

ФАКТОРНИЙ АНАЛІЗ НЕМЕТРИЧНИХ ДАНИХ: ЕВРИСТИЧНИЙ ПОТЕНЦІАЛ КАТЕГОРІАЛЬНОГО АНАЛІЗУ ГОЛОВНИХ КОМПОНЕНТ

Судин Данило Юрійович – кандидат соціологічних наук, асистент кафедри історії та теорії соціології Львівського національного університету імені Івана Франка

У статті розглянуто недоліки факторного аналізу при аналізі неметричних даних, виміряних з допомогою порядкових шкал. Виділено три перешкоди, що роблять статистично некоректним використання факторного аналізу при опрацюванні соціологічних даних: 1) неможливість геометричної інтерпретації матриці даних у випадку неметричних даних; 2) необхідність z-стандартизації даних як умова проведення факторного аналізу, яка неможлива для порядкових змінних; 3) розрахунок частин дисперсії, що пояснюються отриманими факторами, як умова виділення оптимального числа факторів. Розглянуто загальні принципи категоріального аналізу головних компонент, завдяки яким вдається подолати згадані недоліки факторного аналізу, зберігаючи його переваги: квантифікація порядкових змінних та обчислення функції втрат при виділенні факторів. Проведено емпіричне порівняння евристичного потенціалу двох методів опрацювання соціологічних даних на результатах соціологічного опитування.

Ключові слова: категоріальний аналіз головних компонент, факторний аналіз.

В статье рассмотрены недостатки факторного анализа при анализе неметрических данных, измеренных с помощью порядковых шкал. Выделено на три препятствия, которые делают статистически некорректным использование факторного анализа при обработке социологических данных: 1) невозможность геометрической интерпретации матрицы данных в случае неметрических данных; 2) необходимость z-стандартизации данных как условие проведения факторного анализа, которая невозможна для порядковых переменных; 3) расчет частей дисперсии, которые объясняются полученными факторами, как условие выделения оптимального числа факторов. Рассмотрены общие принципы категориального анализа главных компонент, благодаря которым удается преодолеть недостатки факторного анализа, сохраняя его преимущества: квантификация порядковых переменных и вычисление функции потерь при выделении факторов. Проведено эмпирическое сравнение эвристического потенциала двух методов обработки социологических данных на результатах социологического опроса.

Ключевые слова: категориальный анализ главных компонент, факторный анализ.

The article offers analysis of factor analysis shortcomings in the analysis of categorical data (measured with ordinal scales). Three obstacles that make statistically improper use of factor analysis when processing sociological data: 1) inability of geometric interpretation of the data matrix in the case of categorical data; 2) the need for data standardization as a condition of factor analysis, which is impossible for ordinal variables, and 3) calculation of dispersion parts explained by obtained factors as the condition for selection of the optimal number of factors. The general principles of categorical principal components analysis that improve mentioned shortcomings of factor analysis, while maintaining its advantages are analyzed: quantification of ordinal variables and calculating the loss function in the obtaining factors. An empirical comparison of heuristic potential of the two methods of processing sociological data on the results of sociological survey is offered.

Keywords: categorical principal components analysis, factor analysis.

Попри те, що статистичне обґрунтування надійності факторного аналізу при аналізі соціологічних даних відбулося в 30-і рр. ХХ ст., сам метод належить до “недавно відкритих” в пострадянській соціології: ще в 2004 р. В. Паніотто, В. Максименко та Н. Харченко зазначали, що згаданий метод опрацювання та інтерпретації соціологічних даних не отримав значного поширення серед українських соціологів [1, с. 6]. Аналогічна ситуація і в російській соціології, де в 2005 р. О. Орлов вказував факторний аналіз як новий метод аналізу соціологічних даних [2, с. 44–47]. Перевагою цього методу є те, що він дозволяє скоротити кількість змінних, тобто звести їх до кількох факторів, що в подальшому спрощує дослідникові завдання інтерпретації результатів дослідження. Парадокс ситуації полягає в тому, що зараз в західній соціології факторний аналіз піддається критиці як такий, що не може бути застосований до соціологічних даних, оскільки останні є переважно неметричними (виміряними з допомогою порядкових та номінальних шкал),

