ЛЕКЦІЯ 5. ФАКТОРНИЙ АНАЛІЗ

НАВЧАЛЬНІ ПИТАННЯ

Постановка завдання

Методи факторного аналізу

1. Постановка завдання

Нехай є набір вихідних ознак $Z_1, Z_2, \ldots, Z_n$, взагалі кажучи стандартизованих. Як відзначалось раніше, процедура стандартизації є необхідним етапом дослідження ознак, тому в даному випадку будемо вважати її вже проведеною. Необхідно замінити ці ознаки іншими $F_1, F_2, \ldots, F_k$, $k \le n$. Нові ознаки називають факторами. При цьому виходять з припущення, що початкові ознаки є результатом дії деяких спільних чинників, в ролі яких і будуть виступати нові фактори. З іншого боку, кожна з вихідних ознак має свою характерну рису, в ролі якої виступає певний характерний чинник. Тому загальна модель факторного аналізу матиме такий вигляд:

$$\begin{cases} Z_1 = w_{11}F_1 + w_{12}F_2 + \ldots + w_{1k}F_k + w_1U_1 \\ Z_2 = w_{21}F_1 + w_{22}F_2 + \ldots + w_{2k}F_k + w_2U_2 \\ \ldots \\ Z_n = w_{n1}F_1 + w_{n2}F_2 + \ldots + w_{nk}F_k + w_nU_n \end{cases}, \qquad (6.1)$$

Ознаки U_i відображають характерні риси вихідних ознак і називаються характерностями, у той час як фактори F_j відображають їх спільні риси і тому називаються ще спільностями. Величини w_{ij} називаються факторними навантаженнями . Вони показують частку загального фактора F_j у вихідній ознаці Z_i. За цими величинами можна визначити економічний зміст отриманих факторів. Варто відзначити, що значення факторних навантажень коливаються в межах від –1 до 1. Чим ближчі вони за модулем до 1, тим зв'язок між фактором та ознакою щільніший. Якщо величина факторного навантаження додатна, то вплив фактора на ознаку позитивний, інакше – негативний.

Система (6.1) в матричному вигляді записується таким чином

$$Z^{T} = WF^{T} + W_1U^{T}, \qquad (6.2)$$

де

Z – матриця реалізацій вихідних ознак розмірності ($m \times n$);

W – матриця факторних навантажень спільностей розмірності ($n \times k$);

F – матриця реалізацій факторів розмірності ($m \times k$);

W_1 – діагональна матриця факторних навантажень характерностей розмірності ($n \times n$);

U– матриця реалізацій характерностей розмірності ($m \times n$);

Кожне рівняння системи (6.1) можна подати у вигляді

$$\begin{bmatrix} z_{1i} \\ z_{2i} \\ \ldots \\ z_{mi} \end{bmatrix} = w_{i1}\begin{bmatrix} f_{11} \\ f_{21} \\ \ldots \\ f_{m1} \end{bmatrix} + w_{i2}\begin{bmatrix} f_{12} \\ f_{22} \\ \ldots \\ f_{m2} \end{bmatrix} + \ldots + w_{ik}\begin{bmatrix} f_{1k} \\ f_{2k} \\ \ldots \\ f_{mk} \end{bmatrix} + w_i\begin{bmatrix} u_{1i} \\ u_{2i} \\ \ldots \\ u_{mi} \end{bmatrix}. \qquad (6.3)$$

Залежність між компонентами ознак та факторів її можна записати таким чином:

$$z_{ji} = w_{i1}f_{j1} + w_{i2}f_{j2} + \ldots + w_{ik}f_{jk} + w_iu_{ji}, \qquad (6.4)$$

де

z_{ji} – j-те значення i-тої ознаки;

w_{is} – факторне навантаження s-того фактора;

f_{js} – j-те значення s- того фактора;

w_i – факторне навантаження характерності i-тої ознаки;

u_{ji} – j-те значення характерності i-тої ознаки;

$i=1..n$, $j=1..m$, $s=1..k$.

Слід зауважити, що одержані фактори будуються таким чином, щоб вони були взаємно некорельовані між собою та характерностями.

