УДК51-77

# ОБЧИСЛЕННЯ У ТАБЛИЦЯХ «EXCEL» ГОЛОВНИХ КОМПОНЕНТІВ ФАКТОРНОГО АНАЛІЗУ

**КАСЯРУМ Олег Павлович**
*к. ф.-м. н., доцент кафедри інформаційних технологій та вищої математики Черкаського навчально-наукового інституту ДВНЗ «Університет банківської справи»*

**КАСЯРУМ Ярослав Олегович**
*к. пед. н., в.о. доцента кафедри інформаційних технологій та вищої математики Черкаського навчально-наукового інституту ДВНЗ «Університет банківської справи»*

*Анотація. У статті розглядається задача визначення власних значень та власних векторів кореляційної матриці порядку до 10 і більше. Пропонується доступний спосіб обчислень за методом головних компонент факторного аналізу в електронних таблицях «EXCEL». Автори позиціонують статтю, як методичну, для використання в навчальному процесі.*

*Аннотация. В статье рассматривается задача определения собственных значений и собственных векторов корреляционной матрицы порядка до 10 и более. Предлагается доступный способ вычислений по методу главных компонент факторного анализа в электронных таблицах «EXCEL». Авторы позиционируют статью, как методическую, для использования в учебном процессе.*

***Ключові слова:*** *факторний аналіз, метод головних компонент, кореляція, матриця, власний вектор матриці, власні значення матриці, спільність, характерність, навантаження.*

***Ключевые слова:*** *факторный анализ, метод главных компонент, корреляция, матрица, собственный вектор матрицы, собственные значения матрицы, общность, характерность, нагрузки.*

**Постановка проблеми.**

Термін факторний аналіз в навчальній літературі часто використовується для двох різних математичних процедур. Одна з них використовується для підтвердження впливу деяких відомих чинників на досліджувану величину. Їх називають факторами. Таку задачу розглядають у курсах з теорії ймовірності та математичної статистики. Засоби розв'язку цієї задачі добре відомі та доступні для застосування у навчальному процесі. Методи розв'язку легко знайти в таблицях «EXCEL», вони не потребують складних програмних продуктів.

Інша задача, а отже, інша математична процедура, пов'язана із потребою знаходження невідомого числа невідомих факторів, які за припущенням впливають на досліджувані величини. Задача полягає у тому, що для деякого масиву експериментальних даних, які записують у вигляді матриць, потрібно визначити мінімальну кількість факторів, які пояснюють закономірності, що спостерігаються в масиві даних. Такі фактори прийнято називати головними компонентами. Існують інші методи розв'язування другої задачі.. Активний розвиток факторного аналізу відбувся у першій половині 20 століття, але з успіхом використовується в психології та соціології до нині. Для цих галузей науки факторний аналіз залишається одним з головних методів досліджень. Принципи та наукове обґрунтування факторного аналізу можна знайти у класичній тепер монографії Г.Хармана [1].

**Аналіз останніх досліджень та публікацій.** У наш час цікавість до факторного аналізу знову зросла у зв'язку із його використанням у аналізі великої кількості інформації, як для традиційних, так і для задач, які з'явилися відносно недавно. Ось деякі з цих задач:

- Вимірювання латентних (прихованих) величин та побудова нових узагальнених показників (традиційно психологія та соціологія [2, 3]).

- Скорочення числа змінних. Необхідність зменшення розмірності змінних при умові втрати найменшої кількості інформації.

- Виявлення груп взаємозалежних змінних та структури взаємних зв'язків між ними. Факторний аналіз забезпечує лаконічнішу та точнішу модель структури залежностей між змінними.

- Подолання мультиколінеарності змінних у регресійному аналізі.

- Доповнення пропущених значень розріджених матриць для використання у рекомендаційних системах.

