УДК 330.4+519.22

МАТЕМАТИКО-СТАТИСТИЧНІ МЕТОДИ У МОНІТОРИНГОВИХ ЕКОНОМЕТРИЧНИХ ДОСЛІДЖЕННЯХ

І. В. Степахно, кандидат фізико-математичних наук, доцент

Національний університет біоресурсів та природокористування України

e-mail: stir@ukr.net

Анотація. Моніторинг якості в сучасних умовах проводиться з використанням математико-статистичних методів. За допомогою математико-статистичних методів дослідження проводиться обробка великих масивів інформації та встановлюються причинні залежності. Використання статистичних методик сприяє формуванню здатності читати звіти про статистичні дослідження у галузі освіти та економіки, навчає техніці аналізу отриманих даних. У статті застосовано метод багатовимірної статистичної моделі, розглянуто визначення ідентифікації математико-статистичної моделі та розраховані коефіцієнти впливу параметрів моделі на потрібні значення економічних характеристик складної системи. Основним інструментом обробки вхідної інформації є апарат теорії випадкових матриць. Побудова математичних моделей в сучасних умовах при застосуванні дуже великої кількості параметрів потребують розробки таких підходів, у яких можливо врахувати всі необхідні ситуації, не втрачаючи важливої вхідної інформації.

Ключові слова: багатовимірний статистичний аналіз, ймовірність, математична модель, випадкова матриця

Актуальність. Розробка статистичних методів дуже часто потребує отримання об'єктивної інформації для вибору способу прогнозування та оцінювання ситуації. Ставиться задача: показати переваги застосування деяких ймовірносних підходів на прикладах складної економічної системи (оцінки якості освіти).

Аналіз останніх досліджень та публікацій. Теоретико-ймовірносний підхід до рішення прикладних задач, в яких кількість параметрів велика, розглядався у працях В.Л. Гирко [1-4]. Збудований апарат в теорії випадкових матриць апробовано на безрозмірних вхідних даних і дослідження проводилися на прикладах складних математичних моделей.

Мета дослідження. Існує дуже багато прикладних задач у житті суспільства, в яких без використання побудови моделей практично неможливо безпомилково прорахувати множину показників. Оцінювання якості освіти та прогнозування є актуальною задачею, для рішення якої необхідно застосовувати науковий метод математичної статистики. Однак, розглядаючи данні, стає зрозумілим, що вони мають часові та просторові обмеження. Багатовимірність економічних та технічних показників кожної одиниці ускладнює інтуїтивне оцінювання ситуації у кожному конкретному випадку. Проводити багато іспитів не вигідно, і параметри характеристик надходять з відхиленнями. Стає зрозумілим, що застосування багатовимірного статистичного аналізу дозволяє врахувати всі ці складності. Пропонується провести статистичну обробку великої кількості спостережень з урахуванням ймовірносних помилок у рамках припущених обмежень. Для більш точної ідентифікації запропонованої математико-статистичної моделі необхідно зібрати всю інформацію в різні моменти та записати її у відповідну матрицю. Цю інформацію потрібно враховувати постійно.

Матеріали та методи дослідження. Нехай $X_i, i = \overline{1, s}$ – незалежні спостереження над випадковою матрицею $A + \Xi$, де $A = (a_{ij})$ – дійсна матриця, $i = \overline{1, n}, j = \overline{1, m}, \Xi = (\xi_{ij})$ – випадкова матриця тієї ж розмірності.

Позначимо через λ_k сингулярні власні числа матриці A , а через $\mathcal{E}_k$ – сингулярні власні числа матриці $\mathcal{A} = s^{-1} \cdot \sum_{i=1}^s X_i$. Очевидно, що, якщо елементи матриці Ξ незалежні, мають нульові середні та дисперсії $s^{-1}\sigma^2$, то елементи матриці $\mathcal{A}$ також будуть незалежними та матимуть дисперсії $s^{-2}\sigma^2$.

Нехай числа m, n, σ^2, s залежні та виконуються наступні умови :

$$\lim_{m \rightarrow \infty} \sigma^2 s^{-1} n < \infty, \quad \lim_{m \rightarrow \infty} \sigma^2 s^{-1} m < \infty, \quad \lim_{m \rightarrow \infty} mn^{-1} < 1, \quad \lim_{m \rightarrow \infty} mn^{-1} > 0. \quad (1)$$

$$\mathcal{E}_k(A) \leq C < \infty, \quad C = \text{const}, \quad k = 1, 2, \dots, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m. \quad (2)$$

Розглянемо сингулярні спектральні функції

$$\mathcal{F}_m(x) = m^{-1} \sum_{k=1}^m \chi(\mathcal{F}_k < x),$$

де $\mathcal{F}_k$ – корені характеристичного рівняння $\det(I \cdot \lambda - \sqrt{\mathcal{F}'\mathcal{F}}) = 0$.