Таким чином, головним завданням факторного аналізу є пояснення кореляції вихідних ознак за допомогоюнекорельованих факторів. Як правило, кількість відібраних факторів менша за кількість вихідних ознак. Таким чином, за допомогою методів факторного аналізу фактично розв'язується два завдання: зменшення кількості ознак та виявлення латентних (скритих, неявних) ознак, які описують досліджуване явище. Латентні ознаки напряму не піддаються вимірюванню, тому їх кількісна оцінка за допомогою таких методів є досить важливою. З іншого боку, постає проблема економічної інтерпретації одержаних ознак.

Якщо характерності U_i відсутні, то фактори F_j називаються головними компонентами. Метод головних компонент відноситься до методів компонентного аналізу, але його ідея та порядок обчислень мають багато спільного з методами факторного аналізу. З іншого боку він вирізняється простотою . Тому для зручності викладення сутності методів факторного аналізу розглянемо саме випадок побудови головних компонент.

1. Методи факторного аналізу

Метод головних компонент

Якщо характерності U_i відсутні, то фактори F_j називаються головними компонентами. Для зручності викладення сутності методів факторного аналізу розглянемо саме випадок побудови головних компонент. Тоді система (6.2) в матричній формі матиме вигляд

$$Z^{\mathrm{T}} = WF^{\mathrm{T}}, \qquad (6.5)$$

Оскільки головною метою є пояснення кореляції вихідних ознак, то головну роль у факторному аналізі при розрахунку факторних навантажень відіграє кореляційна матриця R, побудована за значеннями вихідних ознак. Її елементи обчислюються за формулою

$$r_{ij} = \frac{1}{m}\sum_{s=1}^{m} z_{si} z_{sj}. \qquad (6.6)$$

Виразимо r_{ij} через факторні навантаження. Для цього скористаємось виразом (6.1), який відображає залежність між вихідними ознаками та факторами:

$$r_{ij} = \frac{1}{m} Z_i Z_j =$$
$$\frac{1}{m}[(w_{i1}F_1 + w_{i2}F_2 + \ldots w_{ik}F_k)(w_{j1}F_1 + w_{j2}F_2 + \ldots w_{jk}F_k)] =$$
$$= \frac{1}{m} m(w_{i1}w_{j1} + w_{i2}w_{j2} + \ldots w_{ik}w_{jk}) = \sum_{s=1}^{k} w_{is}w_{js}. \qquad (6.7)$$

Тут врахований той факт, що $F_iF_j = 0$, $F_iF_i = m$ (внаслідок ортогональності факторів).

Отже, $R = W \times W^T$.

Обчислимо коефіцієнт кореляції між Z_i та F_j.

$$r_{Z_iF_t} = \frac{1}{m}(w_{i1}F_1 + w_{i2}F_2 + \ldots w_{ik}F_k)F_j =$$
$$\frac{1}{m}\sum_{s=1}^{m}\sum_{t=1}^{m}(w_{i1}f_{t1} + w_{i2}f_{t2} + \ldots + w_{ik}f_{tk})f_{js} = = \frac{1}{m} m w_{ij} = w_{ij}. \qquad (6.8)$$

Отже, факторне навантаження дійсно виражає кореляцію між фактором та ознакою.

Виразимо фактори через вихідні ознаки:

$$F = ZB. \qquad (6.9)$$

Оскільки для факторів передбачається рівність дисперсій, то

$$D(F) = D(ZB) = B^{\mathrm{T}}D(Z)B = B^{\mathrm{T}}SB, \qquad (6.10)$$

де S – коваріаційна матриця. Тут врахована та властивість, що матриця коефіцієнтів B є сталою величиною, і тому виноситься за знак дисперсії у квадраті. Припустимо, що матриця B ортогональна, тобто

$$B^{\mathrm{T}}B = E. \qquad (6.11)$$

Оскільки метою факторного аналізу є максимальне пояснення дисперсії, максимізуємо (6.10), враховуючи обмеження (6.11). Для цього скористаємось методом множників Лагранжа:

$$f = B^{\mathrm{T}}SB - \lambda(B^{\mathrm{T}}B - E) \qquad (6.12)$$

$$\frac{\partial f}{\partial B} = 2SB - 2\lambda B = 0. \qquad (6.13)$$

Звідки остаточно маємо:

$$SB - \lambda B = 0. \qquad (6.14)$$

Отже, λ є власним значенням матриці S.