- Латентний семантичний аналіз, який дозволяє вести інформаційний пошук та аналізувати велику кількість документів з метою їх індексації, класифікації і таке інше, там де є потреба виявлення головних факторів з масиву інформаційних даних.

Це далеко не повний список задач, які потребують використання факторного аналізу, а лише задачі більш-менш відомі нам.

**Мета статті.** Наша цікавість цим предметом пов'язана із педагогічною діяльністю. Ми бачимо необхідність ознайомлення студентів із сучасними та перспективними методами аналізу, одним з яких є факторний аналіз. Разом з тим буває, що необхідне програмне забезпечення аналізу важко доступне або потребує спеціальної підготовки користувача. Інколи, також існує проблема придбання необхідних програмних засобів, інколи вивчення мов та принципів програмування студентами не передбачено навчальними планами. Тому ми поставили перед собою задачу знайти відносно простий та зрозумілий метод обчислень головних компонент факторного аналізу для матриць великого порядку. Ми знайшли такий спосіб [4, 5] і хочемо його опублікувати.

**Обґрунтування отриманих наукових результатів.** З точки зору математики, метод головних компонент факторного аналізу представляє собою відому задачу матричної алгебри, а саме, знаходження власних значень та власних векторів квадратної матриці [1]. Як правило, у навчальному процесі використовують приклади знаходження цих величин лише для матриць другого та третього порядків [6]. Оскільки відповідна задача розв'язується методом характеристичних рівнянь, які є степеневими, то розв'язок рівнянь степеню більшого ніж третій досить проблемний. Зрозуміло, що існує достатньо програмних засобів, мов програмування і т.д. але існує також проблема їх доступності, про яку ми згадували вище.

Далі ми покажемо, як користуючись досить обмеженими можливостями електронних таблиць «EXCEL» реалізувати знаходження власних чисел та власних векторів для достатньо великих за розмірами матриць. Раніше ми доповідали результати [4, 5] використання подібного способу обчислень для аналізу результатів навчання студентів (матриця порядку 28) так аналізу якості викладання навчальних дисциплін (матриця порядку11).

Знайдений нами спосіб обчислень слідує теоретичним положенням методу головних компонент факторного аналізу [1]. У цьому ми вбачаємо методичну привабливість способу для навчального процесу. Ми усвідомлюємо, що професійні програмні продукти для проведення факторного аналізу мають перевагу у вирішенні як подібних задач, так і задач іншого рівня. Та ми не ставимо мету заміни цих продуктів, а лише вирішуємо проблему доступності обчислень за методом головних компонент.

Орієнтуючись на читача, який можливо детально не знайомий з положеннями факторного аналізу, а також, виходячи з того, що спосіб обчислень за нашою методою тісно зв'язаний з теорією, дозволимо собі її короткий виклад. Пояснення теорії будемо вести короткими розділами, в кінці яких будемо робити підсумки.

**Параметри статистичного дослідження групи об'єктів.**

Як правило, у статистичному дослідженні розглядають групу об'єктів, що мають певні спільні для них ознаки. Кількісні оцінки цих ознак називають значеннями їх параметрів. Таким чином кожний об'єкт характеризують набором певної кількості значень параметрів $z_j$. Така характеристика об'єкта являє собою багатовимірну випадкову величину з компонентами $z_{ij}$.

Вимірювання $n$ параметрів для $N$ об'єктів можна записати у вигляді таблиці або матриці даних

| № | $X_1$ | $X_2$ | … | $X_j$ | … | $X_n$ |
|---|---|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{12}$ | … | $x_{1j}$ | … | $x_{1n}$ |
| 2 | $x_{21}$ | $x_{22}$ | … | $x_{2j}$ | … | $x_{2n}$ |
| … | … | … | … | … | … | … |
| $i$ | $x_{i1}$ | $x_{i2}$ | … | $x_{ij}$ | … | $x_{in}$ |
| … | … | … | … | … | … | … |
| $N$ | $x_{N1}$ | $x_{N2}$ | … | $x_{Nj}$ | … | $x_{Nn}$ |

Обчислюють:

- Середнє значення кожного параметра.

