556
9                  10  16578.947368  0.238095  1.869436   8868.421053

Стандартизовані дані для показників діяльності:
          V        РП         f         W
0  0.936056  1.407638 -1.819149  2.114227
1  1.441248  0.643909  0.061359  0.960080
2  0.146694 -1.538173 -0.265367  0.190649
3  0.936056 -1.066844  1.866300 -0.302090
4  0.710524  0.720282 -0.392599  0.707078
5 -0.800540  0.478778 -0.543404 -0.397360
6 -1.002617 -0.883549 -0.022353 -0.800988
7 -0.608838 -0.578057  0.909079 -0.899062
8 -1.614460 -0.264309 -0.543404 -1.047516
9 -0.144123  1.080325  0.749536 -0.525019

Кореляційна матриця:
           V        РП         f         W
V   1.000000  0.295412  0.054213  0.782861
РП  0.295412  1.000000 -0.466600  0.542161
f   0.054213 -0.466600  1.000000 -0.565217
W   0.782861  0.542161 -0.565217  1.000000
```

```
Власні значення:
[2.35806299 1.11116633 0.52267245 0.00809824]

Власні вектори:
[[ 0.45020525  0.68245792  0.05991246  0.5726927 ]
 [ 0.4883416  -0.25977258 -0.83299254  0.01281061]
 [-0.40997613  0.66681371 -0.45482949 -0.42474601]
 [ 0.62510582  0.14875775  0.30929606 -0.7010348 ]]

Критерій Уілкса: 35.26298895892571 > 16.811893829770927. Значущий ефект.

Пояснювальна дисперсія для PCA1: 0.5895157472323302
Пояснювальна дисперсія для PCA2: 0.27779158154700695

Матриця навантажень:
[[ 0.69133453  0.71939168  0.04331437  0.05153674]
 [ 0.74989666 -0.27383115 -0.60222103  0.00115283]
 [-0.62955877  0.70290084 -0.32882393 -0.03822299]
 [ 0.95991161  0.15680833  0.22360895 -0.06308628]]

Коефіцієнт інформативності для 1-ї головної компоненти: 0.8319195706892806
Коефіцієнт інформативності для 2-ї головної компоненти: 0.9103893376803672
```

```
Результат PCA:
   Ефективність використання трудових,матеріальних ресурсів і виробничих фондів  \
0                                           4.877443
1                                           2.362213
2                                          -0.701986
3                                          -1.617820
4                                           1.957243
5                                          -0.233732
6                                          -2.110519
7                                          -2.289733
8                                          -1.977754
9                                          -0.265354

   Продуктивність праці і фондовіддача
0                            -0.659217
1                             1.054177
2                             0.370099
3                             2.229979
4                             0.148827
5                            -1.151274
6                            -0.620644
7                             0.218309
8                            -1.635271
9                             0.045014
```