Оскільки за припущенням ці ознаки стандартизовані, то коваріаційна матриця співпадає з кореляційною матрицею, тобто, $S=R$.

З припущення (6.11) випливає, що (6.9) можна переписати у вигляді

$Z = F B^{\mathrm{T}}$, (6.15)

або, виконавши операцію транспонування до обох частин рівняння,

$Z^{\mathrm{T}} = BF^{\mathrm{T}}$. (6.16)

Останнє рівняння співпадає з (6.5). Отже, факторні навантаження є власними векторами коваріаційноїматриці вихідних ознак, тобто, $W=B$.

Дисперсія матриці F буде дорівнювати діагональній матриці Λ, яка складається з усіх власних значень. Тобто,

$$\sum_{i=1}^{n} w_{ij}^2 = \lambda_j = \lambda_j \sum_{i=1}^{n} v_{ij}^2 , \qquad (6.17)$$

де V – матриця нормованих власних векторів.

Звідси одержимо, що факторні навантаження можна обчислити за формулою

$$w_{ij} = \sqrt{\lambda_j} v_{ij} \qquad (6.18)$$

Оскільки матриця R дійсні, симетрична і додатно визначена, то всі її власні значення додатні. З (6.18) випливає, що

$W= V\times\Lambda^{1/2}$. (6.19)

Тоді

$W^{\mathrm{T}}\times W=(V\times\Lambda^{1/2})^{\mathrm{T}}\times(V\times\Lambda^{1/2}.) = \times\Lambda^{1/2}\times\Lambda^{1/2}\ V^{\mathrm{T}}\times V=\Lambda,$ (6.20)

Результат перетворення (6.20) можна використати для обчислення значень факторів F_i . Для цього помножимо (6.5) зліва спочатку на W^{T} , а потім на Λ^{-1}.

$W^{\mathrm{T}}Z^{\mathrm{T}}=W^T(WF) = (W^{\mathrm{T}}W)F^{\mathrm{T}}=\Lambda\ F^{\mathrm{T}}$

$\Lambda^{-1}\ W^{\mathrm{T}}Z^{\mathrm{T}} = F^{\mathrm{T}}$

$\Lambda^{-1}\ (ZW)^{\mathrm{T}} = F^{\mathrm{T}}$

Оскільки $\Lambda^{-1} =[\Lambda^{-1}]^T$, то

$(ZW\Lambda^{-1})^{\mathrm{T}} = F^{\mathrm{T}}$

$F = ZW\ \Lambda^{-1}$. (6.21)

Таким чином, алгоритм обчислень за методом головних компонент укладається в послідовність таких кроків:

1) Обчислення кореляційної матриці R стандартизованих ознак Z_i.
2) Обчислення матриці V нормованих власних векторів матриці R.
3) Обчислення матриці W факторних навантажень за формулою (6.18).
4) Обчислення матриці F значень нових факторів за формулою (6.21) та їх економічна інтерпретація.

Виходячи з аналізу алгоритму обчислення головних компонент можна зробити висновок, що їх кількість буде дорівнювати кількості вихідних ознак, оскільки матриця розмірності (n ´ n) має n різних власних векторів і власних значень. . Але на практиці відбираються не всі головні компоненти, а лише ті, що пояснюють найбільшу частку дисперсії. Для визначення таких компонент власні значення впорядковують, і відповідні їм власні вектори відбирають до тих пір, поки частка суми власних значень не сягне заданої величини δ:

$$v_k = \frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \ge \delta \ . \qquad (6.22)$$

Величина δ показує ступінь пояснення дисперсії вихідних ознак головними компонентами. Як правило, її обирають не меншою 80%.

Приклад. Завдання 6.1.

Нехай маємо дані для 8 ознак:

X_1 – трудомісткість одиниці продукції;

X_2 – питома вага робітників у складі ППП;

X_3 – питома вага придбаних виробів;

X_4 – коефіцієнт змінності обладнання;

X_5 – премії та заохочення одного робітника;

X_6 – питома вага втрат від браку;

X_7 – фондовіддача;

X_8 – середньорічна чисельність ППП;

Значення ознак наведені в табл. 6.1. Потрібно побудувати фактори за методом головних компонент та визначити ті, що пояснюють не менше 80% дисперсії вихідних ознак.