- Відхилення значення параметра від середнього.

- Вибіркову дисперсію параметра.

- Нормоване значення параметра $j$ для об'єкту $i$.

- Коваріацію для будь яких довільних параметрів $j$ та $k$.

- Коефіцієнт кореляції для будь яких довільних параметрів $j$ та $k$.

Зміст цього розділу коротко можна формулювати таким чином. Масив досліджуваних величин потрібно нормалізувати, при цьому математичне очікування нормалізованого параметра $z_j$ – $M(z_j)$=0, а дисперсія – $D(z_j)$=1. Таким чином, першу процедуру яку ми виконуємо, це *нормалізація вихідних даних.*

**Лінійна модель класичного факторного аналізу.**

Завдання факторного аналізу полягає у тому щоби виразити кожний нормалізований параметр $z_j$ у термінах прихованих гіпотетичних факторів. Ці фактори розуміють, як деяку нову систему координат із відповідними масштабами, а $F_1$, $F_2$,…,$F_m$ модулі координатних векторів цієї системи. Найпростішою та вживаною є лінійна модель класичного факторного аналізу

$$z_j = a_{1j}F_1 + a_{2j}F_2 + ... + a_{jm}F_m + d_jU_j \qquad (1)$$

У цій моделі кожний параметр $z_j$ лінійно залежить від $m$ спільних факторів $F_1, F_2,\ldots,F_m$ $(m<<n)$ та одного характерного фактора $U_j$, який відображає ту частину параметра $z_j$, яку не можливо пояснити спільними факторами. Спільні фактори вважають змінними, а $a_{1j}, a_{2j},\ldots,a_{mj}$ – коефіцієнтами при змінних $F_1, F_2,\ldots,F_m$. Те саме стосується характерного фактора $U_j$ та коефіцієнта $d_j$. Лінійна модель класичного факторного аналізу схожа на рівняння регресії, але на відміну від регресії нам невідомо, як вимірювати ці змінні.

На фактори моделі накладають певні вимоги. Фактори мають бути величинами із нормованим нормальним розподілом, а, отже, мати математичне очікування рівне нулю та дисперсію рівну одиниці. Крім того, як вектори нової системи координат, фактори мають бути ортогональними, а, отже, варіації між різними факторами мають бути рівними нулю.

Коефіцієнти при факторах $a_{1j}, a_{2j},\ldots,a_{mj}$ називають *навантаженнями.* Навантаження визначають долю впливу окремого фактора на даний параметр, а їх значення є предметом обчислень у факторному аналізі.

Для $j$ параметра та $i$ об'єкту відповідна лінійна модель набуває вигляду

$$z_{ij} = a_{1j}F_{1i} + a_{2j}F_{2i} + ... + a_{mj}F_{mi} + d_jU_{ij} = \sum_{p=1}^{m} a_{pj}F_{pi} + d_jU_{ij} \qquad \rightarrow \qquad (2)$$

Основною проблемою факторного аналізу є визначення $m \cdot n$ навантажень спільних факторів. Така задача у решті-решт розв'язується процедурою аналізу матриці коефіцієнтів кореляції між параметрами. При такому аналізі проблема статистичної достовірності коефіцієнтів кореляції, як правило, ігнорується.