адже факторний аналіз було розроблено для аналізу метричних даних (вимірних з допомогою інтервальних шкал та шкал відношень), які в соціології майже не використовуються. Відтак, основною дослідницькою проблемою, на вирішення якої спрямована стаття – виявлення евристичного потенціалу методу категоріального аналізу головних компонент як заміни факторному аналізу. Зазначимо, що існує певний дисбаланс у інтересі до цього методу аналізу емпіричних даних: нам вдалося знайти лише дві статті, в яких при аналізі емпіричних даних було використано категоріальний аналіз головних компонент [3], [4], натомість в англійській соціологічній літературі існує низка публікацій, в яких розкривається зміст та призначення згаданого методу аналізу (див. наприклад [5], [6], [7], [8]). Але навіть в них не здійснено порівняння факторного аналізу та категоріального аналізу головних компонент. А тому в пропонованій статті буде здійснено спробу розглянути в порівнянні з факторним аналізом переваги та недоліки категоріального аналізу головних компонент у випадку аналізу неметричних даних¹.

Суть факторного аналізу полягає в зменшенні числа змінних, тобто представлення більшої кількості змінних у формі лінійної комбінації меншої кількості змінних [9, с. 126]. Проте таке трактування – попри те, що точно передає суть факторного аналізу – потребує уточнення, яке краще допоможе зрозуміти недоліки цього методу у випадку його використання щодо соціологічних даних, які переважно є неметричними. З точки зору математичної статистики суть факторного аналізу полягає у проектуванні матриці значень n -вимірного простору у k -вимірний, причому $k < n$, тобто простір, у який проектується матриця значень має меншу розмірність, аніж той, в якому знаходиться сама матриця². В основі такої інтерпретації факторного аналізу лежить представлення результатів опитування як матриці значень $A^{(n \times N)}$, в якій n – кількість змінних, а N – кількість опитаних. Відтак, результати опитування можна представити як комбінацію N точок у n -вимірному просторі. Звідси випливає, що суть факторного аналізу полягає у скороченні числа вимірів, тобто проектуванні матриці значень у простір меншої розмірності [10, с. 563]. Через це факторний аналіз виявляється непридатним для аналізу неметричних даних, адже у випадку порядкових змінних ми маємо справу з монотонно зростаючими функціями, а тому арифметичні різниці між варіантами відповідей не можуть бути геометрично інтерпретовані, тобто не можуть позначати відстань у просторі і слугувати координатами точок (респондентів) у просторі [11, с. 63]. Загалом, проблема полягає в тому, що простір, який задається значеннями порядкових змінних, не є лінійним.

Згаданий підставовий недолік факторного аналізу при опрацюванні неметричних даних доповнюється низкою більш конкретних. Для їхнього аналізу слід розглянути основні етапи факторного аналізу. Загалом, їх виділяють три: 1) підготовка матриці кореляцій між змінними; 2) виділення факторів; 3) обертання матриці факторних навантажень [11, с. 9].

Оскільки в наше завдання не входить розгляд методики факторного аналізу, то ми зосередимося лише на тих його компонентах, які роблять проблематичним застосування цієї статистичної процедури при опрацюванні соціологічних даних³.

Матриця кореляцій між змінними. Перш ніж розраховувати кореляції між змінними, дослідник повинен стандартизувати значення самих змінних [13, с. 58], [9, с. 126]. Формула для z -стандартизації є наступною: $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$, де $j = 1, n$. Проте z -стандартизація не має смислу у випадку порядкових змінних, адже для них некоректно вираховувати стандартне відхилення (s), оскільки останнє передбачає рівність інтервалів (у геометричній інтерпретації – відстаней) між варіантами відповідей. Цей факт становить першу перешкоду при використанні факторного аналізу щодо неметричних даних.

