

АРМ „БАГАТОВИМІРНИЙ СТАТИСТИЧНИЙ АНАЛІЗ”

доц.,к.ф.м.н. Корольов М.Є.-каф.ПМІ

Моделювання та дослідження систем, як правило, потребують експериментальної роботи. З одного боку, експеримент дозволяє перевірити модель й, якщо треба, уточнити її. З іншого боку, модель підказує, який саме експеримент треба проводити, тобто дає інформацію для організації експерименту.

Експериментальна робота пов'язана з вимірюванням різних характеристик системи. Слід зауважити, що сучасне поняття вимірювань значно ширше, ніж класичне, де розглядалися тільки кількісні та однозначні виміри. Його особливостями є:

- 1) виміри якісних характеристик;
- 2) сприймання того факту, що вимір може не знімати невизначеності, якщо вона має розпливчасту природу;
- 3) урахування того, що виміри супроводжуються неминучими похибками;
- 4) величина, що цікавить експериментатора, часто є неспостережною, і можна спостерігати лише деяку функцію від неї.

Вимірювання - це операція, за якою явище, що спостерігається, ставиться у відповідність до одного з елементів певної вимірювальної шкали, необов'язково числової. Шкала може бути порядковою, якщо її елементи порівнянні між собою, або номінальною, якщо будь-які порівняння неможливі. Вимірювальна та експериментальна робота відіграє значну роль у виробничому менеджменті при створенні нових товарів, а також при підвищенні якості товарів, у маркетингових дослідженнях та розробці маркетингової стратегії.

Багатовимірний статистичний аналіз - логічний розвиток методів традиційної статистики, узагальнених у курсі загальної теорії статистики. Принципова відмінність полягає в тім, що об'єкти, соціальні, економічні явища розглядаються з обліком не одного або двох, а одночасно деякої безлічі ознак. Це дозволяє домогтися в дослідженнях повноти теоретичного опису спостережуваних об'єктів й об'єктивності наступних висновків. Тому спільне вивчення значень ознак дозволяє моделювати образ суб'єкта й реально оцінювати його поведінкову реактивність.

До АРМ „багатовимірний статистичний аналіз” входять чотири задачі: метод головних факторів (Алгоритм Хотеллінга), неметричні методи багатовимірного шкалювання, класична модель багатовимірного шкалювання Торгерсона, узагальнена модель пошуку індивідуальних розходжень Такера.

При створенні тестів для даних моделей використовувались електронні таблиці MS Excel, середовище програмування Visual Studio 2010. Головне вікно програми зображене на рисунку :

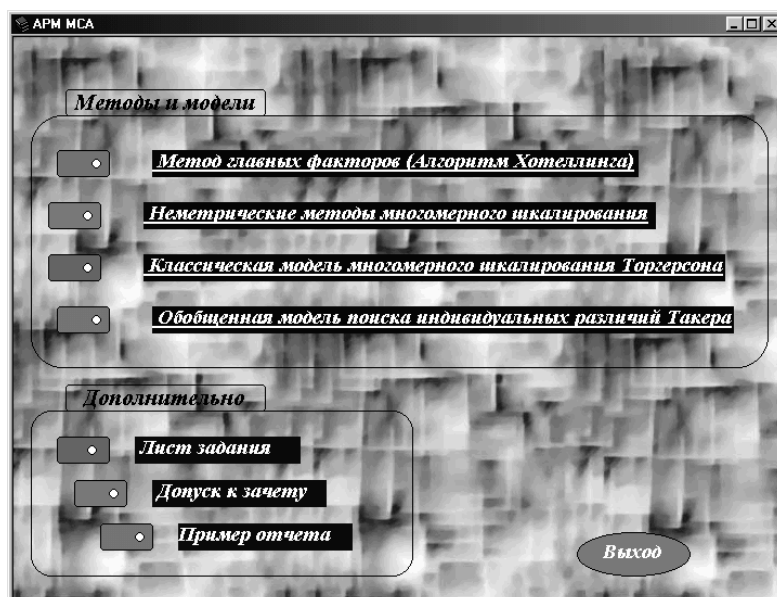


Рисунок - Головне вікно АРМ „Багатовимірний статистичний аналіз”

Метод головних факторів (Алгоритм Хотеллінга)

Постановка задачі має наступний вигляд.

Дана кореляційна матриця R_h :

$$R_h = \begin{pmatrix} 0,687 & 0,58 & 0,69 & 0,25 & 0,211 \\ 0,58 & 0,14 & 0,439 & 0,66 & 0,4 \\ 0,69 & 0,439 & 0,42 & 0,1 & 0,621 \\ 0,25 & 0,66 & 0,1 & 0,97 & 0,9 \\ 0,211 & 0,4 & 0,621 & 0,9 & 0,68 \end{pmatrix}$$

Необхідно визначити факторні навантаження й геометрично представити спостережувані об'єкти в тривіальному просторі латентних (схованих) факторів.

Метод головних факторів використовується для рішення інших трьох завдань, тому його розуміння дуже важливе. Нижче наведемо приклад реалізації цього методу й покажемо як за допомогою тесту перевіряється результат.

Початок рішення - знаходження суми рядків матриці й відношень цих сум до максимальної суми. Заповнюється таблиця 1.

Таблиця 1 - Кореляційна матриця в першому ступені

Ознака	X1	X2	X3	X4	X5	$S_i^{(1)}$	$\alpha^{(1)}$
X1	0,61	0,48	0,51	0,15	0,11	1,86	0,53295
X2	0,48	0,91	0,39	0,66	0,4	2,84	0,81375
X3	0,51	0,39	0,58	0,81	0,21	2,5	0,71633
X4	0,15	0,66	0,81	0,97	0,9	3,49	1
X5	0,11	0,4	0,21	0,9	0,68	2,3	0,65903

Де $S_i^{(k)}$ – сума і-го рядка кореляційної матриці в k-ому ступені, тобто

$$S_i^{(k)} = \sum_j r_{ij} ; \quad (1)$$

$\alpha^{(1)}$ – відношення $S_i^{(1)}$ до S_{\max} ,

$$\alpha^{(1)} = \frac{S_i^{(1)}}{S_{\max}} . \quad (2)$$

Перший цикл ітерації – зведення у квадрат кореляційної матриці – таблиця 2.