**Складові дисперсії лінійної моделі класичного факторного аналізу.**

Відповідно лінійної моделі стандартна дисперсія (нормованої величини) параметру $z_j$ може бути визначена як

$$s_{zj}^2 = \frac{\sum_{i=1}^{N} z_{ij}^2}{N} = \sum_{i=1}^{N} \frac{\left(\sum_{p=1}^{m} a_{pj}F_{pi} + d_jU_{ij}\right)^2}{N} = \sum_{p=1}^{m} a_{pj}^2 \left(\frac{\sum_{i=1}^{N} F_{pi}^2}{N}\right) + d_j^2 \frac{\sum_{i=1}^{N} U_{ij}^2}{N} + 2\sum_{p<q=1}^{m} a_{pj}a_{qj}\left(\frac{\sum_{i=1}^{N} F_{ip}F_{iq}}{N}\right) + 2d_j\sum_{p=1}^{m} a_{pj}\left(\frac{\sum_{i=1}^{N} F_{ip}U_{ij}}{N}\right)$$

Вираз може бути суттєво спрощений завдячуючи двом обставинам.

Перша, стандартна дисперсія дорівнює одиниці, а всі інші величини виразу також записані у стандартній формі. Внаслідок цього маємо

$$s_{zj}^2 = 1 = \sum_{p=1}^{m} a_{pj}^2 + d_j^2 + 2\sum_{p<q=1}^{m} a_{pj}a_{qj}r_{FpFq} + 2d_j\sum_{p=1}^{m} a_{pj}r_{FpUj}$$

Друга, приймається припущення, що всі фактори незалежні один від одного, а отже не корелюють між собою. Отже у решті-решт маємо

$$s_{zj}^2 = 1 = \sum_{p=1}^{m} a_{pj}^2 + d_j^2 = a_{1j}^2 + a_{2j}^2 + ... + a_{mj}^2 + d_j^2 \rightarrow (3)$$

Очевидно, що квадрати факторних навантажень виражають вклади відповідних факторів у дисперсію певного параметру. Суму

$$h_j^2 = \sum_{p=1}^{m} a_{pj}^2 = a_{1j}^2 + a_{2j}^2 + ... + a_{mj}^2$$

називають спільністю $d_j^2$. Величину $d_j^2$ називають характерністю.

Сформулюємо підсумок розділу. Дисперсія параметра $z_j$ складається із двох частин. Перша частина пояснюється впливом на параметр факторів, а друга частина не може бути пояснена впливом факторів, вона існує сама по собі та індивідуальна для кожного окремого параметра.

**Факторне відображення та факторна структура.**

Користуючись лінійною моделлю класичного факторного аналізу (1), яка зв'язує параметр $z_j$ із $m$ спільними факторами,запишемо вирази для усіх $n$ параметрів у вигляді:

$$\begin{cases} z_1 = a_{11}F_1 + a_{21}F_2 + ... + a_{m1}F_m + d_1U_1 \\ z_j = a_{12}F_1 + a_{22}F_2 + ... + a_{m2}F_m + d_2U_2 \\ ... \ ... \ ... \ ... \ ... \ ... \ ... \ ... \ ... \ ... \ ... \ ... \ ... \\ z_n = a_{1n}F_1 + a_{2n}F_2 + ... + a_{mn}F_m + d_nU_n \end{cases}$$

Така система рівнянь називається факторним відображенням або просто відображенням. Її можна записати у матричній формі

$$Z = A \cdot F + D$$

$$Z = \begin{vmatrix} z_1 \\ z_2 \\ ... \\ z_n \end{vmatrix};\ A = \begin{vmatrix} a_{11} & a_{12} & ... & a_{1m} \\ a_{21} & a_{22} & ... & a_{2m} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nm} \end{vmatrix};\ F = \begin{vmatrix} F_1 \\ F_2 \\ \\ F_m \end{vmatrix};\ D = \begin{vmatrix} d_1 & 0 & ... & 0 \\ 0 & d_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & d_n \end{vmatrix}$$

Окрім факторного відображення у факторному аналізі розглядають факторну структуру, необхідну для подальшої побудови теорії. Для знаходження факторної структури достатньо помножити кожне з рівнянь (2) на відповідні фактори, додати по усім N спостереженням для кожного параметра, розділити на кількість значень N і отримати коефіцієнти кореляції параметрів з факторами. Ця процедура враховує, що параметри та фактори – нормалізовані величини. А якщо врахувати також відсутність кореляції між факторами, то виявиться , що відповідні коефіцієнти кореляції параметрів та факторів дорівнюють навантаженням (елементам матриці А).

$$r_{Zj;Fp} = a_{jp}$$

Підсумок розділу. Елементи матриці $A$ – це коефіцієнти кореляції між параметрами та факторами. Матрицю $A$ називають факторною структурою.