Стандартизація значень змінних є необхідною умовою проведення факторного аналізу, а тому для полегшення сприйняття подальших міркувань вважаємо за доцільне нагадати, що після проведення z -стандартизації середнє кожної змінної дорівнює 0, а стандартне відхилення – 1 (тобто $\sum_{i=1}^N x_i = 0$, $s_i = 1$). При проведенні факторного аналізу розраховуються коефіцієнти кореляції Пірсона-Браве: $r = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}$. Для z -стандартизованих змінних вона набуває вигляду: $r = \frac{\sum_{i=1}^N x_i \cdot y_i}{\sqrt{\sum_{i=1}^N x_i^2 \cdot \sum_{i=1}^N y_i^2}}$. Проте розрахунок коефіцієнта кореляції Пірсона-Браве для неметричних даних є некоректним [11, с. 99], що становить другу перешкоду при використанні факторного аналізу для неметричних даних⁴.

¹ Ми зосередимося на одному з різновидів неметричних даних – порядкових змінних, проте категоріальний аналіз головних компонент може бути застосований і щодо номінальних змінних [6, р. 54].

² Завдання ускладнюється тим, що k є невідомим та повинне бути визначеним в ході факторного аналізу, про що йтиметься нижче.

³ Детальний виклад процедури факторного аналізу міститься в працях [11], [12].

⁴ Додамо, що в цьому випадку виникає ще одна проблема, яка міститься в самому формулюванні суті факторного аналізу: зв'язок між змінними повинен бути лінійним. А відтак, у випадку нелінійних залежностей слід розраховувати кореляційне відношення η .

Таблиця 1.

Матриця коефіцієнтів кореляції між змінними

	X_1	X_2	X_3	X_4
X_1	r_{11}	r_{12}	r_{13}	r_{14}
X_2		r_{22}	r_{23}	r_{24}
X_3			r_{33}	r_{34}
X_4				r_{44}

Виділення факторів. В розрахованій таким чином матриці кореляцій є ще одна особливість: її елементи, які знаходяться на головній діагоналі, є ще коефіцієнтами кореляції, а дисперсіями згаданих змінних. Тобто $r_{jj} = s_{jj}^2$ (див. Табл. 1.1). Якщо $r = \frac{\sum_{i=1}^N x_i \cdot y_i}{\sqrt{\sum_{i=1}^N x_i^2 \cdot \sum_{i=1}^N y_i^2}}$, то у випадку діагональних елементів маємо $r_{jj} = \frac{\sum_{i=1}^N (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^N (x_{ij} - \bar{x}_j)^2 \cdot \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2}}$. Якщо врахувати, що $\bar{x}_j = 0$, то згаданий коефіцієнт кореляції дорівнює дисперсії: $r_{jj} = \frac{\sum_{i=1}^N x_{ij}^2}{N} = s_{jj}^2$. А тому $\sum r_{jj} = n$, що також дорівнює загальній варіації всіх змінних [12, с. 152], [14, № 190]. Згаданий показник є важливим при визначенні оптимального числа факторів у випадку проведення факторного аналізу методом головних компонент: фактор повинен пояснювати частку варіації змінних, яка є не меншою 1 (відповідно до критерію Кайзера) [15, с. 103–107]⁵.

Відповідно, запропонований деякими авторами спосіб використання факторного аналізу для неметричних даних (порядкових змінних), розраховуючи матрицю коефіцієнтів кореляції ρ Спірмена або τ Кендалла замість r Пірсона-Брава [15, с. 298], не мають під собою статистичного підґрунтя, адже в отриманій таким чином матриці коефіцієнтів кореляції значення діагональних елементів не дорівнюватимуть дисперсії значень відповідних змінних, тим більше, що для порядкових змінних, як уже згадувалося вище, розрахунок дисперсії не є коректним. А відтак, навіть за умови використання матриці рангових кореляцій, доводиться використовувати дисперсії згаданих змінних (див. [16]). Саме через це автори праці “Факторний аналіз” Дж. Кім та Ч. Мюллер зазначають, що такий спосіб використання факторного аналізу є некоректним. Тим більше, що в порядкових змінних числові значення не можуть бути інтерпретовані як відстані, що – як ми уже зазначали вище – унеможливує проведення факторного аналізу [11, с. 63].