Таблиця 2 – Зведення у квадрат кореляційної матриці

Ознака	$R_h^2 = R_h R_h$					$S_i^{(2)}$	$P_i^{(2)}$	$\alpha^{(2)}$	$d^{(2)}$
	X1	X2	X3	X4	X5				
X1	0,8972	1,0715	0,9387	1,0659	0,576	4,5493	4,5493	0,472	0,061
X2	1,0715	1,8062	1,4445	1,9887	1,3647	7,6756	7,6756	0,797	0,017
X3	0,9387	1,4445	1,4488	1,7784	1,2057	6,8161	6,8161	0,708	0,009
X4	1,0659	1,9887	1,7784	2,8651	1,9356	9,6337	9,6337	1,000	0,000
X5	0,576	1,3647	1,2057	1,9356	1,4886	6,5706	6,5706	0,682	0,023

Де $P_i^{(k)}$ – добуток двох матриць попереднього циклу R_h й $S_i^{(k-1)}$;

$\alpha^{(2)}$ – відношення $P_i^{(2)}$ до P_{\max} , тобто ,

$$\alpha_i^{(2)} = \frac{P_i^{(2)}}{P_{\max}} \quad (3)$$

$d^{(k)}$ – модуль різниці $\alpha^{(k)}$ і $\alpha^{(k-1)}$, тобто

$$d_i = |\alpha_i^{(2)} - \alpha_i^{(1)}| \quad (4)$$

Другий цикл ітерації - кореляційна матриця в четвертому ступені - таблиця 3.

Таблиця 3 - Кореляційна матриця в четвертому ступені

Ознака	$R_h^4 = R_h^2 R_h^2$					$S_i^{(3)}$	$P_i^{(3)}$	$\alpha^{(3)}$	$d^{(3)}$
	X1	X2	X3	X4	X5				
X1	4,30216	7,15847	6,34005	8,92542	6,03144	32,7575	28,9729	0,484	0,012
X2	7,15847	12,3144	10,8898	15,6423	10,7046	56,7095	47,7426	0,798	0,001
X3	6,34005	10,8898	9,68318	13,8788	9,49589	50,2877	42,3656	0,708	0,000
X4	8,92542	15,6423	13,8788	20,2091	13,8992	72,5549	59,8368	1,000	0,000
X5	6,03144	10,7046	9,49589	13,8992	9,61037	49,7414	59,8368	1,000	0,318

Третій цикл ітерації: кореляційна матриця у восьмому ступені – таблиця 4.

Таблиця 3.4 - Кореляційна матриця у восьмому ступені

Ознака	$R_h^8 = R_h^4 R_h^4$					$S_i^{(4)}$	$P_i^{(4)}$	$\alpha^{(4)}$	$d^{(4)}$
	X1	X2	X3	X4	X5				
X1	225,98	392,16	347,72	502,57	344,80	1813,30	1813,30	0,449	0,035
X2	392,16	680,74	603,66	872,55	598,69	3147,84	3147,84	0,780	0,018
X3	347,77	603,68	535,34	773,78	530,92	2791,50	2791,50	0,692	0,016
X4	502,57	872,55	773,78	1118,5	767,53	4035,01	4035,01	1,000	0,000
X5	344,80	598,69	530,9	767,53	526,68	2768,6	2768,64	0,686	0,314

Оцінки S й P підтверджують правильності проведених обчислень таким чином, оцінки компонентів першого власного вектора можна вважати достовірними.

Четвертий цикл ітерації - кореляційна матриця в шістнадцятому ступені – таблиця 5.

Таблиця 5 - Кореляційна матриця в шістнадцятому ступені

Ознака	$R_h^{16} = R_h^8 R_h^8$					$S_i^{(5)}$	$P_i^{(5)}$	$\alpha^{(5)}$	$d^{(5)}$
	X1	X2	X3	X4	X5				
X1	697279,9	1210491	1073460	1551673	1064695	5597599	5597599	0,449	0
X2	1210491	2101434	1863546	2693734	1848330	9717535	9717535	0,780	0
X3	1073460	1863546	1652588	2388796	1639094	8617484	8617484	0,692	0
X4	1551673	2693734	2388796	3452975	2369291	12456469	12456469	1,000	0
X5	1064695	1848330	1639094	2369291	1625711	8547121	8547121	0,686	0

П'ятий цикл ітерації - кореляційна матриця в тридцять другий ступені – таблиця 6.

Таблиця 6 - Кореляційна матриця в тридцять другий ступені

Ознака	$R_h^{32} = R_h^{16} R_h^{16}$					$S_i^{(6)}$	$P_i^{(6)}$	$\alpha^{(6)}$	$d^{(6)}$
	X1	X2	X3	X4	X5				
X1	6,65E+12	1,15E+13	1,02E+13	1,48E+13	1,01E+13	5,33E+13	5,33E+13	0,449	0
X2	1,15E+13	2E+13	1,78E+13	2,57E+13	1,76E+13	9,26E+13	9,26E+13	0,780	0
X3	1,02E+13	1,78E+13	1,57E+13	2,28E+13	1,56E+13	8,21E+13	8,21E+13	0,692	0
X4	1,48E+13	2,57E+13	2,28E+13	3,29E+13	2,26E+13	1,19E+14	1,19E+14	1,000	0
X5	1,01E+13	1,76E+13	1,56E+13	2,26E+13	1,55E+13	8,15E+13	8,15E+13	0,686	0

Власний вектор - це ненормовані значення $\alpha_i^{(5)}$.

Перейдемо до визначення навантажень першого головного фактора – таблиця 7.

Таблиця 7 - Обчислення першого головного фактора

Ознака	$\alpha_i^{(5)} = U_i$	$\beta_1 = R_h^* \alpha_i^{(5)}$	A
X1	0,449	1,227	0,447
X2	0,780	2,130	0,776
X3	0,692	1,889	0,688
X4	1,000	2,730	0,995
X5	0,686	1,873	0,682

Де A - вектор факторних навантажень, що перебуває по формулі

$$A = \frac{u_1 \sqrt{\lambda_1}}{\left(\sum_i u_{1i}^2 \right)^{1/2}} \quad (5)$$

Власне число дорівнює: $\lambda = \max \beta_1 = 2,73$

Визначаємо матрицю парних кореляцій R_h^+ і вирішити питання про необхідність виконання другої ітерації з пошуком другого власного числа λ_2 і вектори факторних навантажень A_2 .

$$R_h^+ = A \cdot A' \quad (6)$$

$$R_h^+ = \begin{pmatrix} 0,4469 \\ 0,7759 \\ 0,688 \\ 0,9945 \\ 0,6824 \end{pmatrix} \cdot \begin{pmatrix} 0,4469 & 0,7759 & 0,688 & 0,9945 & 0,6824 \end{pmatrix} =$$

$$= \begin{pmatrix} 0,200 & 0,347 & 0,308 & 0,445 & 0,305 \\ 0,347 & 0,602 & 0,534 & 0,772 & 0,530 \\ 0,308 & 0,534 & 0,473 & 0,684 & 0,470 \\ 0,445 & 0,772 & 0,684 & 0,989 & 0,679 \\ 0,305 & 0,530 & 0,470 & 0,679 & 0,466 \end{pmatrix}$$

Різниця матриць R_h й R_h^+ покаже залишкову, не пояснену першим головним фактором, кореляцію й допоможе відповісти на запитання про доцільність виділення другого головного фактора.

