

УДК 311

СТАТИСТИЧНИЙ АНАЛІЗ СОЦІОЛОГІЧНИХ ДАНИХ МЕТОДАМИ ФАКТОРНОГО АНАЛІЗУ

Любімов Владислав, Акбаш Катерина

*Центральноукраїнський державний педагогічний університет імені
Володимира Винниченка*

***Анотація.** У статті продемонстровано застосування факторного аналізу для обробки результатів анкетних даних соціологічного опитування з ранговими змінними використовуючи статистичний пакет SPSS, а також покроково сформульовані рекомендації щодо інтерпретації отриманих проміжних та кінцевих результатів факторного аналізу. У ході демонстрації прикладу розкриваються внутрішні методи факторного аналізу та методи обертання факторів. У кінці статті наводиться змістовна інтерпретація виділених факторів, отриманих у ході виконання факторного аналізу.*

***Ключові слова:** соціологічні дані, анкетування, факторний аналіз*

STATISTICAL ANALYSIS OF SOCIOLOGICAL DATA BY FACTOR ANALYSIS METHODS

Lyubimov Vladislav, Akbash Kateryna

Volodymyr Vynnychenko Central Ukrainian State Pedagogical University

***Abstract.** The article demonstrates the application of factor analysis to process the results of personal data of a sociological survey with rank variables using the statistical package SPSS, as well as step-by-step recommendations for the interpretation of the obtained intermediate and final results of factor analysis. During the demonstration of the example, the internal methods of factor analysis and methods of rotation of factors are revealed. At the end of the article is a meaningful interpretation of the selected factors obtained during the factor analysis.*

***Key words:** sociological data, questionnaire, factor analysis*

Постановка проблеми. Для дослідження поведінки людини використовують багато різноманітних математичних методів. У переважній більшості сучасних практичних і наукових літературних джерелах соціально-психологічного напрямку для аналізу досліджуваних проблем застосовують

різні розділи математики, часом надто абстрактні. А один розділ – математичну статистику – використовують найчастіше [1].

Інтенсивне застосування в соціально-психологічних науках методів математичної статистики пов'язано насамперед з природою поведінки людини. Поведінку людини неможливо описати достеменно точно, вона не детермінована. Людина не повною мірою передбачувана у вчинках як на індивідуальному, так і на глобальному суспільному рівнях.

Отже, кожна особа унікально індивідуальна. Не існує двох ідентичних людей. Тому, намагаючись описати найважливіші властивості, притаманній певній групі людей, необхідно знайти в виокремити такі їх характеристики, які найчастіше виявляються в групі.

З огляду на це, поведінку людини і суспільства зазвичай описують за допомогою ймовірнісних підходів. Закономірності стохастичної природи виявляються в багатьох аспектах індивідуального та суспільного життя. А математична статистика забезпечує методи аналізу сукупностей даних ймовірнісної природи.

На сьогодні в соціології, зокрема в соціологічних опитуваннях, широко застосовують основні методи математичної статистики.

Основний акцент даної статті зроблений на характеристику сутності складних методів обробки результатів соціологічних досліджень (зокрема факторного аналізу) за допомогою статистичного пакету SPSS та інтерпретації отриманих результатів.

Аналіз досліджень і публікацій. Аналізуючи наукову літературу, ми виявили ряд математико-статистичних методів, які активно використовуються в соціологічних дослідженнях і ряд методів, які тільки починають використовуватися [2-4]. Активно використовуються статистичні характеристики сукупностей даних і методи їх знаходження, основні статистичні розподіли та їх властивості, вибіркові дослідження (методи формування і основні принципи аналізу вибірок), оцінювання параметрів генеральних сукупностей, основи кореляційно-регресійного аналізу.

На сьогодні соціологи працюють із величезними масивами даних. У процесі проведення досліджень соціолог часто зіштовхується із необхідністю структурувати і типізувати дані. Тобто із великої кількості різнорідних первинних даних потрібно скласти як із мозаїки, загальну картину стану досліджуваної сукупності (суспільства в цілому або окремих його груп). Ось тут на допомогу соціологам можуть прийти такі сучасні методи, які кластерний та факторний аналізи. Подібного роду дослідження наведені у роботі [5].