На основі наведених вище міркувань можемо стверджувати, що перед соціологом постає проблема використання факторного аналізу. З одного боку, його застосування не є коректним до того типу даних, якими оперує соціолог, а з іншого – факторний аналіз є надто потужним статистичним інструментом, щоб від нього відмовлятися. Тому деякі автори зазначають, що факторний аналіз можна використовувати і до неметричних даних, але дослідникові слід пам’ятати, що в такому випадку отримані ним результати не цілком відображатимуть реальний стан справ. Для подолання цієї проблеми було розроблено метод категоріального аналізу головних компонент. Оскільки метою нашої статті є порівняння евристичних можливостей останнього способу зменшення розмірності отриманих соціологічних даних з евристичним потенціалом факторного аналізу, то ми не зосереджуватимемося на аналізі процедури та змісту категоріального аналізу головних компонент, а лише вкажемо на відмінності цього виду опрацювання соціологічних даних від факторного аналізу⁶.

У випадку використання факторного аналізу для аналізу неметричних даних, як ми вже зазначали вище, дослідник стикається з двома проблемами: проектування матриці даних у просторі, що є неможливим для порядкових змінних, а також розрахунок частки варіації, яка пояснюється при проектуванні матриці значень з n -вимірного простору в k -вимірний. Першу проблему в категоріальному аналізі головних компонент вирішено з допомогою використання індикаторних матриць та розрахунку ваг для різних значень порядкових змінних (шляхом розрахунку рівняння регресії). Порядкову змінну h_m можна записати у вигляді результату скалярного множення індикаторної матриці G_m справа на вектор y_m . Матриця G_m складається з N рядків, які відповідають кількості респондентів, та C_m колонок, які відповідають кількості варіантів відповідей для змінної h_m . Відповідно, $G_m = \{g_{im}\}$, $i = 1, N$, причому $g_{im} = \begin{cases} 1, & \text{якщо } i \text{ відповідає } h_m \\ 0, & \text{інакше} \end{cases}$, де c_m – числове значення, яких може набувати згадувана порядкова змінна. А відтак порядкову змінну h_m ми можемо записати як $h_m = G_m \cdot y_m$, де y_m – вектор, що складається зі всіх можливих значень, яких може набувати змінна h_m , тобто з числових відповідників варіантів відповідей на запитання, якому відповідає згадана змінна. Крім того, розраховуються вагові коефіцієнти для кожного з варіантів відповідей на певне запитання (на основі побудови рівняння регресії), що дозволяє квантифікувати порядкові змінні. Адже числові значення, яких може набувати порядкова змінна, описуються монотонно зростаючою функцією, тобто важливим є лише порядок числових значень, а не самі значення. Операція квантифікації дозволяє зберегти первинний порядок значень, але також обчислити “відстані” між ними. Варто наголосити, що така квантифікація здійснюється окремо стосовно

⁵ Саме тому z -стандартизація змінних є необхідною умовою факторного аналізу: лише в такому випадку значення діагональних елементів матриці кореляцій будуть дорівнювати значенням дисперсії згаданих змінних.

⁶ Детальний виклад основ категоріального аналізу головних компонент міститься в таких працях [5], [6].

кожної змінної, яка опрацьовується в категоріальному аналізі головних компонент, а тому має релятивний характер, який визначається структурою відповідей респондентів. Здійснення квантифікації неметричних даних дозволяє розмістити їх в лінійному просторі, тобто виконати основне завдання факторного аналізу – проектування даних з n -вимірного простору у k -вимірний (де $k < n$)⁷.