Таблиця 6.1

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45
2	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30
3	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37
4	12,85	0,17	0,70	0,50	1,42	0,43	0,18	1,65
5	10,63	0,23	0,62	0,40	1,35	0,88	0,15	1,91
6	9,12	0,42	0,76	0,19	1,39	0,57	0,34	0,17
7	25,83	0,31	0,73	0,25	1,16	1,82	0,38	1,94
8	23,39	0,26	0,71	0,44	1,27	1,70	0,09	1,89
9	14,68	0,49	0,69	0,17	1,16	0,84	0,14	1,94
10	10,05	0,36	0,73	0,39	1,25	0,60	0,21	2,06

Розв'язок.

Стандартизуємо ознаки. Результат занесемо до табл. 6.2.

Таблиця 6.2

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
1	-0,19	-0,58	1,41	0,53	0,35	0,26	-0,20	-0,21
2	-0,74	-0,48	0,76	-0,71	1,39	-0,10	1,13	-0,48
3	-0,11	-0,96	-0,76	0,53	0,96	1,35	1,46	-0,35
4	-0,27	-1,15	-0,33	1,42	0,78	-1,26	-0,61	0,15
5	-0,65	-0,58	-2,06	0,53	0,17	-0,40	-0,86	0,61
6	-0,92	1,25	0,98	-1,33	0,52	-0,99	0,71	-2,50
7	2,00	0,19	0,33	-0,80	-1,48	1,39	1,04	0,66
8	1,58	-0,29	-0,11	0,89	-0,52	1,16	-1,36	0,57
9	0,05	1,92	-0,54	-1,51	-1,48	-0,48	-0,94	0,66
10	-0,75	0,67	0,33	0,44	-0,70	-0,93	-0,36	0,88

Обчислимо кореляційну матрицю. Результат запишемо до табл. 6.3

Таблиця 6.3

1	-0,0402	-0,0003	0,0263	-0,5927	0,73995	-0,0789	0,43247
-0,0402	1	0,17894	-0,7876	-0,6206	-0,2863	-0,1336	-0,1248
-0,0003	0,17894	1	-0,2716	0,09653	-0,0195	0,33574	-0,4512
0,0263	-0,7876	-0,2716	1	0,3163	0,05262	-0,3525	0,36169
-0,5927	-0,6206	0,09653	0,3163	1	-0,1803	0,38268	-0,5583
0,73995	-0,2863	-0,0195	0,05262	-0,1803	1	0,31985	0,22334
-0,0789	-0,1336	0,33574	-0,3525	0,38268	0,31985	1	-0,4926
0,43247	-0,1248	-0,4512	0,36169	-0,5583	0,22334	-0,4926	1

За даними цієї матриці обчислимо власні вектори та власні значення. Для обчислень скористаємось методом обертань. Результат запишемо до табл. 6.4 В ній перший рядок являє собою власні вектори, а відповідні стовпці значень, починаючи з другого рядка – нормовані власні вектори. Для зручності власні значення розташовані у порядку спадання їх значень.

Таблиця 6.4

Власні значення

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
2,6081	2,2582	1,7833	0,7417	0,3418	0,1400	0,1086	0,0183

Власні вектори

V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8
0,4552	0,1326	0,4188	0,1747	-0,3252	0,4865	0,3365	0,3371
-0,035	0,6235	-0,2252	-0,0568	-0,0769	-0,167	-0,3533	0,6305
-0,287	0,2161	0,2755	0,83663	0,2564	-0,1597	0,0459	-0,0411
0,1920	-0,5797	-0,0872	0,332	-0,0201	0,1746	-0,637	0,2696
-0,4331	-0,4507	0,0710	-0,0635	-0,1294	-0,2817	0,4092	0,5802
0,2827	-0,0309	0,6244	-0,1548	-0,1849	-0,6334	-0,2531	-0,0767
-0,3333	0,0298	0,5267	-0,3576	0,5079	0,3861	-0,2375	0,1391
0,5419	-0,077	-0,1346	-0,0149	0,7165	-0,2231	0,2575	0,2294