Мета статті: розробка алгоритму використання факторного аналізу для обробки даних соціологічних опитувань з ранговими змінними.

Виклад основного матеріалу дослідження. Для демонстрації застосування факторного аналізу візьмемо приклад анкети Інституту соціології НАН «Соціально-психологічний портрет», із якої оберемо 17 тверджень. В у мовах пандемії ми вирішили використати безпечний вид анкетування, а саме електронне анкетування на основі сервісу Google форми.

Анкета складається із 17 тверджень, які дозволять скласти соціально-психологічний портрет респондентів відносно їх впевненості у собі і рівні самоповаги. Результати анкетування фіксуються по чотирьох бальній шкалі (від «повністю згоден» до «повністю не погоджуюсь»).

Твердження анкети були наступними:

- Я не можу справитись зі своїми проблемами (С-1);
- Іноді я відчуваю, що мною керують у житті (С-2);
- Я мало можу впливати на те, що зі мною відбувається (С-3);
- Я завжди можу виконати задумане (С-4);
- Я часто відчуваю себе безпорадним перед проблемами, що виникають в моєму житті (С-5);
- Те, що зі мною станеться в майбутньому, багато в чому залежить від мене (С-6);
- Те, що я можу зробити, мало що змінить у моєму житті (С-7);
- Я думаю, що я нічим не гірше інших (С-8);

- Я вважаю, що у мене є багато хороших якостей (С-9);
- Загалом, мені здається, що я невдаха (С-10);
- Я можу все робити не гірше за інших (С-11);
- Я думаю, що мені особливо нічим пишатися (С-12);
- Я добре ставлюся до самого (самої) себе (С-13);
- В цілому я задоволений (задоволена) собою (С-14);
- Іноді я відчуваю себе марним (марною) (С-15);
- Я хотів (хотіла) б ставитися до себе з великою повагою (С-16);
- Іноді мені здається, що я нехороша людина (С-17).

Анкета розміщена за наступним посиланням https://docs.google.com/forms/d/e/1FAIpQLScsI2FA7vnosurNhFjAJ1AJwf_9o9JGPHuyGvjFHAIawTgwHQ/viewform.

Тепер покажемо на цьому прикладі, що може дати соціологу такий аналітичний інструмент, як факторний аналіз [6]. У пакеті SPSS потрібно обрати пункт «Факторний аналіз», який розміщений в пункті меню «Аналіз» - «Зниження розмірності» (рис. 1).

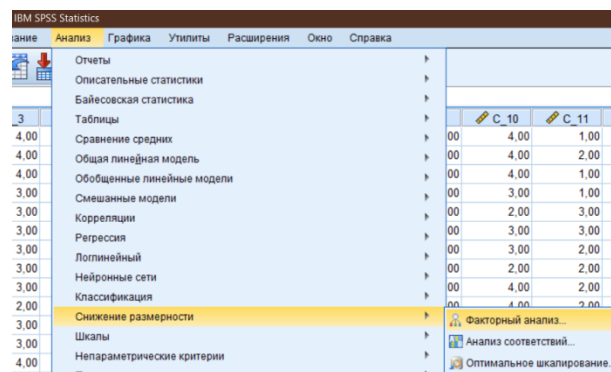


Рис. 1. Вибір пункту «Факторний аналіз» в меню «Аналіз: зниження розмірності»

Для проведення факторного аналізу в SPSS є можливість використати декілька методів:

1. *Метод головних компонент*, заснований на послідовному пошуку найбільш впливових факторів у порядку зменшення їх важливості для пояснення залежностей між змінними. У цьому випадку перший виділений

фактор буде пояснювати найбільшу долю дисперсії ознаки, другий фактор – наступну найбільшу долю дисперсії і т.д.

2. *Незважений метод найменших квадратів*, мінімізуючий суму квадратів різниць між спостережуваною і відтвореною кореляційною матрицею без урахування діагоналей.