$$R_1 = R_h - R_h^+ = \begin{pmatrix} 0,61 & 0,48 & 0,51 & 0,15 & 0,11 \\ 0,48 & 0,91 & 0,39 & 0,66 & 0,4 \\ 0,51 & 0,39 & 0,58 & 0,81 & 0,21 \\ 0,15 & 0,66 & 0,81 & 0,97 & 0,9 \\ 0,11 & 0,4 & & 0,9 & 0,68 \end{pmatrix} -$$

$$= \begin{pmatrix} 0,200 & 0,347 & 0,308 & 0,445 & 0,305 \\ 0,347 & 0,602 & 0,534 & 0,772 & 0,530 \\ 0,308 & 0,534 & 0,473 & 0,684 & 0,470 \\ 0,445 & 0,772 & 0,684 & 0,989 & 0,679 \\ 0,305 & 0,530 & 0,470 & 0,679 & 0,466 \end{pmatrix}$$

$$= \begin{pmatrix} 0,410 & 0,133 & 0,202 & -0,295 & -0,195 \\ 0,133 & 0,308 & -0,144 & -0,112 & -0,130 \\ 0,202 & -0,144 & 0,107 & 0,126 & -0,260 \\ -0,295 & -0,112 & 0,126 & -0,019 & 0,221 \\ -0,195 & -0,130 & -0,260 & 0,221 & 0,214 \end{pmatrix}$$

Матриця перших залишкових коефіцієнтів кореляції містить ще досить великі величини й цілком допускає оцінку другого головного фактора. Наступне виконання другої ітерації аналогічно першій, тільки обчислення виробляються на дані матриці залишків R_1 . Її рішення аналогічно першій ітерації, тому ми наводити його не будемо.

Випишемо тільки факторні напрямки й надамо візуалізацію результату в просторі латентних факторів (таблиця 8, рисунок 2).

Таблиця 8 - Підсумкова таблиця

Головний фактор (факторні навантаження)		
	F1	F2
X1	0,625	0,602
X2	0,629	0,083
X3	0,603	0,441
X4	0,859	-0,561
X5	0,816	-0,263

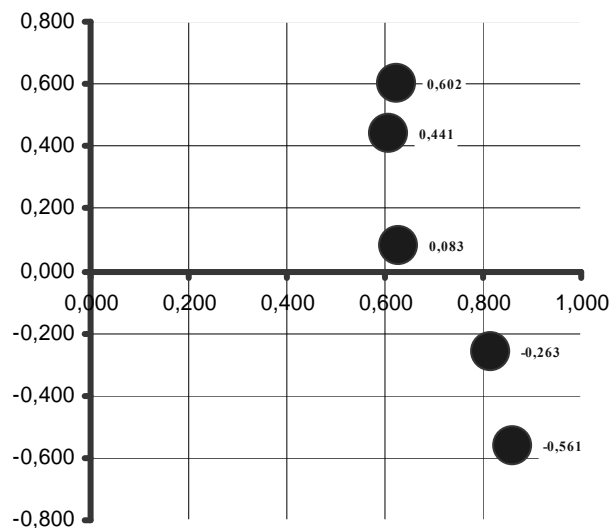


Рисунок 2 - Геометричне подання п спостережуваних об'єктів у тривіальному просторі латентних факторів.

Перевірка результатів відбувається за допомогою наступного діалогового вікна тесту – рисунок 3.

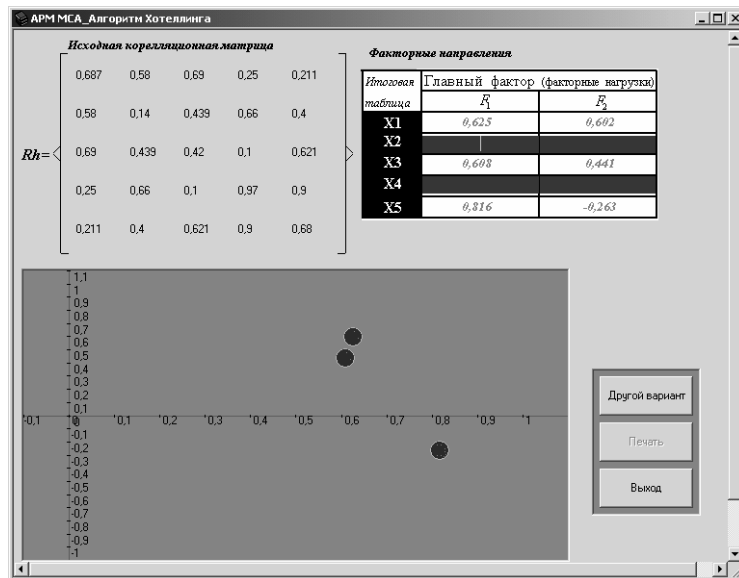


Рисунок 3 – Вікно перевірки результатів реалізації метода головних факторів

В верхньому лівому куті рисунку 3.3 знаходиться початкова кореляційна матриця R_h . Перевірка факторних навантажень відбувається зразу при вводі значень в таблиці. В низу на графіку геометрично відображаються спостережувані об'єкти в тривіальному просторі латентних (схованих) факторів. Програмний код тесту наведено в додатку А 3.2.

Неметричні методи багатовимірного шкалювання

Задача має наступну постановку.

Експертам пропонується шкала з деяким числом розподілів (10), що дозволяють оцінювати кожену пару об'єктів по ступені їхньої подібності (рисунок 3.4).

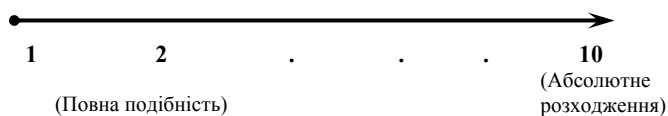


Рисунок 4 – Шкала

За результатами експертного оцінювання одержують $n(n-1)/2$ пар об'єктів, упорядкованих по рангових характеристиках подібностей, у результаті чого одержується матриця розходжень, що містить рангові дані - характеристики несхожості аналізованих об'єктів.

Необхідно досліджувати дану модель неметричними методами багатовимірного шкалювання.[11]

Реалізація даної задачі в АРМ відбувається по кроках (1 - 6), які представлені на схемі рисунку 5.