**Фундаментальна факторна теорема.**

На підставі рівнянь (2), які дають значення параметрів, можна обчислити коефіцієнти кореляції між ними, як добуток правих частин рівнянь, враховуючи нормованість та некорельованість факторів між собою. Нехтуючи деталями отримаємо вирази виду:

$$r_{ij} = a_{1i}a_{1j} + a_{2i}a_{2j} + ... + a_{mi}a_{mj}, \ \ (i, j = 1,2,...,n)$$

Перейдемо до матричної форми цього результату

$$Z = A \cdot F \Rightarrow R = \frac{Z \cdot Z'}{N} \Rightarrow R = \frac{1}{N}(A \cdot F \cdot F' \cdot A') \Rightarrow R = A \cdot A'$$

Оскільки, фактори нормовані та некорельовані між собою, то:

$$\frac{F \cdot F'}{N} = I \ \Rightarrow I = \begin{vmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 1 \end{vmatrix} \Rightarrow A \cdot I = A$$

Таким чином, підсумком розділу є співвідношення $R = A \cdot A'$ яке виражає фундаментальну факторну теорему. Знайдене співвідношення зв'язує між величини, які доступні для вимірювання $R$ і величини $A \cdot A'$, які безпосередньо не вимірюються.

**Метод головних факторів.**

Основна задача факторного аналізу знаходження навантажень $a_{ij}$ факторів. Існує багато методів [1] розв'язку цієї задачі (деякі з них досить архаїчні). Ми скористаємося методом головних факторів розробленим Г. Томсоном.

Вище, виведена формула (3)

$$s_{zj}^2 = 1 = \sum_{p=1}^{m} a_{pj}^2 + d_j^2 = a_{1j}^2 + a_{2j}^2 + ... + a_{mj}^2) + d_j^2$$

яка показує структуру дисперсії показника, що має дві складові: $h_j^2 = \sum_{p=1}^{m} a_{pj}^2 = a_{1j}^2 + a_{2j}^2 + ... + a_{mj}^2$ та $d_j^2$. Перша, спільність $h_j^2$ - показує яка частина дисперсії показника пояснюється факторами, а друга, характерність $d_j^2$ - показує яку частину дисперсії показника не можна пояснити факторами. Чим більша перша частина та менша друга, тим є кращим результат факторного аналізу. Отже умова, за якою послідовно визначають фактори,

вимагає максимуму вкладів кожного з факторів у сумарну спільність.

Для певного фактора $F_j$ ця сума:

$$V_j = \sum_{i=1}^{n} a_{ij}^2 = a_{1j}^2 + a_{2j}^2 + \ldots + a_{nj}^2$$

На першій стадії обчислень шукають навантаження при першому факторі таким чином,щоб була максимальною сума вкладів цього фактора у сумарну спільність, при цьому має бути забезпечена умова, що $r_{jk} = \sum_{p=1}^{m} a_{jk} \cdot a_{kp}$

Це задача на умовний екстремум функції багатьох змінних, яку розв'язують методом множників Лагранжа. Так, для першого фактора маємо:

$2L = \sum_{i=1}^{n} a_{ij}^2 - \sum_{i=1}^{n} \lambda_{jk} r_{jk} = \sum_{i=1}^{n} a_{ij}^2 - \sum_{i=1}^{n} \sum_{p=1}^{m} \lambda_{jk} a_{jk} a_{kh}$ - функція Лагранжа,

$\frac{\partial L}{\partial a_{i1}} = a_{i1} - \sum_{i=1}^{n} \lambda_{1i} a_{1i} = 0$ - частинні похідні функції Лагранжа рівні нулю.