Відбір власних значень для пояснення дисперсії

Сума	2,60812	4,86631	6,64964	7,39134	7,73314	7,87315	7,9817	8
Частка	33%	61%	83%	92%	97%	98%	100%	100%

З аналізу табл. 6.4 випливає, що для заданого ступеня пояснення дисперсії достатньо залишити перших три головних компоненти. Вони пояснюють 83% дисперсії вихідних ознак. Обчислимо факторні навантаження перших трьох головних компонент за формулою (6.18). Результат запишемо до табл. 6.5 Значення головних компонент. Обчислені за формулою (6.22) занесемо до табл. 6.6.

Факторні навантаження дозволяють визначити зв'язок головних факторів і вихідних ознак. Так, з табл. 6.5 видно , що перша головна компонента має позитивний зв'язок з першою, четвертою шостою та восьмою ознакою, і негативний з рештою.

Таблиця 6.5

0,7352	0,1993	0,5593
-0,0565	0,9369	-0,3008
-0,4635	0,3248	0,3679
0,3101	-0,8711	-0,1165
-0,6994	-0,6773	0,0948
0,4565	-0,0465	0,8338
-0,5383	0,0448	0,7033
0,8752	-0,1157	-0,1797

Таблиця 6.6

F_1	F_2	F_3
-0,3064	-0,3613	0,3784
-1,1987	-0,2501	0,5741
-0,2518	-0,9911	1,2268
-0,0778	-1,3242	-0,8513
0,5239	-0,8913	-1,1477
-1,9152	1,0971	-0,1139
1,0543	1,0121	1,6296
1,3908	-0,2617	0,3840
0,6211	1,7059	-1,0621
0,1598	0,2645	-1,0179

Метод головних факторів

При знаходженні факторів за методом головних факторів передбачається , що характерності вже відомі. Це означає, що в кореляційній матриці R діагональні елементи виражають лише спільності, а тому менші одиниці.

Метод головних факторів полягає в знаходженні такої матриці факторних навантажень W, яка задовольняє співвідношенню

$$WW^T = R - H_1^2 = R^{¢}, \qquad (6.23)$$

де H – діагональна матриця, що виражає характерність,

$R^{¢}$ – редукована матриця кореляцій.

Відмітимо, що завдання знаходження матриці W не є однозначним, оскільки матриця факторних навантажень виділяє в просторі Z підпростір F загальних факторів, але не фіксує в ньому систему координат. Тобто, матриця факторних навантажень знаходиться з точністю до ортогонального перетворення. Це дозволяє при її обчисленні накладати додаткові вимоги на неї. Одним з них є те, що загальні фактори обчислюються за впорядкованістю їх внесків у пояснення дисперсії вихідних ознак. Якщо при цьому виявляється, що серед знайдених лише декілька мають великі внески, то для практичного використання можна обмежитись лише ними.

На практиці редукцію вихідної кореляційної матриці можна провести декількома способами:

- *методом найбільшої кореляції* – діагональний елемент замінюється на найбільше по відповідному рядку (стовпчику) недіагональне значення коефіцієнта кореляції;

- *методом Барта* – для кожного рядка спочатку знаходиться середнє значення коефіцієнта кореляції; якщо воно порівняно велике, то діагональний елемент замінюється трохи більшим за найбільше по рядку значення коефіцієнта кореляції, а якщо воно порівняно мале, то діагональний елемент замінюється трохи меншим за найбільше по рядку значення коефіцієнта кореляції;

- *методом тріад* – діагональний елемент обчислюється за формулою

$$h_i^2 = \frac{r_{ik} r_{is}}{r_{ks}}, \qquad (6.24)$$

де r_{ik} та r_{is} – найбільші по рядку значення коефіцієнта кореляції.

