

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ  
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ВНУТРІШНІХ СПРАВ**

*Факультет № 6  
Кафедра соціології та психології*

**ТЕКСТ ЛЕКЦІЇ**

з навчальної дисципліни

**«Комп'ютерні методи практичної психології»**  
обов'язкових компонент освітньої програми  
першого (бакалаврського) рівня вищої освіти

*053 Психологія (практична психологія)*

**Тема №10. Аналіз психологічних даних в STATISTICA, SPSS:  
факторний аналіз**

**Харків 2023**

**ЗАТВЕРДЖЕНО**

Науково-методичною радою  
Харківського національного  
університету внутрішніх справ  
Протокол від 30.08.2023 р. №7

**СХВАЛЕНО**

Вченою радою факультету №6  
Протокол від 25.08.2023 р. №7

**ПОГОДЖЕНО**

Секцією Науково-методичної  
ради ХНУВС з гуманітарних та  
соціально-економічних дисциплін  
Протокол від 29.08.2023 р. №7

Розглянуто на засіданні кафедри соціології та психології  
Протокол від 15.08.2023 р. №8

**Розробник:**

Професор кафедри соціології та психології факультету №6  
д-р соціол. н., професор Нечитайло Ірина Сергіївна

**Рецензенти:**

1. Керівник психологічної служби Харківського гуманітарного університету «Народна українська академія», доцент кафедри соціології та гуманітарних дисциплін, к. психол. н., Гога Н. П.;
2. Доцент кафедри соціології та психології факультету №6, к. психол. н., доцент Філоненко В. М.

## ТЕМА №10. АНАЛІЗ ПСИХОЛОГІЧНИХ ДАНИХ В STATISTICA, SPSS: ФАКТОРНИЙ АНАЛІЗ

### План лекції

**10.1.** Основні поняття факторного аналізу.

**10.2.** Деякі специфічні аспекти процедури факторного аналізу.

### Рекомендована література

#### *Основна*

1. Нечитайло І. С., Бірюкова М. В. Математичні методи в соціології : підручник для студентів ВНЗ / Нар. укр. акад., [каф. соціології]. Харків : Вид-во НУА, 2012. 243 с.

2. Татьянчиков А. О. Математичні методи в психології: навчально-методичні рекомендації (в допомогу до самостійної роботи для здобувачів вищої освіти ступеня бакалавра факультету психології, політології та соціології) ; кафедра психології НУ «Одеська юридична академія». Одеса : Фенікс, 2021. 48 с.

#### *Допоміжна*

3. Катаєв Є.С. Використання статистичних методів обробки даних у дослідженнях “я-концепції” особистості. Вісник Національного університету оборони України. 2012. №2 (27) /2012. С. 171-176.

4. Салюк М. А Статистична обробка даних експериментального дослідження. Методичний посібник з курсу «Експериментальна психологія» / за ред. Е.Л. Носенко. Дніпропетровськ: Інновація, 2010. 26 с.

5. Старушенко Г. А. Статистична обробка даних в системі публічного управління : навч. посіб. Дніпро : ГРАНІ, 2018. 144 с.

6. Татьянчиков А.О. Методичні рекомендації до виконання лабораторних робіт з курсу «Методи психологічного дослідження: математичні методи в психології». Одеса : Вид-во Університету Ушинського, 2019. 38 с.

### ТЕКСТ ЛЕКЦІЇ

#### 10.1. Основні поняття факторного аналізу

*Факторний аналіз* – статистичний метод аналізу впливу окремих факторів (чинників) на результативний показник. В останні 30-40 років факторний аналіз набув значної популярності в психологічних і соціологічних дослідженнях. Багато в чому цьому сприяла розробка Раймондом Кеттеллом (Raymond B. Cattell) відомого 16-факторного особистісного опитувальника (16PF). Саме за допомогою факторного аналізу йому вдалося звести близько 4500 найменувань особистісних особливостей до 187 питань, які, в свою чергу,

дозволяють виміряти 16 різних властивостей особистості.