3. *Узагальнений метод найменших квадратів*, схожий на попередній, але відрізняється від нього тим, що кореляції зважуються величинами, оберненими характеристикам, так що змінні з високою характерністю отримують менше ваги, ніж змінні з високою.

4. *Метод максимальної правдоподібності*, при якому в якості оцінок параметрів обираються ті, для яких спостережувана кореляційна матриця найбільш правдоподібна. При цьому методі кореляції зважуються значеннями, оберненими до характеристик змінних.

5. *Метод факторизації на головній осі*, дозволяє виділити фактори з вихідної матриці кореляцій з квадратами коефіцієнтів множинних кореляцій по діагоналі в якості початкових оцінок спільностей. У ході факторного аналізу за допомогою цього методу наступні значення виділених спільностей заміщують початкові до тих пір, поки не буде знайдений найбільш прийнятний розв'язок.

Для нашого прикладу ми використаємо самий розповсюджений метод - метод головних компонент. Проінтерпретуємо отримані результати.

У таблиці 1 міститься інформація про те, яка частина дисперсії кожної з врахованих змінних може бути пояснена запропонованою факторною моделлю.

Наприклад, ми бачимо, що відношення респондентів до твердження «В цілому я задоволений (задоволена) собою» (С-14), згода або незгода з ним, пояснюється даною факторною моделлю на 87,1% (це максимальне значення), тоді як відношення до твердження «Я добре ставлюся до самого (самої) себе» (С-13) тільки на 57%.

Таблиця 1. Спільності

	Початкова	Виділена
C_1	1,000	,734
C_2	1,000	,626
C_3	1,000	,716
C_4	1,000	,632
C_5	1,000	,668
C_6	1,000	,683
C_7	1,000	,772
C_8	1,000	,779
C_9	1,000	,740
C_10	1,000	,743
C_11	1,000	,691
C_12	1,000	,617
C_13	1,000	,570
C_14	1,000	,871
C_15	1,000	,703
C_16	1,000	,596
C_17	1,000	,721

Метод виділення факторів: метод головних компонент.

Ця таблиця може бути корисна для того, щоб виключити дуже «непередбачувані» змінні з аналізу, тим самим підвищити загальну прогностичну успішність факторної моделі. У нашому випадку жодна змінна-запитання не підлягає виключенню, так як всі значення більші за 50%.

У таблиці 2 видно, яку саме долю загальної дисперсії може пояснити кожен із виділених факторів окремо і вся побудована факторна модель уцілому. У нашому випадку запропонована комп'ютером на основі розрахунків факторна модель може пояснити 69,8% дисперсії. Причому перший фактор передбачає 34,36% загальної дисперсії, другий 45,95%, третій 55,18%, четвертий 62,81%.

Таблиця 2. Повна пояснююча дисперсія

Компонента	Початкові власні значення			Виявлення суми квадратів навантажень		
	Всього	% дисперсії	Суммарний %	Всього	% дисперсії	Суммарний %
1	5,842	34,363	34,363	5,842	34,363	34,363
2	1,970	11,589	45,952	1,970	11,589	45,952
3	1,568	9,225	55,177	1,568	9,225	55,177
4	1,298	7,636	62,813	1,298	7,636	62,813
5	1,184	6,965	69,777	1,184	6,965	69,777
6	,972	5,715	75,492			
7	,748	4,403	79,895			
8	,661	3,890	83,785			
9	,580	3,410	87,195			
10	,518	3,045	90,240			
11	,475	2,795	93,035			
12	,322	1,894	94,929			
13	,289	1,699	96,628			
14	,197	1,162	97,790			
15	,158	,932	98,722			
16	,129	,757	99,478			
17	,089	,522	100,000			

Метод виділення факторів: метод головних компонент.

Така таблиця може бути корисною для прийняття дослідником кінцевого рішення про кількість виділених факторів, так як не має смислу використовувати фактор, який може пояснити дуже малу долю дисперсії. Такий фактор краще прийняти із загального числа факторів. Це спростить модель і дозволить її правильно інтерпретувати.