Іншою проблемою факторного аналізу є розрахунок частки варіації, яка проектується на той чи інший вимір. Згадану проблему в категоріальному аналізі головних компонент вирішено з допомогою обчислення функції втрат (loss function), яку можна сприймати як альтернативу методу найменших квадратів⁸. Загалом цей метод зберігає в собі всі переваги аналізу головних компонент, проте позбавлений його недоліків, які з'являються при аналізі неметричних даних.

В такий спосіб в категоріальному аналізі головних компонент обчислюється матриця факторних навантажень змінних стосовно кожного з факторів, статистична інтерпретація якої є аналогічна до інтерпретації матриці факторних навантажень, яку дослідник отримує внаслідок проведення факторного аналізу. Проте до недавнього часу недоліком категоріального аналізу головних компонент було нехтування задачею обертання згаданої матриці, хоча згадана процедура переважно є необхідною при смисловій інтерпретації отриманих факторів. Лише в 2010 р. розробники згаданого методу опрацювання неметричних даних визнали необхідність та можливість такого обертання. Причому важливо зазначити, що згадана задача вирішується тотожно до вирішення в факторному аналізі, адже обертання – наприклад, методом варімакс – стосується обертання виключно матриці факторних навантажень, незалежно від того, з допомогою якого варіанту факторного аналізу її було отримано⁹ [5, р. 104].

Для перевірки евристичної цінності категоріального аналізу головних компонент в соціологічних дослідженнях нами було здійснено порівняння результатів цього методу опрацювання даних та факторного аналізу¹⁰. Для цього ми використали результати омнібусу, який було організовано та проведено кафедрою історії та теорії соціології Львівського національного університету ім. Івана Франка у Львові та Львівській області в березні 2011 р. ($n=400$, вибіркова сукупність формувалася методом квот та є репрезентативною за критеріями віку та статі). Зокрема, ми використали відповіді на запитання, які були спрямовані на визначення етносоціальної дистанції стосовно низки етнічних та расових груп (всього 25 груп), яка вимірювалася з допомогою шкали соціальної дистанції Е. Богардуса.

Порівняння результатів аналізу ми здійснюватимемо на основі наступних критеріїв:

- частка варіації ознак, яка пояснюється отриманими факторами;
- кількість факторів та їхня статистична надійність;
- рівень наблизеності отриманої факторної структури до простої факторної структури;
- смислова інтерпретація отриманих факторів.

Внаслідок здійсненого аналізу, було отримано наступні дані: фактори, отримані внаслідок проведення факторного аналізу, пояснюють 59,4% варіації змінних, натомість фактори, отримані внаслідок проведення категоріального аналізу головних компонент, пояснюють 61,2% варіації змінних. Отже, у випадку пояснення варіації ознак метод категоріального аналізу головних компонент є більш точним, порівняно з факторним аналізом.

Порівняння двох методів аналізу даних за критерієм кількості факторів та їхньої надійності, дало наступні результати, представлені в Табл. 1.2 і Табл. 1.3.

Таблиця 2.

Значення коефіцієнту надійності α Кронбаха для факторів, отриманих за допомогою факторного аналізу

№ фактора	α Кронбаха
1	0,82
2	0,83
3	0,87
4	0,64
5	0,8

⁷ Фактично, операцію квантифікації можна розглядати як проектування значень з нелінійного простору на лінійний, причому розраховані рівняння регресії дозволяють виявити “кривизну” нелінійного простору, а тому і виконати згадану операцію проектування.

⁸ Більш детальний виклад обчислень та математичне доведення функції втрат міститься в [6, р. 49–55, 60–63].

⁹ Згадане уточнення є важливим, оскільки проведення факторного аналізу або категоріального аналізу головних компонент неможливе без використання обчислювальної техніки. Наскільки нам відомо, здійснення категоріального аналізу головних компонент можливе завдяки модулю SPSS, який має назву CatPCA, проте в ньому відсутня можливість обертання отриманої матриці факторних навантажень. Тому розробники цього методу пропонують здійснювати процедуру обертання з допомогою модулю факторного аналізу [5, р. 104].