Факторний аналіз дає можливість кількісно визначити те, що безпосередньо не вимірюється, виходячи з декількох доступних для вимірювання змінних.

Наприклад, характеристики «відвідує розважальні заходи», «багато розмовляє», «охоче йде на контакт з будь-якою незнайомою людиною» можуть слугувати компонентами такої якості, як «товариськість», яка безпосередньо не піддається кількісному вимірюванню.

Факторний аналіз дозволяє встановити для великого числа вихідних ознак порівняно вузький набір «властивостей», що характеризують зв'язок між групами цих ознак, званих факторами. Процедура факторного аналізу складається з чотирьох основних стадій:

1. Обчислення кореляційної матриці для всіх змінних, що беруть участь в аналізі;
2. Виведення факторів;
3. Обертання факторів для створення спрощеної структури;
4. Інтерпретація факторів.

*Обчислення кореляційної матриці.* Перша операція, яка проводиться при виконанні факторного аналізу, – це обчислення кореляційної матриці для змінних, що беруть участь в аналізі. Вже за наявності такої початкової дії можна зробити висновок про те, що факторний аналіз заснований на взаємодії змінних. Для проведення факторного аналізу зовсім не обов'язково спеціально будувати кореляційну матрицю: при необхідності програма SPSS створить її сама на основі даних файлу. Іноді не потрібні навіть вихідні дані – достатньо мати кореляційну матрицю, яка в цьому випадку вводиться в командний файл SPSS. При необхідності можна звернутися до керівництва користувача програми SPSS.

*Виведення факторів.* Ця процедура являє собою наступний етап факторного аналізу. З математичної точки зору виведення чинників має певну аналогію з множинним регресійним аналізом. Першим кроком у множинному регресійному аналізі є вибір тієї незалежної змінної, яка обумовлює найбільшу частку дисперсії залежної змінної. Потім операція повторюється для незалежних змінних до тих пір, поки частка дисперсії не перестане бути значимою. При факторному аналізі реалізується аналогічна процедура.

Виведення фактору починається з підрахунку сумарного розкиду значень всіх що беруть участь в аналізі змінних (дана величина чимось схожа на загальну суму квадратів). Для цього «сумарного розкиду» непросто підібрати логічну інтерпретацію, однак він є цілком строго певною математичною величиною.

Першим завданням факторного аналізу є вибір взаємодіючих змінних, чия взаємна кореляція обумовлює найбільшу частку загальної дисперсії. Ці змінні утворюють перший фактор.

Потім перший фактор виключається, і з решти безлічі змінних знову вибираються ті, чия взаємодія визначає найбільшу частку загальної дисперсії, що залишилася після побудови першого фактору. Ці змінні утворюють другий

фактор.

Процедура вилучення факторів продовжується до тих пір, поки не буде вичерпана вся загальна дисперсія змінних. За замовчуванням у процедурі факторного аналізу кожна змінна має одиничне значення спільності. Цей показник дорівнює частці дисперсії змінної, обумовленої сукупним впливом факторів. Спільність можна порівняти з множинним коефіцієнтом кореляції  $R$ , що приймають значення 0 у випадку, якщо фактори не впливають на змінну, значення 1 у випадку, якщо дисперсія змінної цілком визначається виділеними факторами. Перед початком вилучення факторів одиничне значення спільності встановлено за промовчанням для змінних, які беруть участь у факторному аналізі.

Після того, як процедура отримує перший фактор, навпроти його номера з'являється його власне значення, наприклад поруч з числом 1 з'являється значення 5,13312. Власне значення фактору є пропорційним частці в загальній дисперсії, обумовленій даним фактором (зверніть увагу, що тут мова йде про фактор, а не про змінну, як було у випадку спільності).

Перше власне значення завжди є найбільшим і перевищує одиницю, що визначається алгоритмом процедури, згідно з якою фактори вилучаються у порядку зменшення їх впливу на дисперсію змінних.