Вектор-стовбці матриці $\mathcal{F}'$ стохастично незалежні. Для матриці $\mathcal{F}'\mathcal{F}$ доведена наступна теорема.

Теорема. Нехай виконуються умови (1), (2), а випадкові елементи ξ_{ij} матриці Ξ для кожного n незалежні, $M\xi_{ij} = 0$, $D\xi_{ij} = s^{-1}\sigma^2$. Тоді

$$\text{plim}_{m \rightarrow \infty} m^{-1} \text{sP} \left[\left(-Iz + \mathcal{F}'\mathcal{F} \right)^{-1} - M \left(-Iz + \mathcal{F}'\mathcal{F} \right)^{-1} \right] = 0.$$

Доведення цієї теореми базується на стандартних формулах збурень для резольвент матриць [1].

Використання математичної статистики при проведенні моніторингових досліджень у природничих та соціальних галузях умовно поділяється на періоди:

1. Опис статистичних даних – XVII - XVIII ст.
2. Аналіз статистичних закономірностей у природничих і соціальних науках – початок XIX ст.
3. Використання математичної статистики для прогнозування суспільних подій – початок XX ст.
4. Розробка і використання апарату прикладної статистики, поширення комп'ютерних засобів – кінець XX, початок XXI ст.

Сучасні розробки математичного апарату прикладної статистики пов'язані з використанням комп'ютерних засобів для розв'язання різних типів задач за якими можна порівнювати декілька програм за показниками оцінок успішності учнів в досліджуваних класах, встановлювати лінію передбачення коефіцієнта інтелекту (IQ) учнів ЗНЗ (WAIS) та учителів (WISC) за шкалами Стенфорда – Біне, Векслера та ін. До загального класу задач математичної статистики відносять задачі статистичного висновку, задачі оцінювання для вимірювання нелінійних зв'язків

між змінними, одно і двофакторного дисперсійного аналізу (ANOVA), множинних порівнянь у двофакторному аналізі та ін.

У більшості наукових досліджень використовують задачі статистичного висновку за допомогою емпіричної функції розподілу. Задачі статистичного висновку використовують для дослідження первинного статистичного матеріалу. Вони передбачають дослідження великих класів предметів у порівнянні з малими класами. Задачі статистичного висновку – це міркування від часткового до загального. Після чого робиться висновок і наводяться міркування у зворотному напрямі від загального до часткового (сілогічне міркування). Всяка велика сукупність предметів, яку ми хочемо досліджувати або відносно якої ми робимо висновок, називається генеральною сукупністю. Термін „генеральна сукупність” набуває смислу при визначенні поняття „вибірка із сукупності”. Вибірка із сукупності є частиною або підмножиною сукупності яку ми проводимо спеціально для визначення властивостей генеральної сукупності. Значення різноманітних описових вимірів, обчислених для генеральної сукупності, називаються параметрами. Параметр описує вибірку так само, як проводиться статистична обробка вибірки. Для вибірок ті ж самі описові виміри називають *статистиками*. Отже, параметр описує сукупність так само, як статистика – вибірку. Статистику обчислену за вибіркою розглядають як оцінку параметра сукупності. Оцінка параметра сукупності це величина, що дає інформацію про параметр; оцінювач – деяка функція від значення у вибірці. Загальна характеристика емпіричного розподілу. Емпіричний розподіл є формою подання результатів дослідження або первинного статистичного матеріалу. Проста статистична сукупність – це сукупність спостережуваних значень – x називається простою сукупністю. Вона є первинним емпіричним матеріалом, що оформлюється у вигляді первинного статистичного ряду, який зводиться у таблицю, де $x \dots x_i$ відповідає порядковий номер.

Припустимо, що ми проводимо моніторингове дослідження, в якому проводимо операції над групою чисел (параметрів) за 11 бальною шкалою. (від 0 до

10). Кожен з параметрів характеризує якість підручника. Бал 0 ставиться, якщо даний показник якості мінімальний; 1 бал, якщо в підручнику наявні окремі елементи відповідного параметра; бали 2– 9 ставляться залежно від повноти реалізації показників якості підручника; 10 балів, якщо даний показник якості послідовно, систематично реалізується в навчальному матеріалі підручника.

