

## §10. Факторен анализ

**1. Увод.** Психологическите изследвания като правило боравят с голям брой конструктори от типа на личностни характеристики, перцепции, нагласи и т.н., които обикновено се измерват посредством специфични стандартизирани тестови скали. Резултатите от измерванията се разглеждат в интервални метрични скали, което позволява пресмятане на средни, дисперсии, и корелационни коефициенти. Линейните трансформации на такива величини не променят корелационните коефициенти по между им а само центъра и мащаба на съответната скала.

Основната цел на факторния анализ е намаляване размерността на пространството от наблюденията без съществена (или малка) загуба на информация. Размерността  $p$  на пространството на наблюденията е равна на броя на наблюдаваните метрични величини  $X_1, X_2, \dots, X_p$ . Искаме да определим известен брой *латентни* (ненаблюдаеми) величини – фактори  $F_1, F_2, \dots, F_k, 1 \leq k \leq p$ , които да бъдат съществено по-малко на брой от първоначалните  $p$ , за които се предполага, че детерминират адекватно характера на наблюдаваните величини. Броят на тези фактори зависи съществено от корелациите между наблюдаваните величини, като в типичния случай на приложение на факторния анализ този брой е много по-малък, обикновено един, два или три фактора. Факторният анализ позволява за всеки от факторите да се пресмятат и факторни резултати, което фактически означава, че в крайна сметка с факторите можем да боравим както с обикновените наблюдавани величини. Факторните резултати се получават в  $z$ -стойности, със средно нула и стандартно отклонение единица.

По този начин факторният анализ представлява начин за идентифициране димензиите на латентното пространство, което обуславя резултатите от наблюденията. Съществуват различни техники на факторен анализ, които са сходни по между си, но имат и съществени различия. Тук ще разгледаме само най-разпространения вид факторен анализ по *метода на главните компоненти (principal components)*. При този вид факторен анализ факторите се получават с нулеви линейни корелации по между си, което дава възможност да ги интерпретираме като независими.

**2. Факторен анализ по метода на главните компоненти.** Броят  $k$  на факторите в този случай се определя от броят на собствените значения (*eigenvalues*) на корелационната матрица

$$(10.1) \begin{pmatrix} 1 & r(X_1, X_2) & \cdots & r(X_1, X_p) \\ r(X_2, X_1) & 1 & \cdots & r(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ r(X_p, X_1) & r(X_p, X_2) & \cdots & 1 \end{pmatrix},$$

които са по-големи от единица. Тези собствени значения са  $p$  на брой положителни числа със сбор, точно равен на  $p$ . Собствените значения показват приноса на съответния фактор при обясняване общата вариация в наблюдаваните величини.

Факторният модел се определя най-вече от факторните тегла (*loadings*)  $l_{ij}$ , които представляват *стандартизираните коефициенти на регресия* на наблюдаваните величини върху групата фактори,

$$X_1 \sim l_{11}F_1 + l_{12}F_2 + \cdots + l_{1k}F_k,$$

$$X_2 \sim l_{21}F_1 + l_{22}F_2 + \cdots + l_{2k}F_k,$$

...

$$X_p \sim l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pk}F_k.$$

Факторните тегла представляват същевременно и **корелационните коефициенти** между съответните наблюдавани величини и фактори, понеже при метода на главните компоненти факторите се избират с нулеви корелации (независими).

Сполучливият факторен модел изисква интерпретация на намерените фактори в рамките на предметната област. За улесняване на тази интерпретация се извършва ротация, обикновено по метода *varimax*. Факторната ротация има за цел увеличаване на дадени факторни тегла са сметка на други, като по този начин стремежът е всяка от наблюдаваните величини да намери обяснение с помощта на по възможност по-малък брой фактори.

Пресмятането на собствените значения, факторните тегла, извършването на ротация и намирането на факторните резултати представляват сложни изчислителни процедури, които се извършват с помощта на компютър, снабден с подходящ софтуер.

Да разгледаме следния пример. Група от  $n=175$  лица – обучаеми във ВУЗ се изследват по 10 на брой скали, свързани с техния емоционален статус. Тези скали имат условни наименования "активация", "враждебност", "вътрешен интерес", "безпокойство", "страх", "неувереност", "потиснатост", "дружелюбност", "самота", "стеснителност".

Таблица 10.1. Коефициенти на корелация

	{1}	{2}	{3}	{4}	{5}	{6}	{7}	{8}	{9}	{10}
активация {1}	1.00	-0.53	0.76	-0.29	-0.35	-0.50	-0.41	0.67	-0.37	-0.27
враждебност {2}	-0.53	1.00	-0.49	0.45	0.33	0.37	0.50	-0.51	0.46	0.35
вътрешен интерес {3}	0.76	-0.49	1.00	-0.15	-0.26	-0.34	-0.31	0.74	-0.29	-0.14
безпокойство {4}	-0.29	0.45	-0.15	1.00	0.56	0.66	0.71	-0.12	0.62	0.64
страх {5}	-0.35	0.33	-0.26	0.56	1.00	0.65	0.60	-0.23	0.51	0.54
неувереност {6}	-0.50	0.37	-0.34	0.66	0.65	1.00	0.71	-0.25	0.64	0.67
потиснатост {7}	-0.41	0.50	-0.31	0.71	0.60	0.71	1.00	-0.30	0.79	0.70
дружелюбност {8}	0.67	-0.51	0.74	-0.12	-0.23	-0.25	-0.30	1.00	-0.30	-0.13
самота {9}	-0.37	0.46	-0.29	0.62	0.51	0.64	0.79	-0.30	1.00	0.64
стеснителност {10}	-0.27	0.35	-0.14	0.64	0.54	0.67	0.70	-0.13	0.64	1.00

Таблица 10.1 показва, че корелациите между много от двойките величини имат големи абсолютни стойности, което предполага ефективно намаляване на дименсиите с помощта на факторен анализ.