Знаходження умовного екстремуму функції Лагранжа краще вести у матричному форматі і тоді об'єднана система рівнянь із частинним похідними набуде виду характеристичного рівняння:

$$\begin{vmatrix} h_1^2 - \lambda & r_{12} & \ldots & r_{12} \\ r_{21} & h_2^2 - \lambda & \ldots & r_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ r_{m1} & r_{m2} & \ldots & h_m^2 - \lambda \end{vmatrix} = 0 \Rightarrow R \cdot \vec{a} = \lambda \cdot \vec{a}$$

Отже, задача зводиться до знаходження власних векторів $\vec{a}$ та власних значень, так званої, редукованої кореляційної матриці $R$. У редукованої кореляційної матриці на головній діагоналі розміщені значення спільності $h_j$ замість одиниць звичайної кореляційної матриці. Власні вектори $\vec{a}$ - матриця стовбець спільностей відповідного фактора. Власні значення $\lambda$ (коефіцієнти Лагранжа), - сума дисперсій параметрів або значення функції $V_j$, яка підлягала максимізації.

У підсумку цього розділу зауважимо наступне. Нарешті знайдене матричне рівняння $R \cdot \vec{a} = \lambda \cdot \vec{a}$, що зв'язує кореляційну матрицю показників (реальних величин) із величинами (прихованими) та дозволяє їх визначати.

Саме з цим рівнянням пов'язана практична проблема визначення способу розв'язку рівняння з метою обчислень недоступних до вимірюванню факторних параметрів. Також маємо задачу, як знайти реальну редуковану кореляційну матрицю. Основною проблемою тут – є визначення спільностей $h_j$.

**Схема обчислень по методу головних факторів.**

Порядок обчислень наступний:

- Для знаходження параметрів першого головного компонента $F_1$ розв'язуємо рівняння $R \cdot \vec{a}_1 = \lambda_1 \cdot \vec{a}_1$, що максимізує суму

$$V_1 = \sum_{i=1}^{n} a_{i1}^2 = a_{11}^2 + a_{21}^2 + \ldots + a_{n1}^2$$

знаходимо перший корінь характеристичного рівняння $\lambda_1$ та один із можливих розв'язків вектор $\vec{b_1}$ =($b_{11}$, $b_{21}$,…, $b_{n1}$). Щоб задовольнити співвідношення для $V_1$ обчислюємо $a_{i1}$:

$$a_{i1} = \frac{b_{i1}\sqrt{\lambda_1}}{\sqrt{b_{11}^2 + b_{21}^2 + \ldots + b_{n1}^2}} \Rightarrow (4)$$

і отримаємо навантаження для першого фактора, яке запишемо у вигляді матриці-стовпця $\vec{a}_1$.

- Щоб знайти розв'язок для другого за величиною компонента $F_2$ потрібно вирахувати із редукованої кореляційної матриці внесок від компонента $F_1$. Для цього знаходимо добуток матриці $\vec{a}_1$ на транспоновану матрицю $\vec{a}_1'$ та отримаємо матрицю $\widetilde{R} = \vec{a} \cdot \vec{a}'$, після чого знайдемо наступну редуковану матрицю $R_1$:

$R_1 \cdot \vec{a}_2 = \lambda_2 \cdot \vec{a}_2$

Далі процедура знаходження навантажень та наступного кореню характеристичного рівняння повторюється. Розв'язуємо рівняння $R_1 \cdot \vec{a}_2 = \lambda_2 \cdot \vec{a}_2$, що максимізує суму $V_2 == a_{12}^2 + a_{22}^2 + \ldots + a_{n2}^2$, знаходимо другий корінь характеристичного рівняння $\lambda_2$ та матрицю-стовпець $\vec{a}_2$.