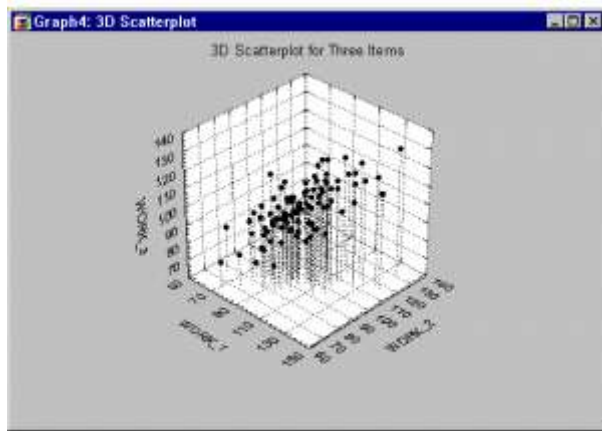
Потім обчислюється відсоток дисперсії, обумовлений даним фактором і рівний відношенню власного значення фактору до кількості змінних, а також відповідний кумулятивний (накопичений) відсоток. З отриманням кожного нового фактору власні значення зменшуються, а кумулятивний відсоток наближається до 100.

Зверніть увагу, що в контексті процедури вилучення факторів жодного разу не прозвучав термін «значимість»: на відміну від регресійного аналізу витяг факторів відбувається до тих пір, поки не буде вичерпана вся дисперсія змінних, незалежно від того, чи є значимим вплив фактору на дисперсію чи ні.

## 10.2. Деякі специфічні аспекти процедури факторного аналізу

*Метод головних компонент.* Приклад, у якому дві корельовані змінні об'єднані в один фактор, показує головну ідею факторного аналізу або, точніше, аналізу головних компонент. Якщо приклад із двома змінними поширити на більшу кількість змінних, то обчислення стають складнішими, проте основний принцип – уявлення двох або більше залежних змінних одним фактором – залишається чинним.

У тому випадку, коли є більше двох змінних, можна вважати, що вони визначають тривимірний «простір» так само, як дві змінні визначають площину. Якщо ви маєте три змінні, можете побудувати 3М діаграму розсіювання (Див. Рис. 10.1).



Для випадку більше трьох змінних, стає неможливим уявити точки на діаграмі розсіювання, проте логіка обертання осей з метою максимізації дисперсії нового фактору залишається незмінною.

*Деякі ортогональні фактори.* Після того, як ми знайшли лінію, для якої дисперсія є максимальною, навколо неї залишається деякий розкид даних. Процедурі потрібно повторити. В аналізі головних компонент саме так і робиться: після того, як перший фактор виділений, тобто після того, як перша лінія проведена, визначається наступна лінія, що максимізує залишкову варіацію (розкид даних навколо першої прямої тощо). Таким чином, фактори послідовно виділяються один за одним. Оскільки кожен наступний фактор визначається так, щоб максимізувати мінливість, що залишилася від попередніх, фактори виявляються незалежними один від одного. Іншими словами, некорельованими чи ортогональними.

Нагадаємо, що аналіз основних компонент є способом скорочення чи редукції даних, тобто шляхом скорочення числа змінних. Виникає природне питання: скільки факторів слід виділяти? Зазначимо, що в процесі послідовного виділення факторів вони включають все менше і менше мінливості. Рішення про те, коли слід зупинити процедуру виділення факторів, головним чином залежить від погляду на те, що вважати мінімальною «випадковою» мінливістю. Це рішення досить довільне, однак є деякі рекомендації, що дозволяють раціонально вибрати число факторів. У якості прикладу розглянемо деякі стандартні результати аналізу головних компонент. При повторних ітераціях виділяються фактори з меншою і меншою дисперсією.

