

5.2. Обробка результатів експертного оцінювання

5.2.1. Коефіцієнт конкордації

Дуже важливим етапом у підведенні результатів дослідження є прогнозування. Часто прогнозування передбачає визначення значень економічних або соціологічних показників у майбутньому. Наприклад, прогнозування думки респондентів щодо рекламованого товару, ціни на товари, об'єм їх продаж.

Прогнозування є початковим етапом планування і включає в себе попередній і кінцевий (формальний) прогнози, для яких розробляється один або декілька сценаріїв майбутніх подій.

Методи експертних оцінок використовуються для прогнозування майбутніх подій, якщо відсутні статистичні дані або їх недостатньо. Вони також застосовуються для кількісного вимірювання таких подій, для яких не існує інших способів вимірювання.

Припускається, що експерт формує своє судження на аналізі групи факторів, оцінюючи ймовірності їх реалізації та впливу на результативну ознаку об'єкта вивчення. Але, при цьому отримані висновки та оцінки пов'язані з особистістю експерта, тому інший експерт, використовуючи ту саму інформацію, може дійти інших висновків. Тому вважається, що при розв'язанні проблем в умовах невизначеності, думка групи експертів дає більш надійні результати, ніж думка одного експерта.

Після отримання експертних оцінок проводиться їх обробка та оцінюється достовірність. Обробку результатів експертного оцінювання можна проводити за **коефіцієнтом конкордації**, який показує ступінь згоди думок експертів. Найбільш достовірні оцінки отримуються за умов узгодженості думок експертів.

Коефіцієнт конкордації W розраховується за формулою:

$$W = \frac{12}{m^2(n^3 - n)} \sum_{j=1}^n \left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2, \quad (5.6)$$

де n – кількість об'єктів оцінювання;

m – кількість експертів;

R_{ij} – ранг j -го об'єкта, представленого i -м експертом.

Якщо об'єкти оцінювання мають однакові ранги, то коефіцієнт конкордації розраховується за формулою:

$$W = \frac{12}{\frac{1}{12} m^2 (n^3 - n) - m \sum_{j=1}^n T_j} \cdot \sum_{j=1}^n \left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2. \quad (5.7)$$

При цьому T_j обчислюється за формулою:

$$T_j = \frac{\sum_{i=1}^{L_i} (n_i^2 - n_i)}{12}, \quad (5.8)$$

де L_i – кількість груп однакових рангів,

n_i – кількість елементів i -тої групи для j -го експерта.

Статистична значущість коефіцієнта конкордації перевіряється порівнянням величини $n(m-1) \cdot W$ з табличним значенням розподілу χ^2 при рівні значущості $\alpha = 0,001$ та $n - 1$ степенях свободи.

Якщо коефіцієнт конкордації виявляється не значущим, то використовується методика виведення експерта, думка якого не узгоджується з думкою інших експертів. Для цього будується матриця коефіцієнтів кореляції Пірсона ($r(k, i)$) або рангових коефіцієнтів кореляції Спірмена ($r_s(k, i)$) та виявляється експерт, оцінка якого підкоряється умові:

$$r_j(k, i) = \min_{i=1, \dots, m} \{r(k, i)\}, \quad (5.9)$$

що означає, що думка цього експерта найменше узгоджується з думкою інших експертів. Бали, подані таким експертом, у подальших розрахунках не враховуються.

Алгоритм повторюється, поки коефіцієнт конкордації не стає значущим.

Приклад 5.3. Група експертів з 3 осіб оцінювала час, що необхідний для виконання робіт певного проекту. Результати оцінювання подано у табл. 5.2. Перевірити степінь узгодженості думок експертів.

Таблиця 5.2

Експерти	Час, необхідний для робіт			
	Робота 1	Робота 2	Робота 3	Робота 4
1-й	6	5	2	4
2-й	4	7	3	9
3-й	5	7	3	6

Розв’язок. Здійснимо перевірку за коефіцієнтом конкордації, для чого знайдемо ранги робіт проекту окремо за оцінками кожного з експертів (табл. 5.3).

Таблиця 5.3

Експерти	Ранги робіт			
	Робота 1	Робота 2	Робота 3	Робота 4
1-й	6	5	2	4
2-й	4	7	3	9
3-й	5	7	3	6

У групах рангів оцінок, наданих окремими експертами, немає однакових, тому коефіцієнт конкордації розраховуємо за формулою 5.6.

Обчислимо величини, що не залежать від індексів сум, враховуючи, що: n – кількість робіт, $n = 4$; m – кількість експертів, $m = 3$.

Отримаємо:

$$\frac{n+1}{2} = \frac{4+1}{2} = 2,5; \quad \frac{12}{m^2(n^3 - n)} = \frac{12}{3^2(4^3 - 4)} = \frac{12}{540} \approx 0,022.$$

Подальші обчислення для зручності представимо у вигляді таблиці (табл. 5.4).

Таблиця 5.4

Розрахункові формули	Результати розрахунків			
	Робота 1	Робота 2	Робота 3	Робота 4
$R_{ij} - \frac{n+1}{2}$	1,5	0,5	-1,5	-0,5
	-0,5	0,5	-1,5	1,5
	-0,5	1,5	-1,5	0,5
$\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right)$	0,5	2,5	-4,5	1,5
$\left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2$	0,25	6,25	20,25	2,25
$\sum_{j=1}^n \left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2$	29			

Отже, коефіцієнт конкордації:

$$W = \frac{12}{m^2(n^3 - n)} \sum_{j=1}^n \left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2 = 0,022 \cdot 29 = 0,638$$

Перевіримо його значущість: $n(m-1) \cdot W = 4(3-1) \cdot 0,638 = 5,104$;

критичне значення χ^2 : ХИ2ОБР (0,001; 4 - 1) = 16,27. Оскільки величина $n(m-1) \cdot W$ менша критичного значення χ^2 , то коефіцієнт конкордації не є значущим та думки експертів не узгоджені.

Виокремимо експерта, оцінки якого є найбільш неузгодженими. Для цього побудуємо матрицю парних коефіцієнтів кореляції Пірсона (табл. 5.5).

Таблиця 5.5

Експерти	1-й	2-й	3-й
1-й	1		
2-й	0,23035	1	
3-й	0,657143	0,797366	1

З таблиці 5.5 видно, що найменшим є значення коефіцієнта кореляції, який показує узгодженість думок першого та другого експертів, тому одного з них необхідно вивести з експертизи. Доцільно вивести першого експерта, тому що його оцінки є менш узгодженими з оцінками третього експерта.

Розрахуємо коефіцієнт конкордації (табл. 5.6), враховуючи відсутність оцінок першого експерта.

$$\text{Отже, } \frac{n+1}{2} = \frac{4+1}{2} = 2,5 ; \quad \frac{12}{m^2(n^3 - n)} = \frac{12}{2^2(4^3 - 4)} = \frac{12}{240} = 0,05$$

Таблиця 5.6

Розрахункові формули	Результати розрахунків			
	Робота 1	Робота 2	Робота 3	Робота 4
$R_{ij} - \frac{n+1}{2}$	-0,5	0,5	-1,5	1,5
	-0,5	1,5	-1,5	0,5
$\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right)$	-1	2	-3	2
$\left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2$	1	4	9	4
$\sum_{j=1}^n \left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2$	18			

Коефіцієнт конкордації:

$$W = \frac{12}{m^2(n^3 - n)} \sum_{j=1}^n \left(\sum_{i=1}^m \left(R_{ij} - \frac{n+1}{2} \right) \right)^2 = 0,05 \cdot 18 = 0,9$$

Отже, значення коефіцієнта конкордації після виведення першого експерта значно збільшилося. Але воно теж не є значущим. Однак це пов'язано не з тим, що думки експертів не погоджуються, а з тим, що кількість експертів надто мала.

Час, необхідний для виконання робіт проекту, розраховується як середнє арифметичне експертних оцінок.

5.2.2. Коефіцієнт компетенції

Використання коефіцієнта конкордації засновано на припущенні, що чим більш узгоджені думки експертів, тим достовірнішими є їх оцінки. Але практика показує, що це не завжди вірно, і експерт, який не згоден з думками більшості, може дати найточніші оцінки.

Якщо дослідник бажає врахувати думки всіх експертів, то обробку результатів експертного оцінювання слід виконувати за **коефіцієнтом компетентності** експерта.

Цей метод базується на використанні попередньої оцінки компетентності експертів, які приймають участь у дослідженні. Оцінка експертів проводиться за критеріями компетентності, серед яких можуть бути: рівень освіти; загальний стаж роботи; стаж роботи за проблемою дослідження; посада тощо. Крім того, важливим критерієм є оцінка надійності експерта, яка розраховується як відношення його правильних оцінок до всіх проведених експертиз. Правильними вважаються ті оцінки, які з часом підтвердилися практикою.

При розрахунку коефіцієнтів компетентності експертів необхідно використовувати єдину для всіх критеріїв шкалу оцінювання. У противному випадку оцінки потрібно буде нормалізувати, тобто привести до однієї шкали.