Метод головних факторів будується як метод розв'язання впорядкованого ланцюга пов'язаних одна з одною екстремальних задач. На першому кроці розглядається завдання максимізації критерія

$$V_1 = \sum_{i=1}^{n} w_{i1}^2 = \sum_{i=1}^{n} r_{z_i F_1}^2, \qquad (6.25)$$

як функціонала, що залежить від факторних навантажень першого фактора. Нехай в результаті обчислень одержаний шуканий вектор факторних навантажень $W_1 = \{w_{11}, w_{21}, \ldots, w_{n1}\}^T$. Далі обчислюється матриця залишкових кореляцій

$$R_1 = R¢ - W_1 W_1^T. \qquad (6.26)$$

В цій матриці виключений вплив першого фактора на вихідні ознаки. Для матриці залишкових кореляцій процедура обчислень повторюється. Обчислювальний процес завершується тоді, коли відібрана достатня кількість факторів для пояснення дисперсії вихідних ознак. Ступінь достатності можна визначити за значущістю матриці залишкових кореляцій: якщо вона значуща, то виділення факторів продовжується.

З обчислювальної точки зору метод головних факторів можна розглядати як модифікацію методу головних компонент. Відмінності полягають в тому, що:

по-перше, обчислення базуються не на кореляційній матриці вихідних ознак, а на редукованій;

по-друге, фактори виявляються один за одним, за зменшенням частки пояснюваної ними дисперсії вихідних ознак.

Якщо λ_j – власне значення кореляційної матриці, а V_j – відповідний йому нормований власний вектор, то факторні навантаження фактора F_j обчислюються за формулою

$$w_{ij} = \sqrt{\lambda_j}\, v_{ij}. \qquad (6.27)$$

При використанні методів факторного аналізу виникає питання, наскільки щільні кореляційні зв'язки між вихідними ознаками. Тобто, наскільки значуща кореляційна матриця вихідних ознак. Адже незначущість такої матриці може означати, що для вихідної сукупності ознак взагалі не існує спільностей, і застосування методів факторного аналізу не є виправданим. В такому випадку виділені фактори фактично будуть являти собою вихідні ознаки.

Оцінка значущості кореляційної матриці може проводитись за критерієм Уілкса χ^2 .

Спостережене значення критерію обчислюється за формулою

$$\chi^2_{спост} = -\left(m - \frac{1}{6}(2n+5)\ln[\det(R)]\right). \qquad (6.28)$$

Це значення порівнюється з критичним для χ^2 - розподілу, знайденим при заданому рівні значущості α та кількості ступенів вільності $\nu = \frac{n(n-1)}{2}$. Значущість кореляційної матриці підтверджується, якщо спостережене значення перевищує критичне: $\chi^2_{спост} > \chi^2_{kp}(\alpha, \nu)$.

Отже, алгоритм обчислень за методом головних факторів має наступний вигляд:

1) обчислення матриці стандартизованих ознак : $X \to Z$;
2) обчислення кореляційної матриці стандартизованих ознак: $Z \to R$;
3) обчислення редукованої кореляційної матриці: $R \to R$¢ ;
4) обчислення першого власного значення λ_1 та відповідно йому нормованого власного вектора V_1редукованої кореляційної матриці: R¢$\to(\lambda_1, V_1)$;
5) обчислення вектора факторних навантажень першого фактора F_1 : $(\lambda_1, V_1) \to W_1$;
6) обчислення матриці залишкових кореляцій : (R¢, W_1) $\to R_1$;
7) перевірка значущості матриці залишкових кореляцій. Якщо вона значуща, то ітераційний процес продовжується.

Інтерпретація одержаних факторів здійснюється наведеному розглянутому підходу при розгляді методу головних компонент.

Ступінь достатності можна визначати різними способами. Одним з них є метод кам’янистого осипу. Сутність методу полягає у графічному зображенні обчислених власних значень на площині, для якої вісь X відображаєномери значень за порядком їх обчислення, а вісь Y – їх величини. Одержані таким чином точки з’єднуються ламаною лінією (рисунок 6.1). Візуально ламана «згладжується» на деяких ділянках прямою лінією. Те значення, для якого згладжена пряма має перший очевидний злам, вважається останнім достатнім виділеним фактором. Так, для рисунка 6.1 можна зробити висновок про достатність трьох виділених факторів.

Вважається, що кількість залишених факторів повинна бути щонайменше вдвічі меншою за кількість вихідних ознак. Тоді одержані фактори допускають гарну змістовну інтерпретацію.