¹⁰ Факторний аналіз було проведено методом головних компонент з нормалізацією Кайзера.

Таблиця 3.

Значення коефіцієнту надійності α Кронбаха для факторів,отриманих за допомогою категоріального аналізу головних компонент

№ фактора	α Кронбаха
1	0,85
2	0,87
3	0,81
4	0,8

З даних, представлених в Табл. 1.2 та Табл. 1.3, робимо висновок, що категоріальний аналіз головних компонент є більш статистично точним: з його допомогою було отримано меншу кількість факторів, кожен з яких має високий рівень надійності. Значення α Кронбаха для факторів, отриманих внаслідок категоріального аналізу головних компонент, не менші 0,8, натомість один з факторів, отриманий внаслідок факторного аналізу, має значення 0,64.

Для порівняння двох методів аналізу соціологічних даних за критеріями наближеності отриманої факторної структури до простої факторної структури та критерієм смислової інтерпретації отриманих факторів, звернемося до даних, представлених у Табл. 1.4 та Табл. 1.5.

Таблиця 4.

Обернена матриця факторних навантажень для факторів,отриманих за допомогою факторного аналізу

№	Етнонаціональна група	Фактори				
		1	2	3	4	5
1.	американці			0,74	0,23	0,13
2.	англійці	0,10		0,81	0,22	0,15
3.	араби		0,73	0,26	0,20	
4.	білоруси	0,29		0,29	0,63	
5.	греки	0,42	0,28	0,46	0,20	
6.	грузини	0,25	0,24	0,29	0,55	-0,17
7.	євреї	0,12	0,53		0,45	
8.	іспанці	0,43	0,30	0,61	-0,13	0,13
9.	італійці	0,38	0,27	0,64		
10.	китайці	0,27	0,64	0,21	0,17	
11.	молдавани	0,22	0,42		0,47	
12.	негри	0,17	0,78			0,10
13.	німці	0,49	0,11	0,27	0,21	0,15
14.	поляки	0,41		0,27	0,36	0,31
15.	португальці	0,50	0,30	0,49		0,15
16.	росіяни	0,12	0,26	-0,15	0,53	0,29
17.	румунни	0,57	0,44		0,20	
18.	словаки	0,82			0,21	
19.	турки	0,46	0,54	0,19	0,15	
20.	угорці	0,67	0,27	0,19	0,30	
21.	українці	0,12		0,13		0,84
22.	українці в ін. країнах	0,13		0,20		0,84
23.	французи	0,55	0,11	0,47		0,17
24.	цигани		0,75			-0,13
25.	чехи	0,67	0,11	0,25	0,12	0,23

Примітка: в матриці вказані значення факторних навантажень, більші 0,1

Таблиця 5.

Обернена матриця факторних навантажень для факторів,отриманих за допомогою категоріального аналізу головних компонент

№	Етнонаціональна група	Фактори			
		1	2	3	4
1.	американці		0,66	0,19	0,16
2.	англійці		0,71	0,19	0,20
3.	араби	0,90	0,13		
4.	білоруси		0,18	0,70	
5.	греки	0,20	0,53	0,46	-0,11
6.	грузини	0,23	0,18	0,58	-0,15

7.	євреї	0,76		0,18	
8.	іспанці	0,22	0,78	0,18	
9.	італійці	0,19	0,78	0,17	
10.	китайці	0,93	0,12		
11.	молдавани	0,63	0,11	0,17	
12.	негри	0,84			0,10
13.	німці	0,14	0,38	0,47	0,21
14.	поляки		0,31	0,49	0,41
15.	португальці	0,19	0,64	0,38	
16.	росіяни	0,30	-0,14	0,47	0,34
17.	румуні	0,77	0,14	0,19	
18.	словаки		0,32	0,73	
19.	турки	0,67	0,23	0,32	-0,12
20.	угорці	0,17	0,33	0,72	
21.	українці		0,16		0,84
22.	українці в ін. країнах		0,20		0,84
23.	французи		0,61	0,47	0,12
24.	цигани	0,76			-0,14
25.	чехи	0,11	0,41	0,58	0,22