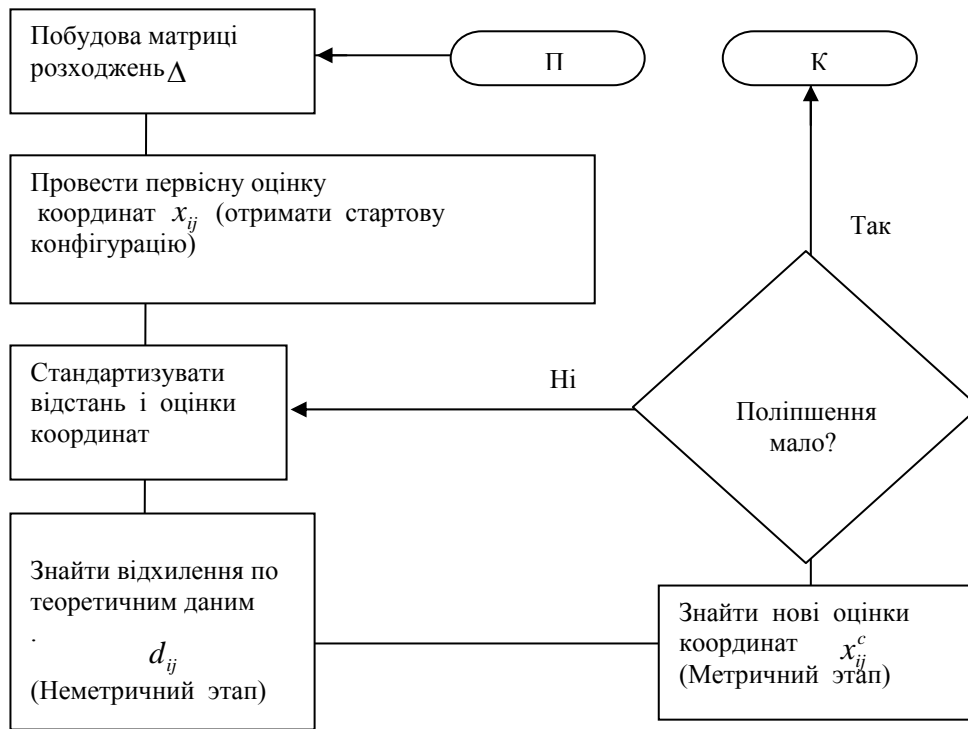


Рисунок 5 - Схема алгоритму неметричного багатовимірного шкалювання

Головне вікно тесту зображено на рисунку 6.

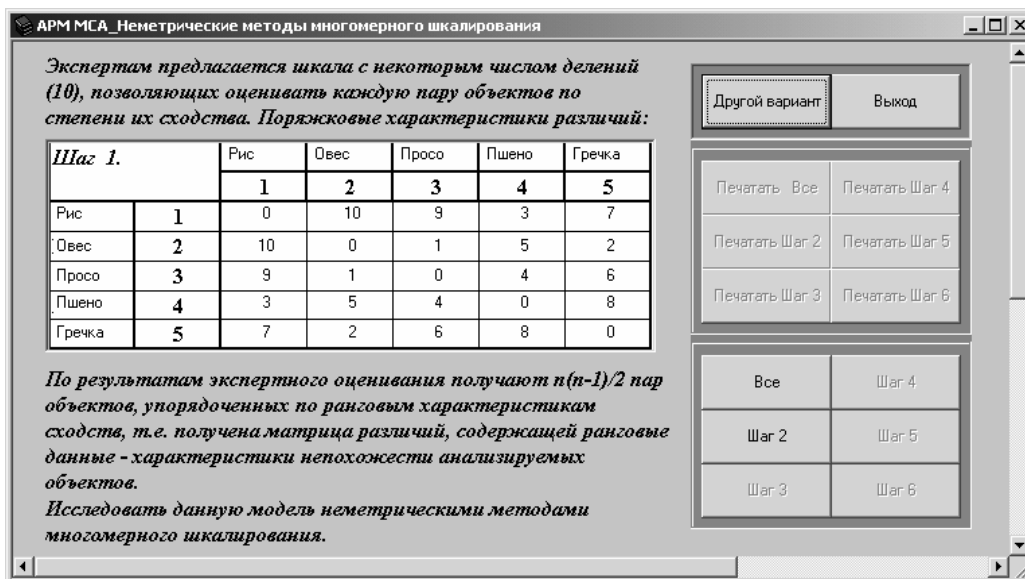


Рисунок 6 – Головне діалогове вікно тесту неметричних методів багатовимірного шкалювання

Перший крок алгоритму дуже кропіткий, тому початкові дані даються одразу. Прикладом таких даних служить таблиця 9 - характеристики несхожості аналізованих об'єктів.

Таблиця 9 – Порядкові характеристики розходжень

		A	B	C	D	E
		1	2	3	4	5
A	1	0	10	9	3	7
B	2	10	0	1	5	2
C	3	9	1	0	4	6
D	4	3	5	4	0	8
E	5	7	2	6	8	0

Крок 2 - пошук стартової конфігурації. Перші приблизні оцінки координат обчислюються методом головних факторів. Результатом чого є нестандартизовані характеристики по першим двом координатним осям - таблиця 10.

Таблиця 10 - Підсумкова таблиця

	Головний фактор (факторні навантаження)	
	F ₁	F ₂
X ₁	2,567946	-2,49988
X ₂	1,84871	1,930278
X ₃	2,030293	1,337738
X ₄	1,89213	-1,11952
X ₅	2,180663	1,034308

Вікно в якому перевіряється реалізація метода головних факторів зображене на рисунку 7.

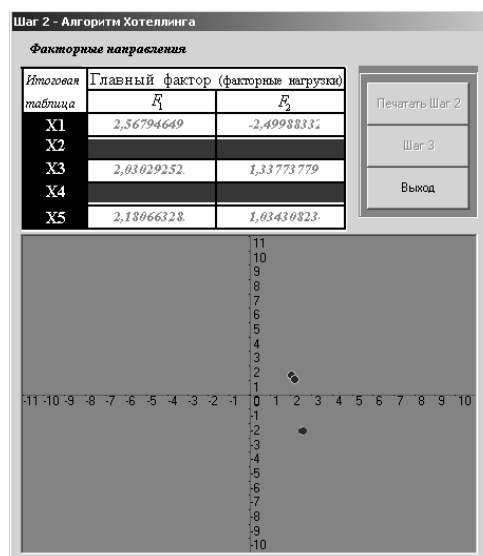


Рисунок 7 – Вікно перевірки знайдених нестандартизованих характеристик по першим двом координатним осям

Перевірка рішення перебуває по кроках. Програма дозволяє друкувати як окремі кроки, так і рішення задачі в цілому.

Крок 3 - стандартизація оцінок координат і відстаней. Стандартизація проводиться з метою зберегти пропорції ортонормированого стимулювання простору і для запобігання вироджених рішень, коли простір стимулів стискається до розмірів крапки й аналіз не дає скільки-небудь значимих результатів. Наприклад, коли кілька стимулів одержують однакові оцінки координат або їхніх координатних оцінок близькі нулю, тобто розташовані поблизу початки системи координат. Стимули в такому теоретичному просторі шкал як би зливаються й стають нерозрізнені для дослідника.[11]

З великої кількості метричних формул вирішили використовувати евклідову відстань

$$\delta_{ij}^E = \left(\sum_k (v_{ik} - v_{jk})^2 \right)^{1/2} \quad (7)$$

Нижче зображене діалогове вікно (рисунок 8), в якому перевіряються нестандартизовані й стандартизовані оцінки координат і відстаней. Студенти вводять знайдені значення в темні комірки, після чого реалізується четвертий крок.