- Оскільки стандартна дисперсія окремого нормованого параметра дорівнює одиниці, то знаходження наступних компонент припиняють коли $\lambda_j$ стає менше 1.

Отже, схема обчислень визначена - лишається доповнити схему алгоритмом обчислень.

**Формування редукованої матриці.**

Виберемо масив даних для проведення факторного аналізу. Наприклад, такий як у роботі [4]. Це оцінки у стобальній шкалі сесійного контролю навчання 28 студентів по 11 предметах.

Відповідно цьому, оцінки розміщені у матриці розміром 28x11 (28 рядків та 11 стовпців). Для дослідження результатів навчання студентів по окремим дисциплінам можна побудувати матрицю кореляцій оцінок між студентами. Це буде

квадратна кореляційна матриця порядку 28. Тоді ми можемо шукати фактори, що впливають на якість навчання студентів. Якщо ми хочемо порівняти якість викладання дисциплін, то можна побудувати матрицю кореляцій оцінок студентів по предметам. Така матриця буде порядку 11.

Приклад ми навели лише для того, щоб показати, що перший етап роботи це здобуття реальних даних. Далі ми говоритимемо лише про математичні процедури.

- Вибираємо масив даних, нормуємо їх та записуємо прямокутну матрицю $\boldsymbol{X}$ з $\boldsymbol{m}$ параметрами та з $\boldsymbol{n}$ вимірюваннями кожної.

- Транспонуємо матрицю $\boldsymbol{X}$ у $\boldsymbol{X'}$, та обчислюємо кореляційну матрицю $\boldsymbol{R}$ як добуток $\boldsymbol{R=X'X}$. Діагональ цієї матриці складається з одиниць. Тут виникає проблема визначення спільності $\boldsymbol{h_j^2}$, значеннями якої потрібно замінити одиниці діагоналі звичайної кореляційної матриці. Як пише Г.Харман [1], існує декілька методів оцінки спільності і ні один не є надійним. Але при великих розмірах кореляційної матриці похибка в обчисленнях, пов'язаних із цією обставиною, не є суттєвою. Чим більша матриця - тим менша похибка результатів аналізу.

- Нас влаштовує оцінка спільності $\boldsymbol{h_j^2}$ через обчислення коефіцієнта множинної кореляції $\boldsymbol{r_j}$ (КМК) кожного окремого параметра з усіма іншими. У цьому є логіка, бо $\boldsymbol{r_j}$ є мірою того, що спільного у параметра $\boldsymbol{j}$ з усіма іншими параметрами (у літературі вважають, що КМК – лише нижня оцінка спільності $\boldsymbol{h_j^2}$).

- Отже, знаходимо обернену матрицю $\boldsymbol{C=(X'X)^{-1}}$, вибираємо діагональний елемент $\boldsymbol{c_{jj}}$ та обчислюємо КМК як $\boldsymbol{r_j}$=1-1/$\boldsymbol{c_{jj}}$.

- Замінюємо діагональні одиниці кореляційної матриці $\boldsymbol{R}$ коефіцієнтом множинної кореляції $\boldsymbol{r_j}$ та отримаємо редуковану кореляційну матрицю, для якої збережемо те саме позначення $\boldsymbol{R}$.

Таким чином маємо редуковану кореляційну матрицю $\boldsymbol{R}$, з якої починаємо процедуру обчислень.

**Обчислення головних факторів.**

Подальша процедура починається з виділення першого фактора, яка полягає у послідовності ітераційних обчислень. Ітерації припиняємо, коли результат обчислень перестає мінятися у межах допустимих похибок (нехай буде 5-6 знак після коми). Основу обчислень становить співвідношення: $R\cdot\vec{a}_1=\lambda_1\cdot\vec{a}_1$ . Після виділення чергового фактора ітераційні обчислення повторюються для наступного.