Для простоти викладу результатів факторного аналізу, зазвичай, робота починається з матриці, у якій дисперсії всіх змінних дорівнюють 1.0. Тому загальна дисперсія дорівнює числу змінних. Наприклад, якщо маємо 10 змінних, кожна з яких має дисперсію 1, то найбільша мінливість, яка потенційно може бути виділена, дорівнює 10 разів по 1. Припустимо, що при вивченні ступеня задоволеності життям ви включили 10 пунктів для вимірювання різних аспектів задоволеності домашнім життям та роботою. Дисперсія, пояснена послідовними факторами, представлена у таблиці 10.1.

STATISTICA ФАКТОРНИЙ АНАЛІЗ	Власні значення (factor.sta) Виділення: Головні компоненти				
Значення	Власні значення	% загальної дисперсії	Кумулят. власні. значення	Кумулят. %	
1	6.118369	61.18369	6.11837	61.1837	
2	1.800682	18.00682	7.91905	79.1905	
3	.472888	4.72888	8.39194	83.9194	
4	.407996	4.07996	8.79993	87.9993	
5	.317222	3.17222	9.11716	91.1716	
6	.293300	2.93300	9.41046	94.1046	
7	.195808	1.95808	9.60626	96.0626	
8	.170431	1.70431	9.77670	97.7670	
9	.137970	1.37970	9.91467	99.1467	
10	.085334	.85334	10.00000	100.0000	

*Факторний аналіз як метод класифікації.* Термін факторний аналіз тепер буде включати як аналіз основних компонент, так і аналіз основних факторів. Передбачається, що ви знаходитесь у тій точці аналізу, коли загалом знаєте, скільки факторів слід виділити. Ви можете захотіти дізнатися про значущість факторів, тобто, чи можна інтерпретувати їх розумним чином і як це зробити. Щоб проілюструвати, як це може бути зроблено, виробляються дії «у зворотному порядку», тобто, починають із деякої осмисленої структури, та дивляться, як це відбивається на результатах. У таблиці 10.2 наведено кореляційну матрицю для змінних, що відносяться до задоволеності справами на роботі та вдома.

STATISTICA ФАКТОРНИЙ АНАЛІЗ	Корреляції (factor.sta) Построчне видалення ПД n=100					
Змінна	РОБОТА_1	РОБОТА_2	РОБОТА_3	ДОМ_1	ДОМ_2	ДОМ_3
РОБОТА_1	1.00	.65	.65	.14	.15	.14
РОБОТА_2	.65	1.00	.73	.14	.18	.24
РОБОТА_3	.65	.73	1.00	.16	.24	.25
ДОМ_1	.14	.14	.16	1.00	.66	.59
ДОМ_2	.15	.18	.24	.66	1.00	.73
ДОМ_3	.14	.24	.25	.59	.73	1.00

Змінні, що відносяться до задоволеності на роботі, більш корелюють між собою. Змінні, які відносяться до задоволеності домашніми справами, також корелюють між собою, але дещо менше. Кореляції між цими двома типами змінних (змінні, пов'язані із задоволеністю справами на роботі, та змінні, пов'язані із задоволеністю справами вдома) порівняно малі. Тому здається правдоподібним, що є два відносно незалежні фактори (два типи факторів), відображені в кореляційній матриці.

*Факторні навантаження.* Тепер проведемо аналіз основних компонентів і розглянемо рішення з двома факторами. Для цього розглянемо кореляції між змінними та двома факторами (або «новими» змінними), які

були виділені за замовчанням. Ці кореляції називають факторними навантаженнями (див. Табл. 10.3)

STATISTICA ФАКТОРНИЙ АНАЛІЗ	Факторні навантаження (без обертів) Головні компоненти	
Змінна	Фактор 1	Фактор 2
РОБОТА_1	.654384	.564143
РОБОТА_2	.715256	.541444
РОБОТА_3	.741688	.508212
ДОМ_1	.634120	-.563123
ДОМ_2	.706267	-.572658
ДОМ_3	.707446	-.525602
Загальна дисперсія	2.891313	1.791000
Частка загальної дисперсії	.481885	.298500

Вочевидь, перший фактор включає в себе більш корельовані змінні, ніж другий. Це слід очікувати, тому що, як було сказано вище, фактори виділяються послідовно і містять все менше і менше загальної дисперсії.