Примітка: в матриці вказані значення факторних навантажень, більші 0,1

Принцип простої факторної структури був запропонований Л. Терстоуном, проте існує декілька варіантів його інтерпретації, причому сучасні дослідники констатують, що згадана проста структура не може бути отриманою на реальних даних, а тому слід прагнути отримати таку факторну структуру, яка буде максимально наближеною до простої [11, с. 27–28]. Ми скористаємося найбільш простим визначенням простої факторної структури, згідно з яким для смислової інтерпретації матриці факторних навантажень слід, щоб ці навантаження для кожної зі змінних розподілялися наступним чином: для одного фактора це значення було не меншим 0,7, а для інших – 0,2 [1, с. 224–225].

У випадку результатів факторного аналізу значення факторних навантажень для 12 змінних з 25 не відповідають простій факторній структурі (отримано по два значення факторних навантажень, які більші 0,4), а у випадку результатів категоріального аналізу головних компонент – таких змінних всього 7. Також згідно з результатами факторного аналізу, для 6 змінних значення факторних навантажень не більші 0,5, в той час згідно з результатами категоріального аналізу головних компонент таких змінних всього 3. Отже, за критерієм наближеності до простої факторної структури, факторна структура, отримана в результаті категоріального аналізу головних компонент є кращою, аніж факторна структура, отримана завдяки факторному аналізу.

За критерієм смислової інтерпретації факторна структура, отримана внаслідок категоріального аналізу головних компонент, більше надається до смислової інтерпретації. Виділені з його допомогою 4 фактори можна інтерпретувати в наступний спосіб: фактор 1 – соціальна дистанція до стереотипізованих / оціночно маркованих груп, фактор 2 – соціальна дистанція до народів Західної Європи та Північної Америки, фактор 3 – соціальна дистанція до народів, які мали радянське або комуністичне минуле, фактор 4 – соціальна дистанція до українців. Єдина національна група, стосовно якої виникають сумніви, це румуні, які потрапили до фактора 1, хоча вони не видаються оціночно маркованою групою. Втім, однозначно вважати, що соціальна дистанція щодо румунів була результатом помилкової класифікації, не можна.

Фактори, отримані внаслідок факторного аналізу можна інтерпретувати в наступний спосіб: фактор 1 – соціальна дистанція щодо посткомуністичних народів; фактор 2 – соціальна дистанція до стереотипізованих / оціночно маркованих груп; фактор 3 – соціальна дистанція до народів Західної Європи та Північної Америки; фактор 4 – соціальна дистанція до пострадянських народів; фактор 5 – соціальна дистанція до українців. Проте згадана інтерпретація є не зовсім коректною, оскільки існують дві національні групи, які можна вважати помилково класифікованими: португальці та французи. Вони потрапили до фактора 1, тобто до фактора, де зосереджені посткомуністичні народи. Цей факт підважує коректність результатів факторного аналізу. Також варто зазначити, що у випадку факторів, отриманих внаслідок категоріального методу головних компонент, згадані національні групи потрапили до фактора 2, тобто до інших народів Західної Європи та Північної Америки. Відтак, ми можемо стверджувати, що за критерієм смислової інтерпретації отриманих факторів категоріальний аналіз головних компонент є більш точним.

На основі наведених вище теоретичних міркувань, а також порівняння результатів опрацювання емпіричних соціологічних даних з допомогою факторного аналізу та категоріального аналізу головних компонент, виявлено, що з допомогою останнього було отримано результати, що відзначаються більшою

точністю та відповідністю соціальній реальності. Відтак, категоріальний аналіз головних компонент можна розглядати як альтернативний до факторного аналізу спосіб опрацювання соціологічних даних. Перспективою подальших досліджень є розгляд та статистичне обґрунтування або спростування можливості використання згаданого методу при опрацюванні метричних даних, залежності між якими мають нелінійний характер, а тому не можуть бути опрацьовані з допомогою факторного аналізу.