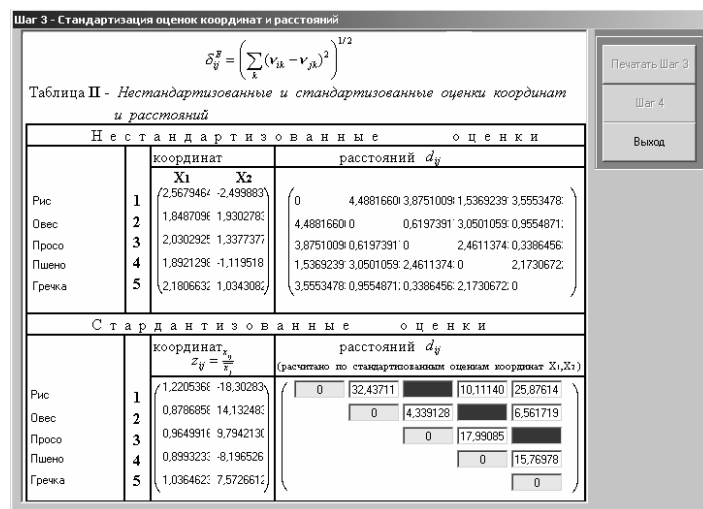


Рисунок 8 – Вікно перевірки нестандартизованих і стандартизованих оцінок координат і відстаней

Крок 4 – неметричний етап. Алгоритмічний крок призначений для впорядкування оцінок відстаней між стимулами. У теоретичному просторі шкал X_k монотонність вихідних даних може порушуватися. Коректування теоретичних величин відстаней d_{ij} здійснюється при незмінних оцінках координат стимулів і таким чином, щоб відновити загальну тенденцію до зростання у вихідних даних розходження.

Наступний графік (рисунок 9) побудований за даними розглянутого приклада і наочно показує невідповідність, що виникла у зміні вихідних і теоретичних рангових оцінок. По осі δ_{ij} відкладені фактичні значення

характеристик розходжень, по осі $\hat{\delta}_{ij}$ - значення, прийняті в теоретичному просторі шкал X1, X2. Лінія L1 - пряма монотонної функції рівномірно зростаючих оцінок, лінія L2- побудована з урахуванням відхилень емпіричних рангових оцінок від теоретичних.

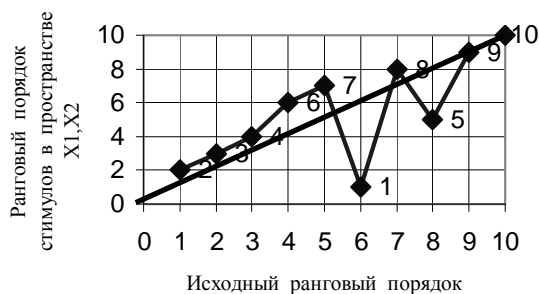


Рисунок 9 - Відносини рангових порядків стимулів по вихідним і теоретичним даним

Для перевірки цього кроку в АРМ використовується вікно, зображене на рисунку 10. Після введення правильних значень в темні комірки програма автоматично вимальовує лінію з відхиленнями емпіричних рангових оцінок від теоретичних.

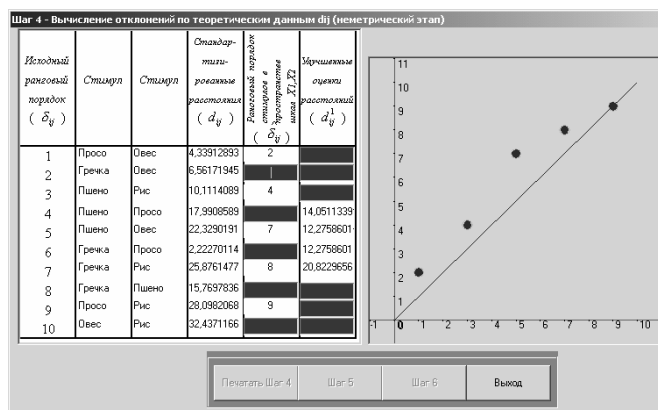


Рисунок 10 – Діалогове вікно перевірки вихідних рангових оцінок розходжень об'єктів і величин відстаней між ними в теоретичному просторі шкал

Крок 5 - метричний етап. На даному етапі наявним вихідним й уточненим величинам відстаней (d_i^c і d_i^{c+1}) знаходять уточнені оцінки координат. Для розрахунків використовується формулу Лінгоса-Роскама

$$x_{ik}^{c+1} = x_{ik}^c - \frac{1}{j} \sum_j \left(1 - \frac{d_{ij}^{c+1}}{d_{ij}^c}\right) (x_{ik}^c - x_{jk}^c) \quad (8)$$

Але перш ніж робити обчислення по формулі Лінгоса—Роскама в АРМ дається змістовна оцінка d_i^c і d_i^{c+1} .

Крок 6 - оцінка відповідностей монотонних рангових емпіричних і теоретичних даних. Перевірці на монотонність підлягають теоретичні дані d_i^c і d_i^{c+1} , розглядається ступінь їхнього поліпшення на минулій ітерації. Якщо поліпшення істотно, ітерація відновлюється після стандартизації отриманих на кроці 5 оцінок координат і відстаней, якщо ж поліпшення мало, ітерації закінчуються, і приступають до інтерпретації підсумків аналізу. Оцінювання відповідностей теоретичних результатів емпіричним даним здійснюється за допомогою спеціальних стресс-формул [11]

$$S_1 = \left(\frac{\sum_{ij} (d_{ij}^0 - d_{ij}^1)^2}{\sum_{ij} d_{ij}^{0^2}} \right)^{1/2}, \quad (10)$$

$$S_2 = \left(\frac{\sum_{ij} (d_{ij}^0 - d_{ij}^1)^2}{\sum_{ij} (d_{ij}^1 - d_{..})^2} \right)^{1/2}, \quad (11)$$

де $d_{..}$ - середнє арифметичне всіх оцінених відстаней

$$d_{..} = \frac{1}{ij} \sum_{ij} d_{ij} \quad (12)$$

Перевірка кроку проходить за допомогою вікна, що зображене на рисунку 11.

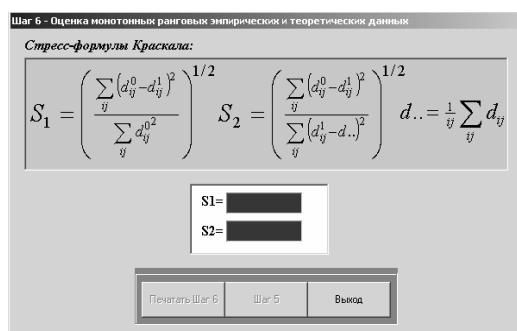


Рисунок 11 – Діалогове вікно перевірки оцінювання відповідностей теоретичних результатів емпіричним даним

Завдання інтерпретації величин, обчислених по стрес-формулах, полегшують відомі заздалегідь стандартні характеристики.

Після чого перевіряється крок 5 (рисунок 3.12), де

$Z = (x_1^0, x_2^0)$ - старі координати стимулів у факторному просторі X_1, X_2 ;

$Z_1 = (x_1^1, x_2^1)$ - нові координати стимулів у факторному просторі X_1, X_2 .