Коефіцієнт компетентності розраховується за формулою:

$$KK_i = \frac{\sum_{j=1}^m k_{ij}}{\sum_{i=1}^n \sum_{j=1}^m k_{ij}}, \quad (5.9)$$

де n – кількість експертів;

m – кількість критеріїв оцінювання експертів;

k_{ij} – бал, отриманий i -м експертом за j -м критерієм.

Приклад 5.4. За вхідними даними прикладу 5.3. знайти час, необхідний для виконання робіт проекту, з урахуванням коефіцієнта компетентності експертів. Бали, отримані експертами, подано у табл. 5.7. Оцінювання проводилося за трьохбальною шкалою.

Таблиця 5.7

Експерти	Бали, отримані експертами		
	Критерій 1. Стаж роботи	Критерій 2. Професіоналіз	Критерій 3. Надійність
1-й	1	2	2
2-й	2	3	3
3-й	2	1	1

Розв'язок. Знайдемо коефіцієнти компетентності експертів за формулою 5.9. Необхідні розрахунки внесемо у табл. 5.8.

Таблиця 5.8

Експерти	Бали, отримані експертами			Сума балів кожного експерта
	Критерій 1	Критерій 2	Критерій 3	
1-й	1	2	2	5
2-й	2	3	3	8
3-й	2	1	1	4
Загальна сума балів				17

Отже:

$$\text{для першого експерта } KK_1 = \frac{\sum_{j=1}^m k_{1j}}{\sum_{i=1}^n \sum_{j=1}^m k_{ij}} = \frac{5}{17} \approx 0,2941;$$

$$\text{аналогічно для другого і третього } KK_2 = \frac{8}{17} \approx 0,4706; \quad KK_3 = \frac{4}{17} = 0,2353.$$

Розрахуємо час, необхідний для виконання робіт проекту, з урахуванням коефіцієнта компетентності експертів за формулою: $t_j = \sum_{i=1}^n KK_i \cdot t_{ij}$; $j = \overline{1, m}$; де t_i – час для i -тої роботи; t_{ij} – оцінка часу i -тої роботи j -м експертом. Результати представлено у табл. 5.9.

Таблиця 5.9

Експерти	KK_i	Час, необхідний для робіт			
		Робота 1	Робота 2	Робота 3	Робота 4
1-й	0,2941	6	5	2	4
2-й	0,4706	4	7	3	9
3-й	0,2353	5	7	3	6
Час з урахуванням KK_i		4,82	6,41	2,71	6,82

5.3. Аналіз часових рядів із сезонною варіацією

Множина даних, отриманих у результаті спостережень, що проводилися регулярно через рівні інтервали часу, називається **часовим рядом**. У часових рядах час є факторною ознакою. Зміна в часі результативної ознаки називається **трендом**.

В процесі господарської діяльності окремі галузі промисловості, торгівля і сфера послуг стикаються з циклічними коливаннями, які викликані сезонним характером виробництва та споживання товарів і послуг. Повторення даних через певний проміжок часу називається **сезонною варіацією**.

Для аналізу тенденції зміни результативної ознаки на основі часового ряду сезонну варіацію даних необхідно виключити (провести десезоналізацію даних). Після цього за допомогою моделі лінійної регресії можна знайти рівняння тренда.

За допомогою рівняння тренда розробляються прогнози на наступні часові періоди. Кожен прогноз містить похибки, які бувають систематичними і випадковими. **Систематичні похибки** виникають внаслідок невірної моделі тренда, порушення сезонної варіації у неналежний бік і т. ін. **Випадкові похибки** – ті, що не можна пояснити моделлю тренда. Похибки обчислюються за рівнянням тренда і фактичними даними за формулами:

$$\text{середнє абсолютне відхилення } MAD = \sum_{i=1}^n \frac{|e_i|}{n}, \quad (5.10)$$

$$\text{середньоквадратична похибка } MSE = \sum_{i=1}^n \frac{e_i^2}{n}, \quad (5.11)$$

де e_i – різниця фактичного і трендового значень.

Як трендові найчастіше використовуються: адитивна модель, мультиплікативна модель та модель експоненційного згладжування.

Для адитивної моделі:

фактичне значення A = трендове значення T + сезонна варіація S + похибка E

Для мультиплікативної моделі:

фактичне значення A = трендове значення T * сезонна варіація S * похибка E

Для моделі експоненційного згладжування:

новий прогноз = α · фактичний результат в останній період +
+ $(1-\alpha)$ · прогноз в останній період

Константу згладжування α вибирають з відрізка $[0; 1]$. В умовах стабільності α належить відріжку $[0,2; 0,4]$, при швидкій зміні результативної ознаки – α вибирають з відрізка $[0,7; 0,9]$.

Побудову та аналіз моделей розглянемо на прикладі.

Приклад 5.5. В табл. 5.12 вказаний об'єм продажу (тис. грн.) за 11 кварталів. На основі цих даних зробити прогноз на наступні 2 квартали.

Таблиця 5.12

Квартал	1	2	3	4	5	6	7	8	9	10	11
Об'єм продажу	4	6	4	5	10	8	7	9	12	14	15

Розв'язок. Побудуємо адитивну модель за формулою:

фактичне значення A = трендове значення T + сезонна варіація S + похибка E

1. Перш за все виключимо вплив сезонної варіації. Користуємось методом ковзкого середнього. Для зручності обчислень заповнимо табл. 5.13.

Таблиця 5.13

Квартал	Об'єми продажу	Ковзке середнє за 4 квартали	Центроване ковзке середнє	Оцінка сезонної варіації
1	4	–	–	–
2	6	–	–	–
3	4	4,75	5,5	–1,5
4	5	6,25	6,5	–1,5
5	10	6,75	7,125	2,875
6	8	7,5	8	0
7	7	8,5	8,75	–1,75
8	9	9	9,75	–0,75
9	12	10,5	11,5	0,5
10	14	12,5	–	–
11	15	–	–	–

Пояснення до таблиці: 1 рік = 4 квартали, тому знайдемо середні значення об'єму продажу за 4 послідовних квартали. Тобто додаємо 4 послідовних числа із 2-го стовпчика і ділимо їх на 4. Результат записуємо у 3-й стовпчик – навпроти третього доданка). Якщо ковзке середнє обчислюється для непарної кількості періодів часу, то результат записується напроти середнього доданка і отримані середні не потрібно центрувати. У нашому випадку середні значення необхідно центрувати, для чого обчислюємо середнє арифметичне двох сусідніх чисел 3-го стовпчика і записуємо у 4-й. Оцінка сезонної варіації (5-й стовпчик) – це різниця фактичних даних і відповідних центрованих ковзких середніх (2-й стовпчик – 4-й стовпчик).

2. Знайдемо сезонну варіацію для кожного з чотирьох кварталів року. Для зручності результати обчислень оформимо у вигляді табл. 5.14.

Оцінки сезонної варіації запишемо у стовпчиках під відповідним номером кварталу. У кожному стовпчику обчислюємо середнє значення. Обчислюємо суму середніх значень (у даному прикладі вона дорівнює –1). Значення сезонної варіації повинні бути скоректовані так, щоб сума середніх значень дорівнювала 0 (середня варіація за рік). Для цього знаходимо коректувальний

коефіцієнт: суму середніх ділимо на 4 (число кварталів у році) та віднімаємо від усіх середніх даних коефіцієнт.

Таблиця 5.14

	Номер кварталу				Сума
	1	2	3	4	
Оцінки сезонної варіації	–	–	–1,5	–1,5	
	2,875	0	–1,75	–0,75	
	0,5	-	–	–	
Середнє	1,7	0,0	–1,6	–1,1	–1
Скоректована сезонна варіація	2,0	0,2	–1,3	–0,9	0,0

3. Виключимо сезонну варіацію із фактичних даних (табл. 5.15).

Таблиця 5.15

Квартал	Об'єми продажу A	Сезонна варіація S	Десезоналізований об'єм продажу $A-S=T+E$
1	4	2	2
2	6	0,2	5,8
3	4	–1,3	5,3
4	5	–0,9	5,9
5	10	2	8
6	8	0,2	7,8
7	7	–1,3	8,3
8	9	–0,9	9,9
9	12	2	10
10	14	0,2	13,8
11	15	–1,3	16,3

Із фактичних даних (2-й стовпчик) віднімаємо сезонну варіацію (3-й стовпчик) і записуємо результат в 4-й стовпчик.

4. Знайдемо рівняння тренда у вигляді лінійної регресійної моделі: $T = ax + v$. Для знаходження коефіцієнтів скористуємося статистичними функціями ОТРЕЗОК та НАКЛОН майстра функцій з пакету Excel. Отримаємо: $a = 1,1$; $v = 1,9$.

Отже, трендові значення об'єму продажу $= 1,9 + 1,1 \cdot$ номер кварталу.

5. Розрахуємо похибки обчислень. Для зручності результати обчислень оформимо у вигляді табл. 5.15.