Опишемо процедуру обчислень.

- Знаходимо добуток матриці $\boldsymbol{R}$ на матрицю стовпчик $\vec{a}_1$. У якості першого наближення стовпчика $\vec{a}_1$ виберемо матрицю, утворену сумою рядочків матриці $\boldsymbol{R}$. В цьому є певна логіка, бо результатом дії буде матриця стовпчик: $\lambda_1\cdot\vec{a}_1$, а максимальний елемент цієї матриці буде близький до факторного коефіцієнта $\lambda_1$ (власного значення).

- Вибираємо найбільший елемент стовпчика $\lambda_1\cdot\vec{a}_1$ та ділимо на нього всі інші елементи. Отримаємо перше наближення власного вектора $\vec{a}_1$.

- Знову знаходимо добуток матриці $\boldsymbol{R}$ на наближення власного вектора $\vec{a}_1$. Повторюємо подальшу процедуру, та обчислюємо друге наближення $\vec{a}_1$. Після деякої кількості ітерацій результат дії перестає змінюватися.

- Формулюємо результат виявлення першої головної компоненти факторного аналізу. Найбільший елемент отриманого стовпчика $\lambda_1\cdot\vec{a}_1$ дорівнює факторній вазі першого фактора $\lambda_1$ Навантаження першого фактора обчислимо за формулою (4) використавши стовпчик $\boldsymbol{b_1}$ з результатами ділення елементів $\lambda_1\cdot\vec{a}_1$ на $\lambda_1$.

- Для виявлення другого фактора $\boldsymbol{F_2}$ із результатів вимірювання параметрів потрібно виключити інформацію зв'язану із фактором $\boldsymbol{F_1}$. Для цього матрицю-стовпчик навантажень $\vec{a'}_1$ першого фактора транспонуємо і отримаємо матрицю рядок $\vec{a'}_1$. Потім знаходимо матрицю $\widetilde{R}=\vec{a}_1\cdot\vec{a'}_1$ та віднімаємо її від редукованої матриці $\boldsymbol{R}$. Маємо $R_1=R-\widetilde{R}$ - наступну матрицю для якої повторюємо всю попередню процедуру. Знаходимо суму рядків $\boldsymbol{R_1}$, повторюємо ітерації та обчислення до знаходження $\boldsymbol{R_2}$.

- Розрахунки для наступних компонент продовжуємо поки не отримаємо $\lambda_{m+1}<1$, тоді останнім компонентом аналізу буде фактор номер $\boldsymbol{m}$.

- Результати обчислень можна оформити у вигляді таблиці.

**Висновки.**

Описана процедура визначення головних компонентів у факторному аналізі кореляційної матриці засобами «EXCEL».

Надано пояснення цієї процедури з позицій теоретичних положень факторного аналізу.

*Список використаних джерел:*

1. Харман Г. Современный факторный анализ. М.: Статистика, 1972. – 486 с.

2. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. – М.: Изд-во Прогресс, 1976. – 496 с.

3. Свиридов А. П. Основы статистической теории обучения и контроля знаний. – М.: Высш. школа, 1981. – 262 с.

4. Касярум О.П Факторний аналіз результатів сесійного контролю рівня знань студентів [Текст]/ О.П. Касярум, С.О. Касярум // Вісник УБС НБУ. – 2008.- №2. – С. 164-167.

5. Касярум О.П Моніторинг навчального процесу за результатами сесійного контролю рівня знань студентів [Текст]/ О.П. Касярум, С.О. Касярум // Вісник Черкаського національного університету. – 2007. – В.98. – С. 44-50.

6. Данко П.Е., Попов А.Г., Кожевникова Т.Я. Высшая математика в упражнениях и задачах. В 2-х частях Ч.1: Учеб. Пособие для втузов.- 5-е изд., испр.-М.:Высш. шк., 1997.-304 с.