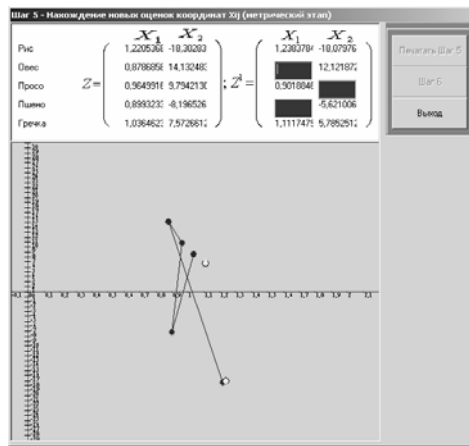


Рисунок 12 – Діалогове вікно перевірки п'ятого кроку

Класична модель багатовимірнього шкалювання Торгерсона

Постановка задачі має наступну інтерпретацію.

Дано матрицю ідентифікацій (таблиця 3.11), рядки якої - перелік об'єктів спостережень, стовпці - характерні ознаки. Використовуючи модель Торгерсона необхідно побудувати симетричну матрицю з подвійним центруванням, потім застосувати метод головних факторів.

Таблиця 3.11 - Матриця ідентифікацій

Об'єкт	Ознака		
	P ₁	P ₂	P ₃
X ₁	11,8	429,6	63,4
X ₂	1,5	88,8	10,1
X ₃	2	124,6	11,5
X ₄	3,1	-17,1	26,8
Середнє значення	4,6	156,475	27,95

При рішенні завдання АРМ використовує наступні етапи:

1. стандартизація вихідних даних;
 2. визначення матриці розходжень за стандартизованими даними;
 3. застосування класичної моделі багатомірнього шкалювання Торгерсона:
 - а) обчислення "середніх" характеристик ;
 - б) перетворення матриці розходжень у матрицю з подвійним центруванням.
 4. застосування методу головних факторів.
- Головне вікно тесту наведено на рисунку 13.

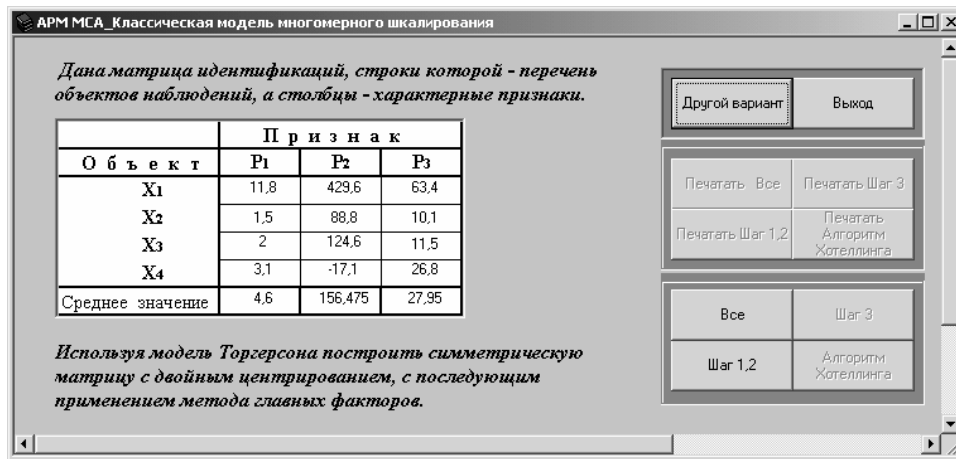


Рисунок 13 – Головне вікно тесту класичної моделі багатомірного шкалювання Торгерсона

Наведемо приклад рішення і перевірки результатів за допомогою тесту.

1) Стандартизуємо вихідні дані, використовуючи наступну формулу:

$$z_{ij} = \frac{v_{ij}}{\bar{v}_j}, \quad (13)$$

де v_{ij} – значення з матриці ідентифікацій;

\bar{v}_j – середнє значення j -го ознаки.

Одержимо

$$z_{ij} = \begin{pmatrix} 2,57 & 2,75 & 2,27 \\ 0,33 & 0,57 & 0,36 \\ 0,43 & 0,80 & 0,41 \\ 0,67 & -0,11 & 0,96 \end{pmatrix}$$

2) Побудова матриці мір розходження профілів.

Використовується евклідова відстань – формула 3.7.

Матриця мір розходження профілів прийме наступний вид

$$\Delta = \begin{pmatrix} 0 & 3,659765 & 3,433094 & 3,666262 \\ 3,659765 & 0 & 0,258203 & 0,967483 \\ 3,433094 & 0,258203 & 0 & 1,084852 \\ 3,666262 & 0,967483 & 1,084852 & 0 \end{pmatrix}$$

На рисунку 14 зображено вікно тесту, де перевіряються описані кроки.

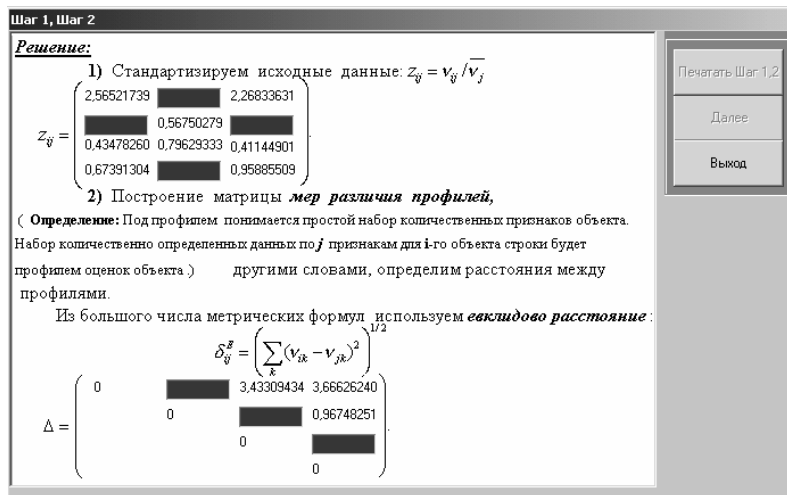


Рисунок 14 – Діалогове вікно тесту - перевірка кроків 1 і 2

3) Значення елементів матриці з подвійним центруванням $\Delta^* = (\delta_{ij}^*)$ знаходять по формулі

$$\delta_{ij}^* = -\frac{1}{2}(\delta_{ij}^2 - \delta_i^2 - \delta_j^2 + \delta_{...}^2), \quad (14)$$

де $\delta_i^2 = \frac{1}{j} \sum_j \delta_{ij}^2$ - середня для характеристик розходжень в j -ому стовпці i -ого рядка, зведених у квадрат;

$\delta_j^2 = \frac{1}{i} \sum_i \delta_{ij}^2$ - середня для характеристик розходжень в i -их рядках j -го стовпця, зведених у квадрат;

$\delta_{...}^2 = \frac{1}{ij} \sum_i \sum_j \delta_{ij}^2$ - середня величина для квадратів характеристик розходжень матриці Δ .

В АРМ для перевірки цього кроку використовується вікно, що зображене на рисунку 15.