Таблиця 5.15

Квартал	Об'єми продажу A	Десезоналізований об'єм продажу $A - S = T + E$	Трендові значення	Похибка e_t	$ e_t $	e_t^2
1	2	3	4	5	6	7
1	4	2	3	–1	1	1
2	6	5,8	4,1	1,7	1,7	2,89
3	4	5,3	5,2	0,1	0,1	0,01
4	5	5,9	6,3	–0,4	0,4	0,16
5	10	8	7,4	0,6	0,6	0,36

Продовження таблиці 5.15

1	2	3	4	5	6	7
6	8	7,8	8,5	-0,7	0,7	0,49
7	7	8,3	9,6	-1,3	1,3	1,69
8	9	9,9	10,7	-0,8	0,8	0,64
9	12	10	11,8	-1,8	1,8	3,24
10	14	13,8	12,9	0,9	0,9	0,81
11	15	16,3	14	2,3	2,3	5,29
Сума					11,6	16,58

$$MAD = \sum_{i=1}^n \frac{|e_i|}{n} = \frac{11,6}{11} \approx 1,1; \quad MSE = \sum_{i=1}^n \frac{e_i^2}{n} = \frac{16,58}{11} \approx 1,5.$$

6. Зробимо прогноз об'ємів продажу на наступні два квартали:

Прогноз об'ємів продажу в 12 кварталі: $(1,9+1,1*12)+(-0,9)=14,2$ тис. грн.

Прогноз об'ємів продажу в 13 кварталі: $(1,9+1,1*13)+2=18,2$ тис. грн.

Приклад 5.6. В табл. 5.16 вказаний об'єм продажу (тис. грн.) за 11 кварталів. На основі цих даних зробити прогноз на наступні 2 квартали.

Таблиця 5.16

Квартал	1	2	3	4	5	6	7	8	9	10	11
Об'єм продажу	63	74	79	120	67	79	88	130	69	82	90

Розв'язок. Побудуємо мультиплікативну модель за формулою:

фактичне значення A = трендове значення T * сезонна варіація S * похибка E

1. Перш за все виключимо вплив сезонної варіації. Користуємось методом ковзкого середнього. Для зручності обчислень заповнимо табл. 5.17.

Пояснення до таблиці: центроване ковзке середнє обчислюється аналогічно як і для адитивної моделі. Оцінка сезонної варіації (5-й стовпчик) – це частка від ділення фактичних даних на відповідні центровані ковзкі середні (2-й стовпчик / 4-й стовпчик).

Таблиця 5.17

Квартал	Об'єми продажу	Ковзке середнє за 4 квартали	Центроване ковзке середнє	Оцінка сезонної варіації
1	63	–	–	–
2	74	–	–	–
3	79	84	84,5	0,935
4	120	85	85,625	1,401
5	67	86,25	87,375	0,767
6	79	88,5	89,75	0,880
7	88	91	91,25	0,964
8	130	91,5	91,875	1,415
9	69	92,25	92,5	0,746
10	82	92,75	–	–
11	90	–	–	–

2. Знайдемо сезонну варіацію для кожного з чотирьох кварталів року. Для зручності результати обчислень оформимо у вигляді табл. 5.18.

Таблиця 5.18

	Номер квартала				
	1	2	3	4	
Оцінки сезонної варіації	–	–	0,935	1,401	Сума
	0,767	0,880	0,964	1,415	
	0,746	–	–	–	
Середнє	0,756	0,880	0,950	1,408	3,994
Скоректована сезонна варіація	0,757	0,881	0,952	1,410	4,0

Оцінки сезонної варіації запишемо у стовпчику з відповідним номером квартала. У кожному стовпчику обчислюємо середнє значення. Обчислюємо суму середніх значень (у даному прикладі вона дорівнює 3,994). Значення сезонної варіації повинні бути скоректовані так, щоб сума середніх значень дорівнювала 4 (4 частки від чотирьох кварталів – середня варіація за рік). Для цього знаходимо коректувальний коефіцієнт: 4 ділимо на суму середніх та множимо усі середні на цей коефіцієнт.

3. Виключимо сезонну варіацію із фактичних даних (табл. 5.19).

Таблиця 5.19

Квартал	Об'єми продажу A	Сезонна варіація S	Десезонізований об'єм продажу $A / S = T * E$
1	63	0,757	83,176
2	74	0,881	83,953
3	79	0,951	83,074
4	120	1,410	85,096
5	67	0,757	88,457
6	79	0,881	89,626
7	88	0,951	92,538
8	130	1,410	92,188
9	69	0,757	91,098
10	82	0,881	93,029
11	90	0,951	94,641

Фактичні дані (2-й стовпчик) ділимо на сезонну варіацію (3-й стовпчик) і записуємо результат в 4-й стовпчик.

4. Знайдемо рівняння тренда у вигляді лінійної регресійної моделі: $T = ax + b$. Для знаходження коефіцієнтів скористуємося статистичними функціями ОТРЕЗОК та НАКЛОН майстра функцій із пакету Excel. Отримаємо: $a = 1,2$; $b = 81,6$.

Отже, трендові значення об'єму продаж = $81,6 + 1,2 * \text{номер квартала}$.

5. Розрахуємо похибки обчислень. Для зручності результати обчислень оформимо у вигляді таблиці 5.20.

$$MAD = \sum_{i=1}^n \frac{|e_i|}{n} = \frac{11,2}{11} \approx 1; \quad MSE = \sum_{i=1}^n \frac{e_i^2}{n} = \frac{17,1}{11} \approx 1,6 - \text{похибки складають}$$

приблизно 1%.

6. Зробимо прогноз об'ємів продажу на наступні два квартали:

Прогноз об'ємів продажу в 12 кварталі: $(81,6+1,2 \cdot 12) \cdot 1,41=135,4$ тис. грн.

Прогноз об'ємів продажу в 13 кварталі: $(81,6+1,2 \cdot 13) \cdot 0,757=73,6$ тис. грн.

Таблиця 5.20

Квартал	Об'єми продажу A	Десезоналізований об'єм продажу $A/S=T \cdot E$	Трендові значення	Похибка e_t	$ e_t $	e_t^2
1	63	83,176	82,8	0,4	0,4	0,16
2	74	83,953	84	0,0	0,0	0,00
3	79	83,074	85,2	-2,2	2,2	4,84
4	120	85,096	86,4	-1,3	1,3	1,69
5	67	88,457	87,6	0,9	0,9	0,81
6	79	89,626	88,8	0,9	0,9	0,81
7	88	92,538	90	2,4	2,4	5,76
8	130	92,188	91,2	1,0	1,0	1,00
9	69	91,098	92,4	-1,3	1,3	1,69
10	82	93,029	93,6	-0,5	0,5	0,25
11	90	94,641	94,8	-0,3	0,3	0,09
Сума					11,2	17,10

Приклад 5.7. В таблиці 5.21 вказаний об'єм продажу (тис. грн.) за 11 кварталів. На основі цих даних зробити прогноз на наступні 2 квартали.

Таблиця 5.21

Квартал	1	2	3	4	5	6	7	8	9	10	11
Об'єм продажу	4	5	5	6	9	9	8	10	11	13	16

Розв'язок. Побудуємо модель експоненційного згладжування за формулою:

$$\text{новий прогноз} = \alpha \cdot \text{фактичний результат в останній період} + (1-\alpha) \cdot \text{прогноз в останній період}$$

Нехай $\alpha = 0,8$, тоді $1-\alpha = 0,2$. Перший прогноз обираємо рівним першому фактичному значенню, далі користуємось формулою експоненційного згладжування.

Експоненційне згладжування зручно проводити за допомогою сервісних функцій Excel. Необхідно викликати **Сервіс – Аналіз даних – Експоненціальне сглаживание – ОК**. У графі **Фактор затухання** вказати значення $1 - \alpha$ (стандартне значення 0,3).

Для зручності результати обчислень оформимо у вигляді табл. 5.22.

Таблиця 5.22

Квартал	Об'єми продажу	Прогноз
1	2	3
1	4	4
2	6	4
3	4	5,6
4	5	4,32

Продовження таблиці 5.22

1	2	3
5	10	4,864
6	8	8,9728
7	7	8,19456
8	9	7,238912
9	12	8,6477824
10	14	11,32955648
11	15	13,4659113
12		13,77

Похибка прогнозу обчислюється аналогічно як і для адитивної і мультиплікативної моделей.

5.4. Елементи факторного аналізу

5.4.1. Елементи факторного аналізу

Факторний аналіз – сукупність моделей і методів, призначених для стискання інформації, яка міститься в кореляційній матриці. Він допомагає виявити приховані фактори, які пояснюють взаємозв'язки між спостережуваними ознаками досліджуваного об'єкта. Кількість ознак може бути великою і зв'язки між ними надзвичайно складними, однак, спостерігаючи за об'єктом, ми можемо виявити невелику кількість факторів, які впливають на досліджувані ознаки. Факторний аналіз передбачає класифікацію ознак, які мають подібний характер зміни при переході від одного об'єкта спостереження до іншого.

Обґрунтована заміна великої кількості ознак, описаних об'єктами спостережень, меншим числом комплексних характеристик (факторів) складають зміст факторного аналізу. Кожний **фактор** – це група взаємопов'язаних ознак, які визначають змістовну інтерпретацію даного фактора. При цьому в один фактор об'єднуються ознаки, які тісно корелюють між собою. Ознаки з різних факторів характеризуються слабким кореляційним зв'язком.