*Обертання факторів.* За дуже рідкісними винятками для дослідника не представляють інтересу всі виявлені фактори. Якщо факторів виявиться стільки ж, скільки вихідних змінних, факторний аналіз втрачає сенс, оскільки його метою є скорочення кількості змінних (включених до вихідного набору). Отже, потрібно прийняти рішення, які з факторів слід залишити для подальшого аналізу. Тут, в першу чергу, рекомендується керуватися здоровим глуздом і залишати ті фактори, які мають зрозумілу теоретичну або логічну інтерпретацію. Однак не завжди представляється можливим заздалегідь встановити призначення кожного фактору, і тому дослідники на першому етапі зазвичай використовують формальні критерії.

За замовчуванням при виконанні команди Factor (Факторний аналіз) всі фактори, *чії власні значення перевищують одиницю*, зберігаються для подальшого аналізу. Оскільки число факторів дорівнює числу змінних, лише для невеликої кількості факторів власні значення виявляються більшими за одиницю, виконання процедури обертання факторів дозволяє радикально скоротити число факторів.

Існують і інші критерії виділення факторів (наприклад, критерій «кам'янистого осипу» Р. Кеттелла). Крім того можна вибрати фактори, виходячи з відомих нам особливостей конкретного масиву даних. У будь-якому випадку остаточне рішення про кількість факторів зазвичай приймається після інтерпретації факторів, отже, факторний аналіз передбачає неодноразове виділення різної кількості факторів.

Таким чином, важливим кроком після виділення факторів є їх обертання. Обертання потрібно тому, що спочатку структура факторів, будучи математично коректною, як правило, є важкою для інтерпретації. Метою обертання є отримання простої структури, яка акумулює великі значення навантажень кожної змінної лише за одним фактором і малі – по всіх інших факторам. Навантаження відображає зв'язок між змінною і фактором, будучи



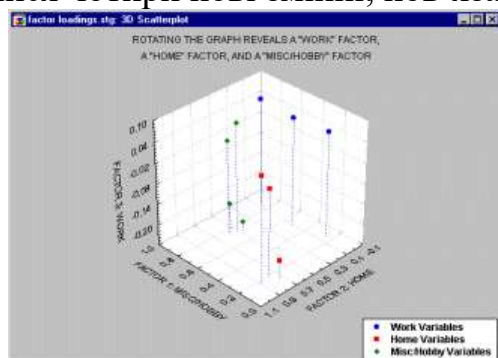
подобою коефіцієнта кореляції. Значення навантаження лежить в межах від -1 до 1. Ідеальна проста структура передбачає, що кожна змінна має нульові значення навантажень для всіх факторів, крім одного, для якого навантаження цієї змінної є близьким до 1 (-1).

Існують різні методи обертання факторів. Метою цих методів є отримання зрозумілої (інтерпретованої) матриці навантажень, тобто факторів, які ясно відзначені високими навантаженнями для деяких змінних та низькими – для інших. Цю загальну модель іноді називають «простою структурою» (більш формальне визначення можна знайти в стандартних підручниках).

Типовими методами обертання є стратегії варімакс, квартімакс та еквімакс. Ідея обертання за методом варімакс була описана вище. Цей метод є досить часто вживаним. Нижче наведено таблицю (10.4) навантажень на фактори, які пройшли процедуру обертання.