Разом з тим метод кам’яного осипу не завжди дає однозначний результат. Так для даних, зображених на рисунку 6.2 можна з однаковою впевненістю залишити для аналізу як три, так і чотири фактори.

Більш надійним способом визначення достатності виділених факторів є перевірка статистичних гіпотез про значущість матриці залишкових кореляцій.

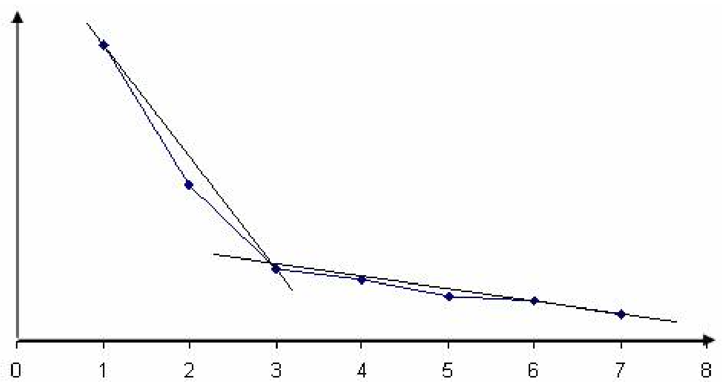

Рисунок 6.1 Зображення зміни власних значень і згладжування ламаної, що з’єднує їх величини.

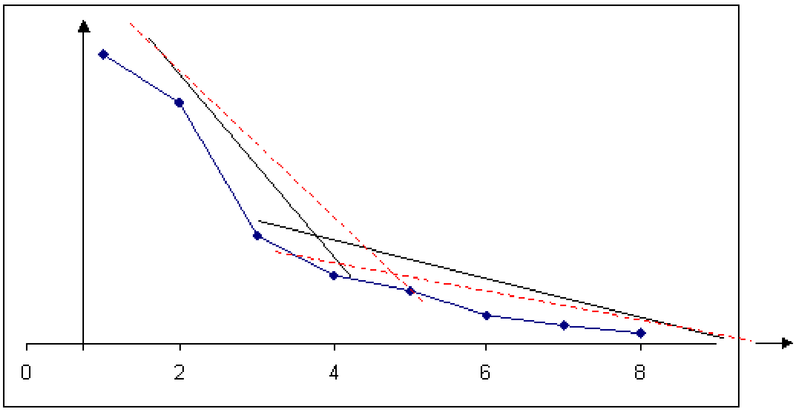

Рисунок 6.2 Неоднозначність визначення зламу згладженої лінії.

Достатність виділених факторів може бути перевірена за критерієм Лоулі χ^2 . Його сутність полягає у перевірці, наскільки матриця $R^{+} = WW^{T}$, яка обчислюється за факторними навантаженнями всіх виділених факторів, відрізняється від вихідної матриці. Нульова гіпотеза полягає в тому, що відтворена матриця не відрізняється від початкової редукованої матриці кореляцій: H_0:R^{+}=R'.

Спостережене значення критерію обчислюється за формулою

$$\chi^2_{\text{спост}} = (m-1)\ln\left(\frac{\det(R^{+})}{\det(R')}\right). \qquad (6.29)$$

Критичне значення знаходиться при заданому рівні значущості α та кількості ступенів вільності $\nu = ((n-k)^2 - n - k)/2 - k$, де k – кількість виділених головних факторів. Нульова гіпотеза приймається, якщо спостережене значення критерію не перевищує критичне: $\chi^2_{\text{спост}} \le \chi^2_{kp}(\alpha,\nu)$. Даний критерій можна застосовувати для великих вибірок, коли кількість спостережень щонайменше на 51 більше за кількість вихідних ознак.

2. Питання для самоперевірки.

1. Поясніть призначення методів факторного аналізу. Який вид діагностичний ознак можна побудувати такими методами?
2. В чому полягає сутність методу головних компонент?
3. Що являє собою матриця факторних навантажень?
4. Поясніть алгоритм обчислень за методом головних компонент.
5. Поясніть алгоритм обчислень за методом головних факторів.
6. Охарактеризуйте основні відмінності між методами головних компонент і головних факторів.