Основними етапами факторного аналізу є:

- 1) збір емпіричних даних і підготовка кореляційної (коваріаційної) матриці;
- 2) виділення початкових факторів і побудова факторної структури (обчислення факторних навантажень): проводиться вибір методу обчислення, визначається кількість факторів на основі змістовних або математичних міркувань;
- 3) обертання факторної структури: вибір критерію обертання;
- 4) змістовна інтерпретація результатів факторного аналізу;
- 5) обчислення факторних значень.

Припустимо, що досліджується деякий масив з n емпіричних ознак. За допомогою методів кореляційного аналізу можна встановити залежності між ними, обчисливши коефіцієнти кореляції. Тоді вся множина з n емпіричних

ознак розіб'ється на окремі групи за величиною коефіцієнтів кореляції. Наприклад, перша ознака тісно пов'язана з четвертою і шостою, а шоста з першою, перша з п'ятою і т. д., причому з іншими ознаками зв'язок виявляється значно слабкішим. Тоді ці взаємопов'язані ознаки утворюють загальну функціональну одиницю, яку ми і називаємо фактором. Наприклад, при аналізі успішності студентів деякий фактор має високу додатну кореляцію з оцінкою з вищої математики, інформатики, розділів фізики і високу від'ємну кореляцію з політологією, філософією, історією – фактор характеризує точне мислення.

Створення математичної моделі факторного аналізу базується на припущенні про те, що усі зміни значень ознак обумовлені зміною деяких прихованих властивостей спостережуваних об'єктів. Ці приховані властивості називаються **загальними факторами** і їх кількість має бути меншою від числа ознак, за допомогою яких вони вимірюються. Кожний такий фактор має окреме значення значущості для різних досліджуваних ознак. Рівень значущості кожного фактора називається його **факторним навантаженням**. Він визначає степінь впливу загального фактора на зміну даної ознаки.

На зміну значень спостережуваної ознаки можуть впливати також деякі суб'єктивні, властиві тільки цій ознаці, зміни. Вони можуть бути викликані випадковими помилками, похибками вимірів і т.д.. Причини усіх таких не взаємообумовлених змін об'єднуються в поняття **специфічного фактора**.

Отже, зміни значень спостережуваних ознак залежать від двох складових: загальних факторів і специфічних. Нехай X_i – i -та емпірична ознака. Позначимо через U_i – загальну частину цієї ознаки (частина змін, викликана впливом загальних факторів) і ε_i – специфічні фактори, зміни яких не пов'язані один з одним і не залежать від змін інших показників зміни ознаки X_i . Тоді зміна ознаки X_i розкладається на суму загальної частини усіх змін і змін, викликаних впливом специфічних факторів.

$$X_i = U_i + \varepsilon_i.$$

Подальший розвиток ідеї факторного аналізу ґрунтується на тому припущенні, що дані n змінних U_i є лінійними комбінаціями меншого числа інших змінних F_j , які називаються факторами, тобто

$$U_i = \omega_{i1}F_1 + \omega_{i2}F_2 + \dots + \omega_{ik}F_k, \text{ де } i = \overline{1, n},$$

ω_{ji} – факторні навантаження факторів F_j , які характеризують степінь впливу j -го загального фактора на i -у емпіричну ознаку. Об'єднуючи вище зазначені формули, отримуємо, що X_i виражається через загальні і специфічні фактори таким чином:

$$X_i = \omega_{i1}F_1 + \omega_{i2}F_2 + \dots + \omega_{ik}F_k + \varepsilon_i, \text{ де } i = \overline{1, n}.$$

Причому:

1. Загальні фактори F_j є або некорельованими випадковими величинами з дисперсією, що дорівнює 1, або невідомими невідповідними параметрами.
2. Специфічні фактори ε_i мають нормальний розподіл, не корелюють між собою і не залежать від загальних факторів.

Тому наступним кроком має бути оцінка значущості загальних і специфічних факторів для зміни значень ознаки X_i . Для цього розглянемо дисперсію ознаки X_i :

$$D(X_i) = \frac{1}{N-1} \sum_{p=1}^N (x_{ip} - \bar{x}_i)^2,$$

де N – кількість об'єктів спостереження; x_{ip} – значення ознаки X_i для p -го об'єкта спостереження ($p = \overline{1, N}$); \bar{x}_i – середнє значення i -ї ознаки. Так як U_i і ε_i не корелюють між собою, то:

$$D(X_i) = D(U_i) + D(\varepsilon_i).$$

Чим більше значення $D(U_i)$, тим більша частина змін ознаки X_i залежить від загальних факторів.

Значення дисперсій загальних факторів $D(F_j)$ дозволяють ранжувати їх за степенями впливу на зміну ознак X_i .

Математична модель факторного аналізу є розділом багатомірної статистики і потребує достатньо глибоких знань у таких розділах вищої математики, як матрична алгебра, основи математичної статистики і математичного аналізу. Тому опишемо змістовну структуру факторного аналізу, опускаючи складні математичні і логічні обґрунтування.

Усі методи оцінки зв'язків ознак можна розділити на дві групи: прямі і непрямі. Прямі – це методи, які дозволяють визначити силу зв'язку до проведення факторного аналізу. До непрямих методів відносяться апостеріорні оцінки, коли до проведення факторного аналізу вони не відомі (беруться довільні оцінки), а в якості додаткової умови береться кількість факторів.

Найпоширеніший прямий метод оцінки зв'язку між ознаками базується на припущенні про те, що усі оцінки зв'язків дорівнюють 1. Тобто, вони дорівнюють діагональним елементам кореляційної матриці, а, отже, дисперсії специфічних факторів дорівнюють 0. У цьому полягає зміст однієї з найвідоміших моделей факторного аналізу – методу головних компонент.

Метод головних компонент. Для оцінки факторних навантажень, в якості критерію, використовується мінімум розбіжності між кореляційною матрицею початкових ознак і тією, яка виходить після оцінювання навантажень. Іншими словами, метод головних компонент здійснює перехід до нової системи координат, яка є системою ортонормованих лінійних комбінацій. Лінійні комбінації є власними векторами кореляційної матриці. Перша головна компонента – це лінійна комбінація, яка має найбільшу дисперсію. Друга компонента має найбільшу дисперсію серед усіх інших лінійних комбінацій, які не корелюють з першою головною компонентою.

Метод головних факторів. Це одна з найпоширеніших моделей факторного аналізу. Метод вимагає попередньої оцінки дисперсій і для нього критерієм оптимальної оцінки факторних навантажень є максимальна наближеність початкових кореляцій ознак до тих, які отримані в моделі після оцінювання навантажень. Для визначення кількості факторів використовуються різні статистичні критерії, за допомогою яких перевіряється гіпотеза про незначущість матриці кореляційних лишків.

Метод максимальної правдоподібності (Д. Лоулі) на відміну від попередніх моделей факторного аналізу ґрунтується не на попередній оцінці дисперсій, а на апріорному визначенні кількості загальних факторів. У разі великої вибірки дозволяє отримати статистичний критерій значущості отриманого факторного рішення.

Метод мінімальних залишків (Г. Харман) ґрунтується на мінімізації недиагональних елементів залишкової матриці кореляцій, і так як метод максимальної правдоподібності, вимагає попереднього вибору кількості факторів, що пояснюють спільні зміни ознак.

Перераховані методи відрізняються за способом пошуку розв'язання основного рівняння факторного аналізу. Вибір методу вимагає великого досвіду роботи. Проте деякі дослідники використовують відразу декілька методів, і виділені в усіх методах фактори вважають найбільш стійкими.

Моделі факторного аналізу називають іноді «прямими» або «початковими» в тому розумінні, що отримувані з їх допомогою оцінки навантажень напряму залежать від значень початкових ознак. Але на практиці може виникнути ситуація, коли навантаження при емпіричних ознаках утворюють таку комбінацію знаків і величин, яка важко інтерпретується. Ці випадки змусили дослідників шукати шляхи зміни початкових навантажень, щоб отримати результат, який краще інтерпретується. Один з методів вирішення цієї задачі – обертання факторів, тобто поворот відповідних факторам координатних осей, який проводиться не в просторі початкових ознак, а в просторі знайдених факторів.

Отже, наступним етапом факторного аналізу є обертання факторів, яке базується на принципах простої структури Терстоуна:

- в кожній стрічці факторної структури має бути хоча б один нуль;
- в кожному стовпчику – принаймні k нулів (k – кількість загальних факторів);
- для кожної пари стовпчиків можна знайти принаймні k параметрів (емпіричних ознак), для яких елементи факторної структури дорівнюють нулю в одному з двох стовпчиків і не дорівнюють нулю – в іншому.

На основі цих принципів побудовано велику кількість аналітичних методів, які максимізують деякий критерій. Суть процесу обертання будь-якої пари векторів полягає в знаходженні такого кута між новим і старим напрямом факторів, який давав би найбільший приріст обраного критерію. Обертання факторів в просторі дає можливість кожній ознаці охарактеризуватися переважачим впливом якогось одного фактора.