STATISTICA ФАКТОРНИЙ АНАЛІЗ	Факторні навантаження (Варімакс нормаліз.) Виділення: Головні компоненти	
Змінна	Фактор 1	Фактор 2
РОБОТА_1	.862443	.051643
РОБОТА_2	.890267	.110351
РОБОТА_3	.886055	.152603
ДОМ_1	.062145	.845786
ДОМ_2	.107230	.902913
ДОМ_3	.140876	.869995
Загальна дисперсія	2.356684	2.325629
Частка загальн. дисп.	.392781	.387605

*Інтерпретація факторної структури.* Як і очікувалося, перший фактор відзначений високими навантаженнями на змінні, пов'язаний із задоволеністю справами на роботі, а другий фактор – із задоволеністю будинком. З цього можна зробити висновок, що задоволеність, виміряна вашим інструментарієм (опитувальником), складена з двох частин: 1) задоволеність справами вдома; задоволеність справами на роботі. Таким чином, зроблено класифікацію змінних. Розглянемо наступний приклад, тут до попереднього прикладу додалися чотири нові змінні, пов'язані із задоволеністю хобі (див. Рис. 10.2).



На цьому графіку факторних навантажень 10 змінних були зведені до трьох факторів: фактор задоволеності роботою (work); фактор задоволеності справами вдома (home); фактор задоволеності хобі (hobby/misc).

Отримана кореляційна матриця (див. Табл. 10.4) для косокутних

факторів піддається подальшому аналізу для того, щоб виділити безліч ортогональних факторів, що розділяють мінливість у змінних на ту, що відносяться до розподіленої або загальної дисперсії (вторинні фактори), і на приватні дисперсії, що відносяться до кластерів або схожих пунктів опитувальника) в аналізі (первинні фактори). Такий ієрархічний аналіз може дати факторні навантаження, представлені у таблиці 10.5:

<b>STATISTICA ФАКТОРНИЙ АНАЛІЗ</b>	<b>Вторинні і первинні факторні навантаження</b>		
<b>Фактор</b>	<b>Вторин. 1</b>	<b>Первин. 1</b>	<b>Первин. 2</b>
РОБОТА_1	.483178	.649499	.187074
РОБОТА_2	.570953	.687056	.140627
РОБОТА_3	.565624	.656790	.115461
ДОМ_1	.535812	.117278	.630076
ДОМ_2	.615403	.079910	.668880
ДОМ_3	.586405	.065512	.626730
ХОБІ_1	.780488	.466823	.280141
ХОБІ_2	.734854	.464779	.238512
ХОБІ_3	.776013	.439010	.303672
ХОБІ_4	.714183	.455157	.228351

Уважне вивчення представленої таблиці дозволяє зробити такі висновки: 1) є загальний (вторинний) чинник задоволеності, до якого, мабуть, схильні всі типи задоволеності, виміряні для 10 пунктів; 2) є ймовірно дві первинні унікальні області задоволення, які можуть бути описані як задоволеність роботою, і задоволеність домашнім життям.

Слід зауважити, що обертання факторів зовсім не впливає на математичну строгість аналізу: взаємне положення зірочок-змінних не змінюється при повороті осей. Спочатку процедура обертання виконувалася вручну, і дослідник самостійно ставив таке положення осей, при якому точки змінних максимально до них наближалися. Тепер SPSS дозволяє виконати кілька варіантів обертання, повертаючи осі так, щоб отримати просту структуру, яка задовольняє той чи інший критерій. Найбільш популярним варіантом обертання є метод Varimax (Варимакс), про що вже йшлося вище.

Варіант обертання Varimax є ортогональним, оскільки при такому обертанні осі зберігають своєвзаємне розташування під прямим кутом. Іноді можна отримати більш зручну просту структуру, якщо змінити кут між осями. Для цього призначені варіанти обертання Direct Oblimin і Promax. Наявність непрямого кута між осями чинників означає, що вони не є повністю незалежними один від одного. Оскільки в реальних дослідженнях чинники дійсно не є абсолютно незалежними, відхилення кута між осями від прямого при обертанні є цілком припустимим.

Неортогональної обертання – досить складна процедура, і для її застосування необхідно чітко уявляти собі суть того, що відбувається. Аналогічну рекомендацію можна дати щодо факторного аналізу взагалі: не слід застосовувати його без попереднього вивчення відповідних розділів статистики.