Следващата таблица съдържа собствените значения.

Таблица 10.2.

	собствено значение	% от цялата вариация	натрупан процент
1	5.236255	52.36255	52.3626
2	1.974954	19.74954	72.1121
3	0.656328	6.56328	78.6754
4	0.481996	4.81996	83.4953
5	0.409913	4.09913	87.5945
6	0.341251	3.41251	91.0070
7	0.284106	2.84106	93.8480
8	0.224858	2.24858	96.0966
9	0.202064	2.02064	98.1173
10	0.188275	1.88275	100.0000

Тук са налице две собствени значения по-големи от 1,  $\lambda_1 = 5.236$  и  $\lambda_2 = 1.975$ , като следващото по големина  $\lambda_3 = 0.656$  е съществено по-малко от 1. Това показва, че трябва да изберем  $k = 2$  на брой фактора, т.е. размерността на латентното пространство в този случай е равна на 2, докато първоначално размерността на пространството на наблюденията е равна на 10, колкото е броят на скалите. Първите два фактора обясняват общо около 72% от общата вариация, което представлява много добър резултат.

Преди извършване на ротация имаме следните факторните тегла.

Таблица 10.3. Факторни тегла без ротация.

	фактор 1	фактор 2
активация	0.683	-0.565
враждебност	-0.674	0.297
вътрешен интерес	0.573	-0.707
безпокойство	-0.751	-0.396
страх	-0.715	-0.234
неувереност	-0.830	-0.226
потиснатост	-0.863	-0.258
дружелюбност	0.539	-0.710
самота	-0.805	-0.231
стеснителност	-0.735	-0.426

Посредством ротация по метода *varimax* получаваме следните факторните тегла.

Таблица 10.4. Факторни тегла след ротация.

	фактор 1	фактор 2
активация	-0.265	-0.846
враждебност	0.403	0.616
вътрешен интерес	-0.095	-0.905
безпокойство	0.846	0.077
страх	0.727	0.193
неувереност	0.819	0.263
потиснатост	0.864	0.255
дружелюбност	-0.065	-0.889
самота	0.801	0.245
стеснителност	0.848	0.043

Интерпретацията на тези два фактора ще извършим въз основа на факторните тегла, приведени в таблица 10.4. Първият фактор условно може да бъде наречен "**отрицателен емоционален статус**". Индикаторите на първия фактор са скалите "безпокойство", "страх", "неувереност", "потиснатост", "самота" и "стеснителност". Вторият фактор след обръщане на знака може условно да бъде наречен "**емоционален социобилитет**". Неговите индикатори са "активация", "вътрешен интерес" и "дружелюбност".

Скалата "враждебност" участва във формирането и на двата фактора, като не може да бъде отнесена строго към нито един от тях.

В по-нататъшните предметни статистически анализи можем да заменим всичките 10 наблюдавани величини с факторните резултати от определените току що само дава фактора.

Описаният подход води не само до икономия в технически план. Ако факторният модел е сполучлив, то самите фактори представляват своеобразен хармоничен синтез от своите индикатори, като по този начин определят количествено някакъв конструкт, който дава поглед "отгоре" върху анализа. Такива обобщаващи стратегии дават възможност за пълноценно разкриване естествените предимства на статистическия подход.

**3. Общи бележки.** Факторният анализ е раздел от статистиката, който е получил интензивно развитие именно поради нуждите на психологията. Възможността за неговото ефективно прилагане се крие дълбоко в основите на експерименталната психология. Самата идея за измерване на психичните конструкти има някои характерни особености. Преди всичко всяко измерване от тип суров бал от айтъмна скала разчита на повърхностен подход в добрия смисъл на думата. Изучаваният конструкт се измерва посредством неговите елементарни прояви. Такава философия изисква конструктът да бъде обхванат подробно и повърхностно от няколко аспекта, които обикновено са заложени като субскали в дадена основна скала. Резултатът от измерването несъмнено ще съдържа примеси от влияния на други психични конструкти. В такъв случай факторният анализ позволява да се извлече максимум полезна информация, която според механизма на самия анализ е заложена във водещия фактор – този с най-голямо собствено значение.

Понижаването на размерността се налага обективно от специфичния начин, по който дадена батерия от скали работи върху конкретна популация. Много добрите факторни модели предполагат високи стойности на коефициента на Кайзер-Майер-Олкин, на практика по-големи от 0.8.

Онова важно което малко се знае или се пренебрегва по технически и административни причини е фактът, че факторните резултати на водещия фактор представляват оптималните резултати от измерването при айтъмните скали, когато разбира се тези скали притежават някаква приемлива надеждност.

Използването на факторен анализ е свързано и с някои проблеми, които не възникнаха в разгледания пример. Добрият факторен модел изисква ясно определяне броя на факторите, както и ясна идентификация на всеки от факторите в рамките на предметната област. Когато има собствено значение по-голямо но близко до 1 възниква въпросът дали съответният фактор да бъде включен в анализа или не. В този случай е за препоръчване да се изследват и двата варианта и да се избере онзи с по-добра интерпретация.