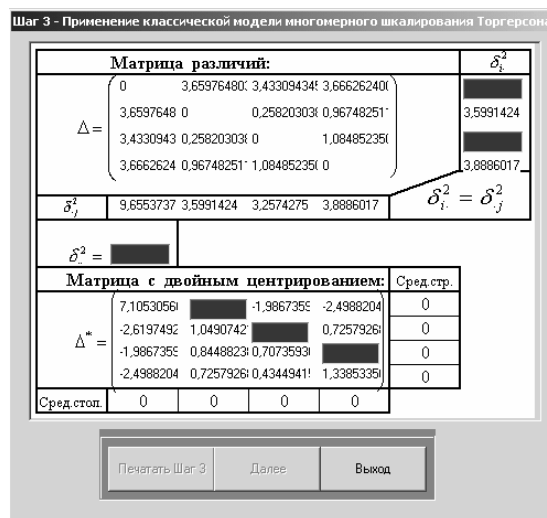


Рисунок 15 – Діалогове вікно перевірки кроку 3

4). Для матриці з подвійним центруванням застосовується класична модель головних факторів з кореляційною матрицею $R_k = \Delta^*$. Результат – таблиця 12.

Таблиця 12 - Підсумкова таблиця

	Головний фактор (факторні навантаження)	
	F ₁	F ₂
X ₁	-2,66551	0,018688
X ₂	0,980761	-0,29526
X ₃	0,74258	-0,39488
X ₄	0,942171	0,671452

Результати цієї таблиці перевіряються тестом (рисунок 16).

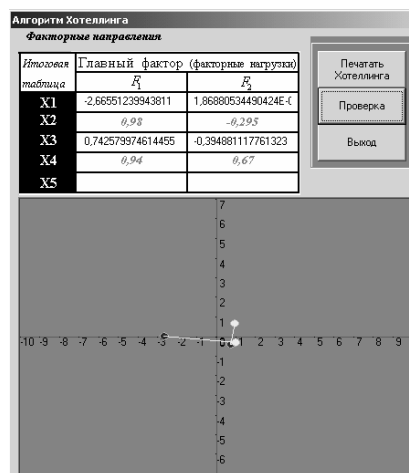


Рисунок 16 – Діалогове вікно тесту - перевірка реалізації методу головних факторів

Перевірка обчислень: для відомої матриці скалярних добутоків по Торгерсону діє рівність $\Delta^* = XX'$ (рисунок 17).

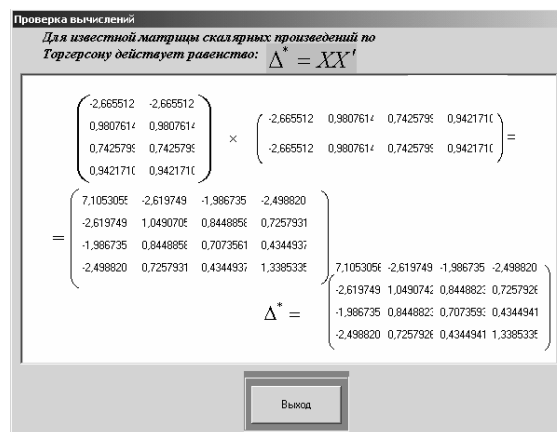


Рисунок 17 – Вікно тесту Перевірка зроблених обчислень

Узагальнена модель пошуку індивідуальних розходжень Такера

Постановка завдання наступна.

Дано вихідні матриці розходжень за результатами експертного оцінювання двома суб'єктами.

$$\Delta_1 = \begin{matrix} & \begin{matrix} \text{n1} & \text{n2} & \text{n3} \end{matrix} \\ \begin{matrix} \text{n1} \\ \text{n2} \\ \text{n3} \end{matrix} & \begin{pmatrix} 0 & 0,5 & 0,9 \\ 0,5 & 0 & 0,7 \\ 0,9 & 0,7 & 0 \end{pmatrix} \end{matrix} \quad \Delta_2 = \begin{matrix} & \begin{matrix} \text{n1} & \text{n2} & \text{n3} \end{matrix} \\ \begin{matrix} \text{n1} \\ \text{n2} \\ \text{n3} \end{matrix} & \begin{pmatrix} 0 & 0,2 & 0,1 \\ 0,2 & 0 & 0,4 \\ 0,1 & 0,4 & 0 \end{pmatrix} \end{matrix}$$

Необхідно реалізувати алгоритм для знаходження шкал і подати в координатному просторі як стимули, так і суб'єктів, їх оцінюючих.

Рішення цієї задачі в АРМ також проходить по крокам. Дано невеликий опис реалізації моделі програмою і вкажемо які параметри перевіряються при тестуванні. Головне вікно програми зображене на рисунку 18.

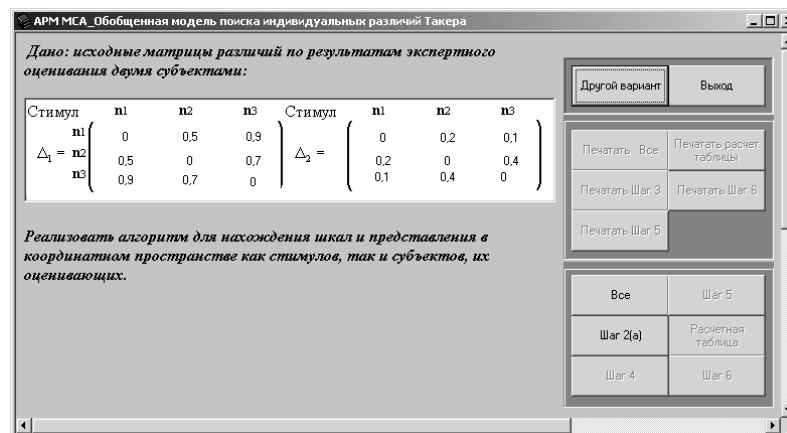


Рисунок 18 – Головне вікно тесту узагальненої моделі пошуку індивідуальних розходжень Такера

Крок 1 - побудова матриць розходжень стимулів Δ_s для кожного із суб'єктів (дається в завданні).

Крок 2 (а) - побудова S матриць скалярних добутків (матриць із подвійним центруванням).

Значення елементів матриці з подвійним центруванням $\Delta_s^* = (\delta_{ijs}^*)$ знаходяться по формулі

$$\delta_{ijs}^* = -\frac{1}{2}(\delta_{ijs}^2 - \delta_{i..s}^2 - \delta_{.js}^2 + \delta_{..s}^2) \quad (15)$$

Де $\delta_{i..s}^2 = \frac{1}{j} \sum_j \delta_{ijs}^2$ - середня для характеристик розходжень в j-их стовпцях i-ого рядка, зведених у квадрат ;

$\delta_{.js}^2 = \frac{1}{i} \sum_i \delta_{ijs}^2$ - середня для характеристик розходжень в і-их рядках j-го стовпця, зведених у квадрат ;

$\delta_{..}^2 = \frac{1}{ij} \sum_i \sum_j \delta_{ijs}^2$ - середня величина для квадратів характеристик розходжень матриці Δ .