У сучасних пакетах статистичної обробки даних найчастіше використовуються методи обертання: варімакс, квартімакс і еквімакс. Обертання методом варімаксу спрощує значення стовпчиків факторної матриці, зводячи їх до 1 або 0. Обертання методом квартімаксу спрощує значення елементів стрічок факторної матриці. І, нарешті, еквімакс займає проміжне положення – при обертанні факторів за цим методом одночасно робиться спроба спростити значення елементів і стовпчиків, і стрічок.

Отримавши після процедури обертання факторне розв'язання (факторну матрицю), можна переходити до інтерпретації і найменування факторів. Цей етап роботи цілком і повністю залежить від інтуїції, рівня обізнаності і практичного досвіду дослідника. Щоб зрозуміти природу конкретного фактора, необхідно проаналізувати зміст ознак, які входять у даний фактор, і спробувати виявити спільні для них риси. Чим більша кількість ознак з великим значенням навантажень у цьому факторі, тим легше розкрити його природу. Для вибору назви фактора немає формалізованих прийомів. В якості попереднього варіанту можна використовувати ім'я ознаки, яка увійшла до фактора з найбільшим навантаженням.

Отже, процес інтерпретації факторних навантажень – це пошук таких загальних властивостей системи спостережень, які б могли бути описані в термінах одночасного збільшення (зменшення) однієї групи ознак на противагу зменшенню (збільшенню) іншої групи або просто збільшенню (зменшенню) якої-небудь частини ознак.

Якщо фактори знайдені і представлені, то на останньому кроці факторного аналізу, додатковим ознакам (додаткові запитання анкети) можна присвоїти значення цих факторів, так звані факторні значення. Тоді для кожного об'єкта спостереження значення великої кількості ознак можна перевести у значення невеликої кількості факторів – нових ознак.

Факторний аналіз є складною процедурою. Як правило, досконале факторне розв'язання (досить просте і таке, що змістовно інтерпретується) вдається отримати щонайменше після декількох циклів його проведення – від відбору ознак до спроби інтерпретації після обертання факторів. Для успішного проведення факторного аналізу необхідно дотримуватись основних вимог:

1) Змінні мають належати до шкали інтервалів (за класифікацією Стівенса). Передбачається, що порядкові змінні підлягають факторному аналізу, якщо їм надати числових значень. Не слід включати дихотомічні змінні в аналіз, якщо завдання не передбачають зменшення кількості ознак.

2) Відбираючи ознаки для факторного аналізу, слід враховувати, що на один фактор має припадати не менше трьох ознак.

3) Не варто включати у факторний аналіз ознаки, які мають дуже слабкі зв'язки з іншими ознаками. Велика ймовірність того, що вони не ввійдуть ні до одного фактора. Якщо в роботі не стоїть завдання сформулювати шкалу нового опитування на основі факторного аналізу або якого-небудь аналогічного завдання, то не слід також включати усі ознаки, які мають між собою дуже тісні зв'язки. Швидше за все, вони утворюють один фактор. Чим більше таких ознак включається у факторний аналіз, тим більша ймовірність того, що вони утворюють перший фактор і до нього приєднається більшість інших ознак.

4) Стійкість виявленої факторної структури (її невипадковість) тим менша, чим більше складових її факторів. Вона також нестійка при малій кількості випробовуваних об'єктів.

5.4.2. Факторний аналіз засобами SPSS

Проілюструємо описаний метод на прикладі складеної і апробованої нами анкети. Етап дослідження був проведений для населення м. Одеси віком 18-35 років, в процесі якого вивчалася думка молоді щодо новинок вітчизняного ринку – продуктів з вмістом ГМО. Респондентам пропонували висловити свою думку до наступних положень:

1. Потрібно досконало вивчити вплив ГМО на організм людини.
2. Необхідно виробити стандарт, який би регулював допустиме відсоткове відношення генно-модифікованих організмів у продуктах харчування.
3. В Україні існує велика кількість натуральних продуктів.
4. Кількість населення і так іде на спад, а тут ще й продукти з ГМО.
5. Варто прислухатися до розвитку інноваційних технологій – природні ресурси вичерпні.
6. Кількість продуктів з ГМО на ринку України варто обмежити.
7. Натуральні продукти зникли.
8. Необхідно вивчити результати досліджень міжнародних компаній щодо вмісту неприродних компонентів.
9. Заборона використання продуктів з вмістом ГМО принесе шкоду економіці країни.
10. Варто агітувати населення країни щодо споживання натуральних продуктів.
11. Ми не маємо достовірної інформації про вміст продуктів, які ми вживаємо.
12. Порушуючи права споживачів, виробники компрометують себе.
13. Ми варті якісних продуктів.
14. Не хочу навіть чути про ГМО.
15. Апробація продуктів з ГМО – поступове скорочення кількості населення.

Оцінки ставилися за семибальною шкалою: від повної незгоди (-3) до повної згоди (3). За результатами опитування 110 респондентів проводився факторний аналіз засобами SPSS. Запитання анкети введені під іменами X_1, X_2, \dots, X_{15} відповідно. Для виявлення факторів п'ятнадцяти ознак опишемо можливість використання програми.

Для цього в меню необхідно вибрати послідовність команд:

- 1) *Анализ* → *Снижение размерности* → *Факторный анализ* ...
Відкриється діалогове вікно *Факторный анализ* (рис. 5.1);

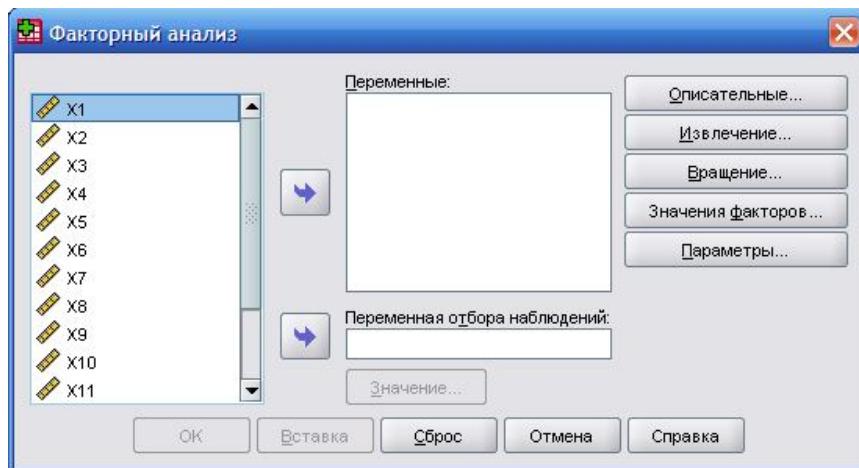


Рисунок 5.1. Діалогове вікно для факторного аналізу

2) Змінні $X1 - X15$ необхідно перенести в поле *Переменные* і ознайомитись з можливостями різних кнопок даного діалогового меню.

Після натискування на кнопку *Описательные...* відкриється діалогове вікно *Факторный анализ: Описательные*, зображене на рис. 5.2, у якому варто залишити виведення первинних результатів, які включають початкові відносні дисперсії простих факторів, власні значення і процентні частки об'єднаної дисперсії. Часто виникає необхідність у виведенні одновимірних статистик і кореляційних коефіцієнтів, яку можна реалізувати за допомогою функцій та можливостей даного діалогового вікна.

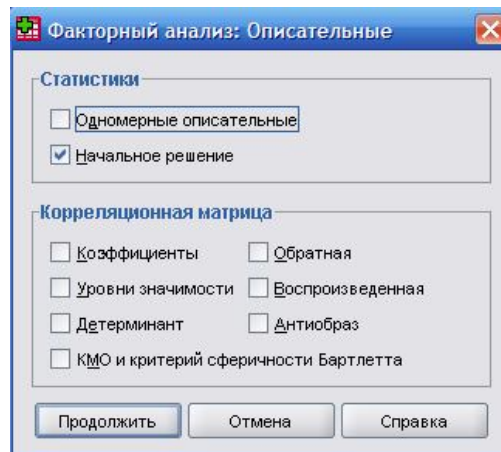


Рисунок 5.2. Діалогове вікно вибору параметрів факторного аналізу

За допомогою кнопки *Извлечение* можна вибрати метод відбору (перелік методів зображено на рис. 5.3) факторів; залишаємо встановлений автоматично аналіз головних компонентів. Кількість вибраних факторів прирівнюється до власних значень, які перевищують одиницю. Величину значення можна відкорегувати відповідною опцією.