У нашому випадку видно, шести факторна модель може дати додаткову інформацію по відношенню до п'яти факторної, але збільшення кількості факторів більше п'яти – не має смислу. Це видно із таблиці 2.

Далі у діалоговому вікні «Факторний аналіз: Виділення факторів» задамо кількість факторів самі. Візьмемо за основу шестифакторну модель з пояснюючою здатністю 69,8%. Для цього оберемо у розділі «Виділити» діалогового вікна «Факторний аналіз: Виділення факторів» фіксовану кількість факторів – 6.

Щоб результат побудови факторної моделі був більш наочний, виконаємо наступні дії. По-перше, видалимо із таблиці значення коефіцієнтів кореляції менших за 0,5 (у кожному випадку при визначенні цього порогу дослідник повинен виходити із логіки дослідження і принципу доцільності). Їх можна не враховувати, так як вони дуже малі. По-друге, скористаємось методом обертання факторів і побачимо наочно його результативність. Як правило, обертання потрібно для того, щоб спростити інтерпретацію факторів чи змінних. У SPSS досліднику запропоновано декілька варіантів обертання факторів – варімакс, облімін, квартимакс, еквімакс, промакс.

Варімакс – метод ортогонального обертання факторів, який мінімізує число змінних з високими навантаженнями на кожен фактор, що суттєво спрощує інтерпретацію факторів.

Прямий облімін – метод косокутного (неортогонального) обертання факторів.

Квартимакс – метод обертання факторів, який мінімізує число факторів, які потрібні для пояснення кожної змінної, що суттєво спрощує інтерпретацію спостережуваних змінних.

Еквімакс – метод обертання факторів, який поєднує методи варімакс, що спрощує фактори, і квартимакс, що спрощує змінні. При його використанні мінімізується число змінних з великими факторними навантаженнями і кількість факторів, які потрібні для пояснення кожної змінної.

Промакс – метод косокутного обертання, який застосовується для факторної моделі, в якій фактори можуть корелювати між собою.

Вибір конкретного способу обертання завжди залишається за дослідником, який відштовхується від задач свого дослідження, тобто із бажаних результатів.

Для нашого прикладу ми оберемо варімакс обертання для того, щоб максимально спростити інтерпретацію виділених факторів. Тепер замість однієї матриці факторних навантажень SPSS виводить дві – просту і повернуту (табл. 3).

Таблиця 3. Матриця повернутих компонент (з кореляціями >0,5)

	Компонента					
	1	2	3	4	5	6
C_17	,858					
C_12	,637	,600				
C_9	-,610		,514			
C_15	,606					
C_4	-,604					
C_7		,763				
C_1		,749				
C_5		,651				
C_10		,529				
C_8			,850			
C_11			,829			
C_13				,857		
C_14				,799		
C_3					,799	
C_2					,695	
C_6						,818
C_16						,727

Метод обертання факторів: метод головних компонент.
 Метод обертання: варімакс з нормалізацією Кайзера.
 а. обертання зійшлося за 9 ітерацій.

Таким чином, на основі виявлення ступеню відповідності всіх тверджень були виявлені наступні 6 факторів:

- Фактор 1 (Ф-1): твердження 17,12,9,15,4 (всього п'ять тверджень):
 - ✓ Іноді мені здається, що я нехороша людина (С-17);
 - ✓ Я думаю, що мені особливо нічим пишатися (С-12);
 - ✓ Іноді я відчуваю себе марним (марною) (С-15);
 - ✓ Я вважаю, що у мене є багато хороших якостей (С-9);
 - ✓ Я завжди можу виконати задумане (С-4).
- Фактор 2 (Ф-2): твердження 7,1,5,10 (всього чотири твердження);
 - ✓ Те, що я можу зробити, мало що змінить у моєму житті (С-7)
 - ✓ Я не можу справитись зі своїми проблемами (С-1);
 - ✓ Я часто відчуваю себе безпорадним перед проблемами, що виникають в моєму житті (С-5);
 - ✓ Загалом, мені здається, що я невдаха (С-10).