Література:

1. Паніотто В. Статистичний аналіз соціологічних даних / В.І. Паніотто, В.С. Максименко, Н.М. Харченко. – Київ : Видавничий дім “КМ Академія”, 2004. – 270 с.
2. Орлов А. Статистические методы в российской социологии (тридцать лет спустя) / А.И. Орлов // Социология: 4М. – 2005. – № 20. – С. 32–53.
3. Виноградов О. Психометричні характеристики інтегрального індексу загальних умінь та навичок. / О.Г. Виноградов, Л.О. Малиш // Наукові записки. Національний університет “Києво-Могилянська академія”. – Том 109. Соціологічні науки. – Київ : Національний університет “Києво-Могилянська академія”, 2010. – С. 13–16.
4. Гаригіна О. Хто формує смаки українського суспільства : застосування теорії інновацій у дослідженнях стилів життя / О.С. Гаригіна // Наукові записки. Національний університет “Києво-Могилянська академія”. – Том 109. Соціологічні науки. – Київ : Національний університет “Києво-Могилянська академія”, 2010. – С. 41–45.
5. Manisera M. Identifying the Component Structure of Satisfaction Scales by Nonlinear Principal Components Analysis / Marica Manisera, Anita J. van der Kooij, Elise Dusseldorp // Quality Technology & Quantitative Management. – 2010. – Vol. 7. – No. 2. – P. 97-115.
6. Meulman J. Principal Components Analysis with Nonlinear Optimal Scaling Transformations for Ordinal and Nominal Data / Jacqueline J. Meulman, Anita J. van der Kooij, Willem J. Heiser // Handbook of Quantitative Methodology for the Social Sciences / Ed. by D. Kaplan. – Thousand Oaks, CA : Sage Publications, Inc., 2004. – P. 49–70.
7. Ferrari P. A comparison between alternative models for environmental ordinal data: Nonlinear PCA vs Rasch Analysis / Pieralda Ferrari, Paola Annoni, Silvia Salini // Statistical Solution to Modern Problems : Proceedings of the 20th International Workshop on Statistical Modelling : Sydney, Australia, 11–15 July 2005. – [s.l.]: University of Western Sydney, 2005.– P. 173–177.
8. Manisera M. Component structure of job satisfaction based on Herzberg’s theory [electronic resource] / Marica Manisera, Elise Dusseldorp, Anita J. van der Kooij. – Retrieved from: http://www.datatheory.nl/pages/fullmanuscript_final_epm.pdf
9. Осипов Г. Методы измерения в социологии / Г.В. Осипов, Э.П. Андреев. – Москва : Издательство “Наука”, 1977. – 183 с.
10. Уилкс С. Математическая статистика / С. Уилкс. – Москва : Главная редакция физико-математической литературы, 1967. – 632 с.
11. Ким Дж.-О. Факторный анализ: статистические методы и практические вопросы / Дж.-О. Ким, Ч.У. Мьюллер // Факторный, дискриминантный и кластерный анализ / Под ред. И.С. Енюкова. – Москва : Финансы и статистика, 1989. – С. 5–77.
12. Харман Г. Современный факторный анализ / Гарри Харман. – Москва : Статистика, 1972. – 486 с.
13. Лоули Д. Факторный анализ как статистический метод / Д. Лоули, А. Максвелл. – Москва : Издательство “Мир”, 1967. – 144 с.
14. Финкель Л. Анализ факторный / Л. Финкель // Социологический справочник / Под. общей ред. В.И. Воловича. – Киев : Политиздат Украины, 1990. – С. 190–193.
15. Иберла К. Факторный анализ / К. Иберла. – Москва : Статистика, 1980. – 398 с.
16. Шафир М. Факторный анализ с использованием корреляционной матрицы Спирмена [електронний ресурс] / Марк Шафир. – Режим доступу: http://forum.gfk.ru/filestore/0030/0029/1162/FA_na_matrice_Spirmena.doc.