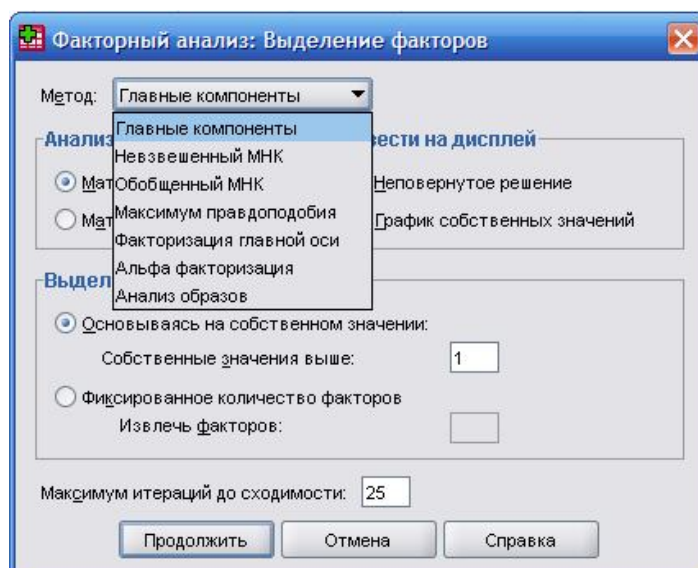


Рисунок 5.3. Методи відбору змінних у фактори

3) У діалоговому вікні *Факторный анализ: Вращение* вибирається метод обертання. Активуємо метод варімаксу і залишаємо активним виведення поверненої матриці факторів. Можна здійснити інтерпретацію факторних навантажень в графічному виді, в якому перші три фактори будуть представлені в тривимірному просторі; у випадку наявності тільки двох факторів – на площині.

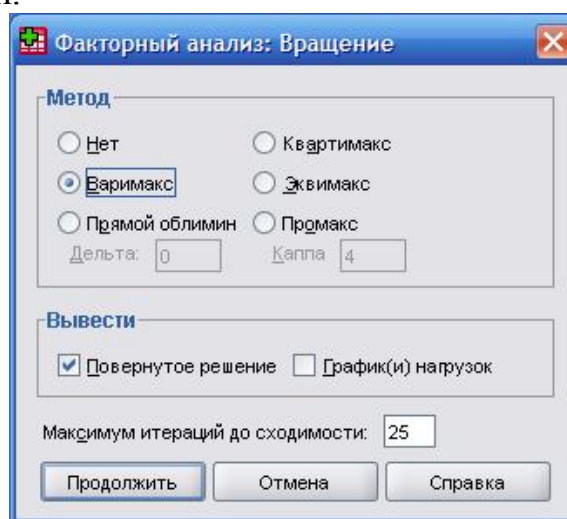


Рисунок 5.4. Діалогове вікно вибору методу обертання

Якщо потрібно знайти значення факторів і зберегти їх у вигляді додаткових змінних, варто натиснути на кнопку *Значения факторов* і відмітити *Сохранить как переменные*. Автоматично встановлений регресійний метод.

Пункт *Параметры* призначений для обробки пропущених значень.

4) Для проведення розрахунків натискаємо на *ОК*. У вікні виводу з'являться результати. Спочатку приводяться первинні статистики, які наведені у табл. 5.23:

Таблиця 5.23

Повна пояснена дисперсія

Компоненти	Початкові власні значення			Суми квадратів навантажень		
	Всього	%	Кумулятивний %	Всього	% дисперсії	Кумулятивний %
1	5,146	34,308	34,308	3,466	23,105	23,105
2	1,945	12,970	47,278	2,536	16,907	40,013
3	1,415	9,433	56,711	2,505	16,698	56,711
4	,990	6,601	63,312			
5	,936	6,238	69,550			
6	,760	5,068	74,617			
7	,693	4,622	79,240			
8	,612	4,083	83,323			
9	,529	3,529	86,852			
10	,473	3,151	90,004			
11	,433	2,889	92,893			
12	,339	2,262	95,1555			
13	,301	2,007	97,161			
14	,245	1,635	98,797			
15	,181	1,203	100,000			

Метод відбору: Аналіз головних компонент

За даними таблиці можна побачити, що значення трьох власних факторів більше за одиницю. Отже, для аналізу відібрано тільки три фактори. Перший фактор пояснює 34,308 % сумарної дисперсії, другий – 12,97 % і третій – 9,433 %. Оскільки ми відмінили виведення неповерненої матриці факторів, то даними табл. 5.24 є значення поверненої матриці.

Таблиця 5.24

Матриця повернутих компонент

	Компонент		
	1	2	3
X1	-.466	.628	-.191
X2	-.141	.657	.215
X3	.327	-.153	.711
X4	.533	-.106	.394
X5	-.362	.783	4.52E-02
X6	-1.2E-02	-3.8E-02	.763
X7	.525	3.58E-02	.543
X8	-.117	.719	-.267
X9	2.56E-02	.551	-8.8E-02
X10	.252	-9.5E-02	.685
X11	.125	.392	-.292
X12	.802	-.199	.108
X13	.685	-.110	.465
X14	.837	-.144	-2.5E-02
X15	.725	-4.8E-02	.144

Метод відбору: Аналіз головних компонент

Метод обертання: Варімакс з нормалізацією

Кайзера

а. Обертання виконано за 8 ітерацій

Найголовніша частина факторного аналізу – пояснення відібраних факторів. Для цього в кожній стрічці поверненої факторної матриці потрібно відмітити те факторне навантаження, яке має найбільше абсолютне значення.

Ці факторні навантаження слід розуміти як кореляційні коефіцієнти між ознаками і факторами. Так положення $X1$ найбільше корелює з фактором 2, величина кореляції складає 0,628; ознака $X2$ також найсильніше корелює з фактором 2 згідно значення 0,657; ознака $X3$ найтісніше – з фактором 3 (0,711) і т. д. У більшості випадків включення окремої змінної в один фактор, здійснюване на основі коефіцієнтів кореляції, є однозначним. В окремих випадках, наприклад, як в ситуації із ознакою $X7$, вона може відноситися до двох факторів одночасно. Можуть бути також і ознаки, в нашому прикладі $X11$, які не можна включити ні в один із вибраних факторів.

Отже, на основі вище зазначеного, ознаки можна віднести в наступному порядку до трьох факторів:

Фактор 1

Кількість населення і так іде на спад, а тут ще й продукти з ГМО.

Натуральні продукти зникли.

Порушуючи права споживачів, виробники компрометують себе.

Ми варті якісних продуктів.

Не хочу навіть чути про ГМО.

Апробація продуктів з ГМО – поступове скорочення кількості населення.

Фактор 2

Потрібно досконало вивчити вплив ГМО на організм людини.

Необхідно виробити стандарт, який би регулював допустиме відсоткове відношення генно-модифікованих організмів у продуктах харчування.

Варто прислухатися до розвитку інноваційних технологій – природні ресурси вичерпні.

Необхідно вивчити результати досліджень міжнародних компаній щодо вмісту неприродних компонентів.

Заборона використання продуктів з вмістом ГМО принесе шкоду економіці країни.

Ми не маємо достовірної інформації про вміст продуктів, які ми вживаємо.

Фактор 3

В Україні існує велика кількість натуральних продуктів.

Кількість продуктів з ГМО на ринку України варто обмежити.

Натуральні продукти зникли.

Варто агітувати населення країни щодо споживання натуральних продуктів.

Так як положення «Натуральні продукти зникли» має однакові значення навантажень як для фактора 1, так і для фактора 3, то воно включається в обидва фактори.

Останнім і вирішальним кроком факторного аналізу є виявлення і опис змістового зв'язку факторів.

Перший фактор зібрав усі положення, які агресивно налаштовані по відношенню до появи на вітчизняному ринку продуктів з вмістом ГМО.

Другий фактор об'єднує положення, які, в деякій мірі, підтримують обмежене споживання продуктів з вмістом ГМО.

До третього фактора увійшли точки зору, які, не відкидаючи споживання генно-модифікованих продуктів, підтримують споживання натуральних продуктів.

Відповідно до порядку висловлювань ці три фактори можна коротко охарактеризувати за допомогою наступних виразів: 1) негативне відношення, 2) підтримка обмеженого споживання продуктів з вмістом ГМО, 3) нейтральна позиція.

Як показує практика, не завжди фактори можна чітко пояснити. Якщо неможливо здійснити обґрунтоване пояснення факторів, то проведений факторний аналіз можна вважати невдалим.

Оскільки зробили розрахунок значень факторів, то відповідно до трьох відібраних автоматично утворилися три нові змінні, під назвою *fac1_1*, *fac2_1* і *fac3_1*, які містять обчислені значення факторів.

Факторні значення, як правило, лежать в межах від -3 до 3 і за своєю величиною пояснюють тісноту зв'язку кожної ознаки з відповідним фактором.

5.5. Основні вимоги до аналізу даних та формування звіту

Етап аналізу отриманої інформації – це зіставлення отриманої про вивчений об'єкт інформації з уже відомим об'ємом знань про нього.

Основною метою аналізу даних дослідження є пояснення змісту окремих результатів, об'єднання і виділення узагальнюючих положень, зведення їх в одну теоретичну систему.

Для отримання надійних і достовірних результатів емпіричного аналізу варто дотримуватися ряду вимог:

1) дослідник повинен мати уявлення про логіку використаних математичних методів;

2) застосовувати правильно підібрані математичні методи для аналізу даних досліджень, адже множина математично-статистичних методів буває достатньо різноманітною в залежності від типу дослідження: емпіричного, прикладного або теоретичного.

3) варто попередньо провести пробну обробку на невеликій кількості масиву даних.

У процесі аналізу та узагальнення результатів умовно можна виділити кілька етапів.