- Фактор 3 (Ф-3): твердження 9,8,11 (всього три твердження);
 - ✓ Я вважаю, що у мене є багато хороших якостей (С-9);
 - ✓ Я думаю, що я нічим не гірше інших (С-8);
 - ✓ Я можу все робити не гірше за інших (С-11).
- Фактор 4 (Ф-4): твердження 13,14 (всього два твердження);
 - ✓ Я добре ставлюся до самого (самої) себе (С-13);
 - ✓ В цілому я задоволений (задоволена) собою (С-14).
- Фактор 5 (Ф-5): твердження 2,3 (всього два твердження);
 - ✓ Іноді я відчуваю, що мною керують у житті (С-2);
 - ✓ Я мало можу впливати на те, що зі мною відбувається (С-3).
- Фактор 6 (Ф-6): твердження 6,16 (всього два твердження);
 - ✓ Те, що зі мною станеться в майбутньому, багато в чому залежить від мене (С-6);
 - ✓ Я хотів (хотіла) б ставитися до себе з великою повагою (С-16).

Отже, ми виокремили фактори і подивилися, які змінні охопили ці фактори.

Можна виділити три блоки факторів:

- Блок факторів, які характеризують позицію соціального безсилля, коли людина пасує перед життєвими складнощами і обирає рухатися за течією «річки життя». Тобто це позиція соціальної пасивності і невіра у можливість щось змінити у своєму житті, а самі фактори, що входять до цього блоку фіксують ступінь активності життєвої позиції. Таким чином блок цих факторів матиме шкалу: соціальна пасивність (соціальне безсилля) – соціальна активність (ініціативність). До цієї групи можна віднести фактори 2 і 5.

- Наступний блок факторів складається із суджень, які характеризують в цілому позицію, направлену на самоповагу і самозадоволеність. Таким чином по групі цих факторів можна визначити рівень самозадоволеності респондентів, тобто він охопив би твердження, які в цілому характеризують позицію, направлену на самоповагу і самозадоволення. Шкалу оцінки цього критерію можна визначити від «повністю не задоволений собою» до «повністю задоволений собою». До цієї групи можна віднести фактор 3.

- Блок факторів, який в цілому відображає характер самоповаги респондентів, оснований на суспільній корисності. До цієї групи можна віднести фактори 1,4,6.

Висновки та перспективи подальших пошуків у напрямі дослідження. На даному прикладі ми продемонстрували можливість використання факторного аналізу у практиці соціологічних досліджень. Досліднику слід розуміти, що сам по собі факторний аналіз беззмістовний, якщо його результати не можна логічно проінтерпретувати. Тому при проведенні факторного аналізу дослідник повинен зважувати всі «за» і «проти» на кожному етапі виконання даного виду аналізу – від відбору змінних і визначення кількості факторів до вибору методу аналізу і способів обертання (у випадку необхідності) матриць факторних значень. Тільки послідовне і продумане здійснення всіх цих етапів даного виду аналізу може дати змістовний результат.

Список використаної літератури

1. Телейко А.Б. Математико-статистичні методи в соціології та психології: Навч. Посіб. / А.Б. Телейко, Р.К. Чорней. – К.: МАУП, 2007. – 424 с.
2. Максименко В.С., Паніотто В.І., Харченко Н.М. Статистичний аналіз соціологічних даних. – К.: Видав. Дім «КМ Академія», 2004.
3. Захаров В.П. Применение математических методов в социально-психологических исследованиях. – Л.:Изд-во ЛГУ, 1985.
4. Кендалл М.Дж., Стюарт А. Статистические алгоритмы в социологических исследованиях. – Новосибирск: Наука, 1985.
5. Фарахутдинов, Ш. Ф. Обработка и анализ данных социологических исследований в пакете SPSS 17.0. Курс лекций : учебное пособие / Ш. Ф. Фарахутдинов, А. С. Бушуев. – Тюмень : ТюмГНГУ, 2011. – 220 с.
6. Лупан І.В., Авраменко О.В., Акбаш К.С. Комп'ютерні статистичні пакети: навчально-методичний посібник. – 2-ге вид. – Кіровоград: «КОД» 2015. – 236 с.