Крок 2 (б) - для індивідуальних матриць скалярних добутоків розраховується матриця середніх.

$$\Delta^* = \frac{1}{2}(\Delta_1^* + \Delta_2^*) = \begin{pmatrix} 0,09 & 0,00 & -0,09 \\ 0,00 & 0,06 & -0,06 \\ -0,09 & -0,06 & 0,15 \end{pmatrix} .$$

Знайдені таким чином матриці скалярних добутоків (матриці 3 подвійним центруванням), матриця середніх програма заносить у вікно, що зображене на рисунку 19, скриваючи частину з них.

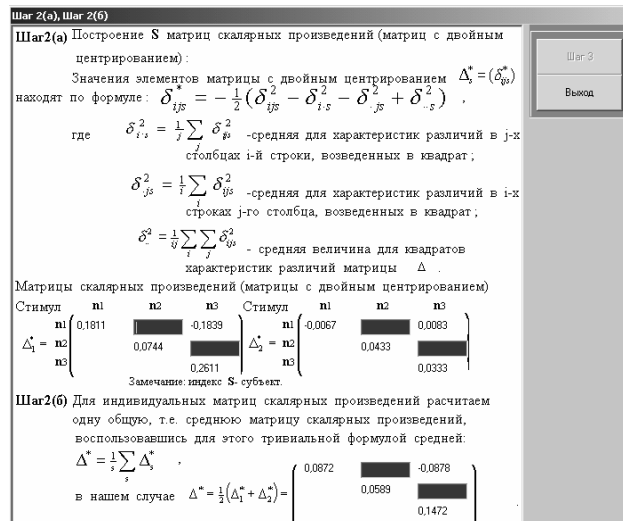


Рисунок 19 – Вікно тесту для перевірки кроків 1, 2(а) і 2(б)

Крок 3 - пошук наближених оцінок координат стимулів одержаних методом головних факторів.

Крок 4 - оцінка вагових коефіцієнтів, що показують, як сильно експерти віддають перевагу окремим шкалам, поліпшення оцінки координат стимулів, у наявному (нестандартизованому) шкальному просторі визначається розташування стимулів з урахуванням думки кожного із суб'єктів. Тестування проводиться за допомогою вікна, що зображене на рисунку 20.

Крок 5 - розрахунок F-критерію якості ітеративних оцінок скалярних добутоків, тобто висновки по виконанню першого кроку застосування евклідової метрики - потрібно чи виконувати друге застосування? Після проміжних обчислень знаходиться значення критерію F=0,01.

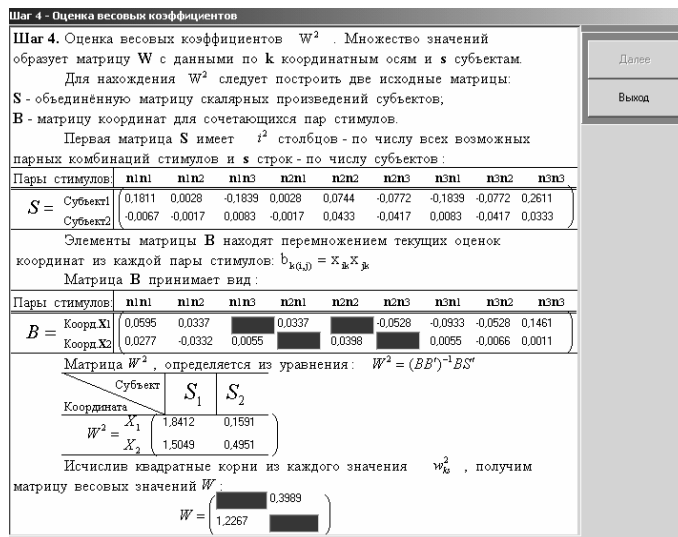


Рисунок 20 – Диалогове вікно тесту – перевірка четвертого кроку

Ціль ітеративного алгоритму складається в мінімізації значення критерію F , тобто сума квадратів різниць між фактичними й оціненими скалярними добутками повинна бути найменшою. Ітерації повторюються доти, поки при переході до наступної ітерації величина F не стане незначною, наприклад, менше 0,001.

У нашому прикладі $F=0,01 > 0,001$ - критерій якості не влаштовує, отже дослідження потрібно продовжити.

Вікно рисунку 21 перевіряє цей етап.

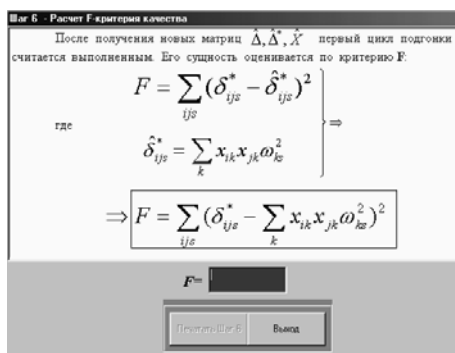


Рисунок 21 – Вікно тесту – перевірка п'ятого кроку

Перелік посилань

1. Иберла К. Факторный анализ / Пер. с нем. В. М. Ивановой. – М.: Статистика, 2001. – 398 с.: ил.
2. Многомерный статистический анализ в экономике: Учеб. пособие для вузов / Л. А. Сошникова, В. Н. Тамашевич, Г. Уебе, М. Шевер. – М.: ЮНИТИ – ДАНА, 1999. – 598 с.
3. Чекотовський, Едуард Васильович.
Графічний метод у статистиці на основі програми Excel: Навч. посіб. – К.: Знання, 2000. – 518 с.: іл.
4. Дисперсионный анализ / Пер. с англ. Б. А. Севастьянова. – 2-е изд. – М.: Наука, 2007. – 512 с.
5. Мидлтон, Майкл Р. Анализ статистических данных с использованием Microsoft Excel для Office XP: Практикум: Пер. с англ. – М.: БИНОМ, 2005. – 296 с.: ил.
6. Клейнен Дж. Статистические методы в имитационном моделировании: Вып. 1 / Пер. с англ. – М.: 2009. – 224 с.
7. Клейнен Дж. Статистические методы в имитационном моделировании: Вып. 2 / Пер. с англ. – М.: 2003. – 336 с.
8. Маркова Е. В. Комбинаторные планы в задачах многофакторного эксперимента / Е. В. Маркова, А. Н. Лисенков. – М.: Наука, 2004. – 346 с.
9. Себер Дж. Линейный регрессионный анализ / Пер. с англ. – М.: Мир, 2001. – 456 с.
10. Сильвестров, Дмитрий Сергеевич.
Пакеты прикладных программ статистического анализа / Д. С. Сильвестров, Н. А. Семенов, В. В. Марищук. – К.: Тэхніка, 2005. – 176 с.
11. Хейс Д. Причинный анализ в статистических исследованиях / Пер. с англ. М.: Финансы и статистика, 2007. – 256 с.: ил.