Перший – це етап впорядкування, класифікації, групування даних у відповідності з дослідницькими гіпотезами.

Другий етап аналізу та інтерпретації – це узагальнення даних, перевірка значущості і достовірності числових характеристик.

Третій етап – перевірка дослідницьких гіпотез за допомогою отриманих числових характеристик.

Результати дослідження, зокрема соціологічного, завжди відображаються у звіті. Звіт повинен бути основою для подальшої теоретичної роботи і формою впровадження результатів дослідження в практику.

Структура змісту звіту залежить від типу дослідження: теоретичного чи прикладного. У звіті описуються проблеми, цілі і завдання, об'єкт та предмет, інтерпретуються основні поняття, подається стан вивчення та дослідження проблеми у сучасній науці, обґрунтовується вибірка, методи збору інформації, аналізуються результати і визначається степінь розв'язання поставлених завдань.

Основні вимоги до звіту:

1. У звіті повинні бути відображені всі взаємопов'язані групи проблем у відповідності з логікою наукового пошуку.

2. Кожний розділ звіту має складатися із двох частин: у першій – проблеми та результати, у другій – висновки.

3. Звіт формується незалежно від послідовності запитань в анкеті чи опитуванні.

4. Правильне оформлення звіту.

Звіт має:

– містити кількісні параметри вивченого об'єкта;

– описувати основні тенденції і темпи розвитку явищ;

– розкривати взаємозв'язок між ознаками та явищами за допомогою отриманих показників.

Універсального рецепту для правильної інтерпретації кількісних показників у науці немає. Це залежить від професіоналізму, компетентності, культури, ціннісних орієнтацій і установок дослідника, науковця.

До звіту, як правило, додається пояснювальна записка, в якій описуються результати, і додатки з таблицями, графіками.

Варто пам'ятати, що великий об'єм цифрового матеріалу дезорієнтує споживача інформації, тому ефективнішою є графічна інтерпретація результатів дослідження.

5.6. Професійне маніпулювання результатами дослідження

Дослідник-практик перед проведенням спостереження або експерименту періодично вивчає додаткові графіки та діаграми допоміжних матеріалів. Вивчаючи готову інформацію, можна проаналізувати помилки, неточності та непорозуміння, які були допущені попередніми дослідниками. Це, з однієї сторони, змушує уважніше відноситись до представлення результатів дослідження, а з іншої, набувати досвіду маніпулювання даних для ефективного відображення необхідних висновків. Адже не виключено, що коли-небудь доведеться виконати роботу, результати якої можуть зачепити чий-небудь інтереси або стануть предметом гострих суперечок.

Правильно відображена вихідна інформація допоможе уникнути звинувачень в упередженості і некомпетентності зі сторони осіб, зацікавлених в протилежних результатах дослідження.

Відомо, що методи графічного відображення даних часто стають об'єктами свідомої або неусвідомленої фальсифікації. Так, яскраво виражені негативні дані або, навпаки, дані, що навіюють нестримний оптимізм, передбачають упушення в графіку (таблиці) якого-небудь фактора, що істотно впливає на результати.

Маневрування статистичними даними часто виникає від незнання прийомів статистики та невміння правильно використовувати відповідні методи. Особливо часто це спостерігається в журналістиці, зокрема, в період передвиборних перегонів.

Річ у тому, що основна проблема полягає не в статистичних даних як таких, а в їх інтерпретації. Одну і ту ж вихідну інформацію можна по-різному тлумачити – від критичної оцінки до схвальної. Такі трактування базуються на прийомах впливу на свідомість публіки. У своїй основі ці прийоми зводяться до декількох способів роботи з матеріалами.

Для простоти та чіткості ми проілюструємо різні способи маніпулювання графічно представленою статистичною інформацією.

Спосіб перший: приховування одиниць вимірювання. Яскравим прикладом використання даного способу можуть бути результати дослідження неспішності студентів. Статистика кількості незадовільних оцінок з вищої математики у деякому ВУЗі показує жахливі результати (табл. 5.24, рис. 5.5).

Таблиця 5.24

Відомості про незадовільні результати здачі іспитів з вищої математики

	2004	2005	2006	2007	2008
Технічні спеціальності	90	93	107	123	165
Економічні спеціальності	85	86	90	97	102
Гуманітарні напрями	67	68	72	89	96
Всього	242	247	269	309	363

Візуальна інтерпретація даних формує лише жахливе уявлення про якість знань студентів. Адже, починаючи з 2005 р. крива графіка стрімко зростає. Навіть у випадку, якщо графічний матеріал супроводжується відповідною таблицею (як це зроблено в даному випадку), більшість читачів орієнтується тільки на графік.

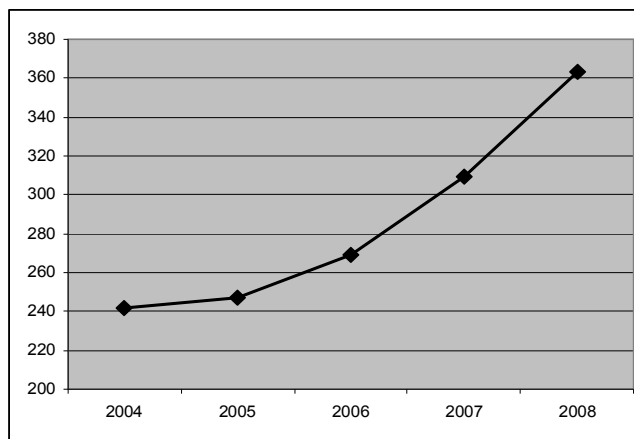


Рисунок 5.5. Результати здачі іспитів з вищої математики: незадовільні оцінки

Причин такого негативного кількісного аналізу успішності з вищої математики може бути кілька: опущено фактори, які впливають на значення показників (загальна кількість студентів ВУЗу зросла); некоректно вибрано одиниці вимірювання.

При представленні результатів необхідно доповнити таблицю хоча б відносними одиницями виміру.

Вводячи в таблицю 1 додатковий параметр – відношення кількості негативних до загальної кількості усіх оцінок з предмету (у відсотках), ми отримуємо реальнішу ситуацію досліджуваної ознаки (табл. 5.25).

Таблиця 5.25

Відомості про незадовільні оцінки з вищої математики

	2004		2005		2006		2007		2008	
Технічні спеціальності	90	8,6%	93	9%	107	9,5%	123	8,8%	165	9,2%
Економічні спеціальності	85	10%	86	9,5%	90	9,5%	97	9%	102	9%
Гуманітарні напрями	67	12%	68	10%	72	11%	89	12%	96	11%
Всього	242		247		269		309		363	

Різкий приріст показників неспішності пояснюється збільшенням кількості студентів у ВУЗі. Обчисливши середнє значення відсотків зі спеціальностей для кожного року, отримуємо такі дані:

$$\frac{8,6 + 10 + 12}{3} = 10,2; \quad \frac{9 + 9,5 + 10}{3} = 9,5; \quad \frac{9,5 + 9,5 + 11}{3} = 10; \quad \frac{8,8 + 9 + 12}{3} = 9,9;$$

$$\frac{9,2 + 9 + 11}{3} = 9,7.$$

Вони не лише не ідентичні попереднім, а й показують абсолютно інші тенденції змін неспішності студентів за роками (рис. 5.6).

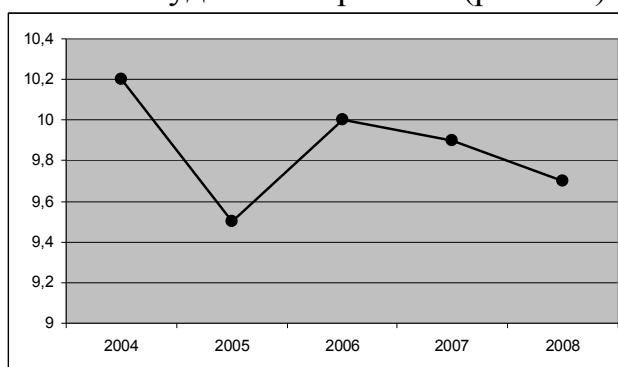


Рисунок 5.6. Відносні частки незадовільних оцінок з вищої математики по роках

Доповнюючи емпіричні дані одним показником, ми змінили характер поведінки вивченої ознаки. Хоча теоретично створити аналітичний матеріал, який усесторонньо схарактеризує об'єкт і предмет дослідження – це мистецтво досвідченого дослідника-практика.

Другий спосіб полягає в умінні маніпулювання осями графіків. Два наступні графіки показують, як по-різному може сприйматися одна і та ж інформація (рис. 5.7).

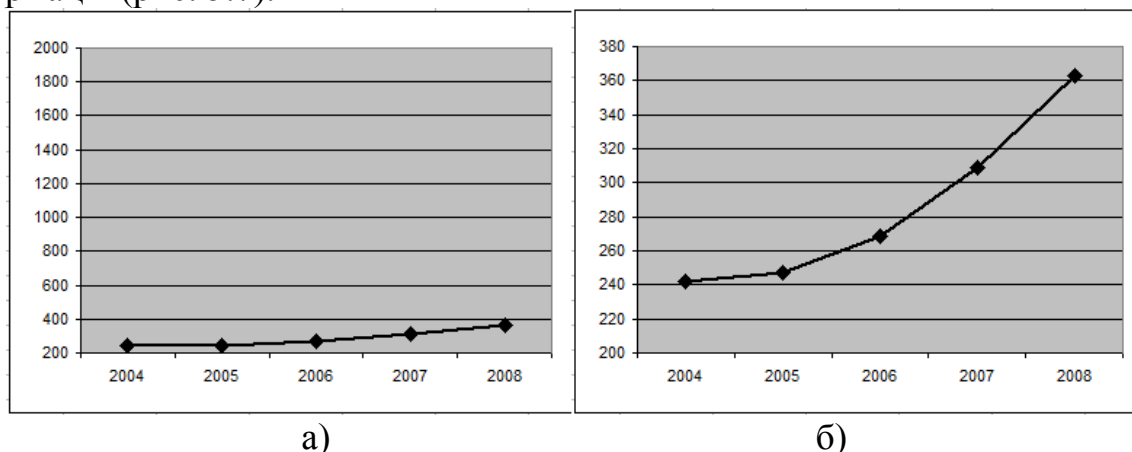


Рисунок 5.7. Результати здачі іспитів з вищої математики: незадовільні оцінки

Графіки побудовані на основі ідентичних даних, а їх візуальне зображення інтерпретується по-різному. На рис. 5.7а) ми спостерігаємо допустиму тенденцію росту кількості незадовільних оцінок протягом п'яти років. На відміну від цього рис. 5.7б) демонструє катастрофічну ситуацію спаду успішності студентів. Такі результати отримані шляхом стискування-розтягування осі ординат відповідного графіка.

Інколи використовується масштабування, яке теж допомагає маніпулювати результатами, гіперболізуючи або мінімізуючи їх. Знявши числові поділки на шкалі можна досягнути ще кращого ефекту. Максимально звужуючи вісь OX та розтягуючи OY , спостерігатимемо тенденцію до деякого збільшення числового показника – це стане картиною нестримного росту (катастрофічного падіння) характеристики досліджуваної ознаки.

Третій спосіб передбачає усвідомлений або випадковий аналіз частини даних для отримання бажаних для замовника результатів. Тут часто керуються принципом «якщо факти не підходять до теорії – тим гірше для фактів» [1, 178].

За допомогою вищеописаного прикладу графічно продемонструємо спад відносної частки незадовільних оцінок студентів з вищої математики (рис. 5.8). Звідси можна прийти до висновку, що із кожним навчальним роком рівень успішності студентів даного ВУЗу зростає.

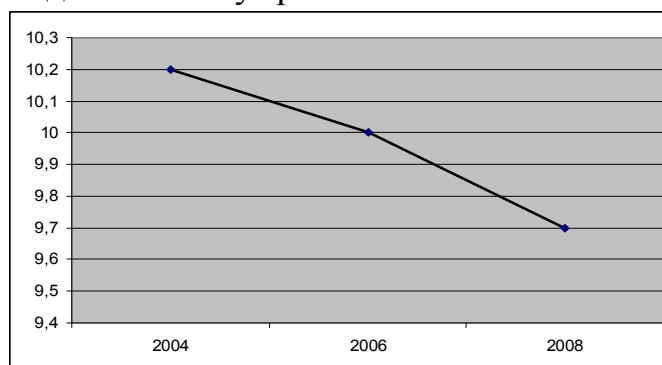


Рисунок 5.8. Відносні частки незадовільних оцінок з вищої математики

Таким чином, дослідник може підвищити репутацію та рейтингову оцінку навчального закладу.

Отже, ми на одному прикладі продемонстрували різні можливості інтерпретації результатів дослідження. Це ні в якому разі не фабрикування результатів, а лише незначне маневрування значеннями і графіками. Адже, раціональність мислення людини часто поступається емоційному сприйманні навколишньої дійсності.

Питання для самоконтролю

1. Що таке тренд?
2. Що таке сезонна варіація?
3. Як розраховується ковзке середнє?
4. Який алгоритм побудови адитивної моделі?
5. Який алгоритм побудови мультиплікативної моделі?
6. Який алгоритм побудови експоненційного згладжування?
7. Як здійснюється прогноз за методом експоненційного згладжування?
8. Яким чином константа згладжування впливає на прогноз?
9. Як розраховується помилка прогнозу?
10. Як обґрунтовується оптимальність моделі для прогнозування?
11. Що таке фактор?
12. Що таке факторне навантаження?
13. Чим відрізняються загальні та спеціальні фактори?
14. Які методи оцінки зв'язків факторного аналізу ви знаєте?
15. Які методи обертання факторів використовую статистичні пакети програм?
16. Що таке інтепретація факторів?
17. Для чого знаходяться факторні значення?
18. Які вимоги до звіту ви знаєте?

ЛІТЕРАТУРА

1. Агабекян Р. Л. Математические методы в социологии. Анализ данных и логика вывода в эмпирическом исследовании: учеб. пособ. для вузов / Р. Л. Агабекян, М. М. Кириченко, С. В. Усатилов. – Ростов н/Д: Феникс, 2005. – 192 с.
2. Бутник О. М. Економіко-математичне моделювання перехідних процесів у соціально-економічних системах: [монографія] / Бутник О. М. – Х.: Видавничий Дім „ИНЖЕК”; СПД Лібуркіна Л. М., 2004. – 304 с.
3. Бююль А. SPSS: искусство обработки информации. Platinum Edition, пер. с нем. / А. Бююль, П. Цёфель. – СПб.: ООО «ДиаСофтЮП», 2005. – 608 с.
4. Валентинов В. А. Эконометрика: практикум / Валентинов В. А. – М.: РДЛ, 2007. – 436 с.
5. Дослідження операцій: Навч. посіб. / М. Г. Медведєв, О. В. Колодінська. – [2-ге вид., перер. і доп.]. – К.: Вид-во Європ. ун-ту, 2006. – 158 с.
6. Дубина А. Г. Excel для экономистов и менеджеров / А. Г. Дубина, С. С. Орлова, И. Ю. Шубина, А. В. Хромов. – СПб.: Питер, 2004. – 295 с.
7. Екимов С. В. Нетрадиционные подходы в экономико-математическом моделировании: [монография] / Екимов С. В. – Днепропетровск: Наука и образование, 2004. – 240 с.
8. Ермолаев О. Ю. Математическая статистика для психологов: [учебник] / Ермолаев О. Ю. – [2-е изд. испр.]. – М.: Московский психолого-социальный институт Флинта, 2003. – 336 с.
9. Карагодова О. О. Дослідження операцій: навч. посіб. / О. О. Карагодова, В. Р. Кігель, В. Д. Рожок. – К.: Центр учбової літератури, 2007. – 256 с.
10. Лапач С. Н. Статистика в науке и бизнесе / С. Н. Лапач, А. В. Чубенко, П. Н. Бабич. – К.: МОРИОН, 2002. – 640 с.
11. Макаренко Т. І. Моделювання та прогнозування у маркетингу: навч. посіб. / Макаренко Т. І. – К.: Центр навчальної літератури, 2005. – 160 с.
12. Минько А. А. Статистический анализ в MS Excel / Минько А. А. – М.: Изд. дом «Вильямс», 2004. – 448 с.
13. Наследов А. Д. SPSS: Компьютерный анализ данных в психологии и социальных науках / Наследов А. Д. – СПб.: Питер, 2005. – 416 с.
14. Невежин В. П. Сборник задач по курсу «Экономико-математическое моделирование» / В. П. Невежин, С. И. Кружилов. – М.: ОАО «Изд. дом „Городец”», 2005. – 320 с.
15. Просветов Г. И. Эконометрика: Задачи и решения: учебно-метод. пособ. – [4-е изд., доп.] / Просветов Г. И. – М.: Издательство РДЛ, 2007. – 192 с.
16. Таха Х. Введение в исследование операций. – [7-е изд.] / Таха Х., пер. с англ. – М.: Изд. дом "Вильямс", 2007. – 912 с.
17. Шимко П. Д. Статистика / П. Д. Шимко, М. П. Власов. – Ростов н/Д: Феникс, 2003. – 448 с. – [Серия «Учебники, учебные пособия»].
18. Экономико-математические методы и прикладные модели: учеб. пособ.; Под ред. В. В. Федосеева. – М.: ЮНИТИ, 1999. – 321 с.

Навчальне видання

**ВАСИЛЕНКО Оксана Анатоліївна,
СЕНЧА Ірина Анатоліївна**

Математично-статистичні методи аналізу в прикладних дослідженнях

Навчальний посібник

Редактор

Г. Ю. Греля

Комп'ютерне верстання

Є. С. Корнійчук

Здано в набір 9.04.2012 Підписано до друку 10.05.2012.

Формат 60/88/16 Зам. № 4847.

Тираж 100 прим. Обсяг: 10,5 ум. друк. арк.

Віддруковано на видавничому устаткуванні фірми RISO
у друкарні редакційно-видавничого центру ОНАЗ ім. О.С. Попова
ОНАЗ, 2012