Множина чисел розташовується в таблиці з рядками і стовпчиками, де кожен елемент такого розташування описується за значенням груп стовпців і їх розміщення в рядку $x \dots x_i$. Операції над групою чисел проводимо за допомогою символів для введення формули в Excel. Послідовність $x \dots x_i$ представляє собою групу із n чисел, кожне число якої можна записати як X_i .

Емпіричний розподіл подається у вигляді первинного статистичного ряду, який зводиться у таблицю (порядковому номеру досліду і відповідає значення x_i). Упорядковані за значенням результати спостережень $x_1 x_2 \dots x_i$, де $x_i < x_{i+1}$, називають варіаційним рядом або рядом розподілу, а окремі значення $x_1 x_2 \dots x_n$ – варіантами варіаційного ряду. Різниця між найбільшим x_n і найменшим x_1 варіантами варіаційного ряду має назву розмаху емпіричного розподілу та визначаються як $R = x_n - x_1$; Розгляд властивостей явищ, суспільних процесів можна проводити за допомогою математико-статистичного методу сукупності даних.

Сукупності даних (вибірки) поділяються на:

- генеральні сукупності;
- якісні генеральні сукупності ;
- кількісні генеральні сукупності;
- вибіркові сукупності ;
- репрезентативні сукупності даних.

Генеральною сукупністю даних називають множини, які включають у себе всі однорідні об'єкти, котрим притаманні або не притаманні певні якості та кількісні ознаки.

Якісні генеральні сукупності це сукупності, де властивості якостей одиниць відбору є ознаками.

Вибірковою сукупністю (вибіркою) вважається частина генеральної сукупності.

Кількісною генеральною сукупністю називають таку сукупність, де властивості відбору оцінюються змінними. Змінна – математична величина яка може набувати в перебігу дослідження різних значень.

1. Аналіз і інтерпретація кількісних даних

Учень	Оцінка (бали)	Учень	Оцінка (бали)
1.	90	13	100
2.	60	14	75
3.	100	15	71
4.	86	16	93
5.	57	17	91
6.	78	18	68
7.	44	19	99
8.	69	20	93
9.	46	21	80
10.	55	22	67
11.	58	23	89
12.	81	24	79

Репрезентативною до генеральної сукупності є така вибірка, що має пропорції, відповідні до пропорцій генеральної сукупності. Для репрезентативності вибірки мають бути виконані умови:

перша умова – кожна з одиниць генеральної сукупності повинна мати рівну імовірність потрапляння до вибірки;

друга умова - вибірка змінних здійснюється незалежно від досліджуваної ознаки;

третя умова – відбір здійснюється при достатньому обсязі вибірки;

четверта умова – відбір здійснюється з однорідних за статистичними характеристиками сукупностей (близькими показниками середніх або однаковими).

Забезпечення умов репрезентативності вибірки виключає упередженість відносно генеральної сукупності.

Аналіз і інтерпретацію кількісних даних розпочинаємо з їх узагальнення. Наприклад, приводимо результати анкетування учнів, проведеної на початку навчального року (табл. 1).

Упорядкування представлених даних за величиною від максимальної до мінімальної називають не згрупованим рядом. Проводимо інтерпретування кожної оцінки в термінах рангів (табл. 2). Вірогідно, що мають місце і рівні оцінки. У такому випадку присвоюють їм однакові ранги.

2. Інтерпретування оцінки в термінах рангів

Учень	Оцінка (бали)	Ранг	Учень	Оцінка (бали)	Ранг
1	90	5	13	100	(2)1
2	60	16	14	75	11
3	100	(2)1	15	71	12
4	86	7	16	93	(2)3
5	57	18	17	91	4
6	78	10	18	68	14
7	44	21	19	99	2
8	69	13	20	93	(2)3
9	46	20	21	80	(2)8
10	55	19	22	67	15
11	58	17	23	89	6
12	80	(2) 8	24	79	9

3. Упорядкування за величиною(ранжування і табулювання)

Зменшувана послідовність	Ранг	Табулювання без наступного групування	
		Оцінка	Частота
100	1 (2)	100 (max)	2
99	2	99	1
93	3(2)	93	2
91	4	91	1
90	5	90	1
89	6	89	1
86	7	86	1
80	8	80	2
81	9	81	1
79	10	79	1
78	11	78	1
75	12	75	1
71	13	71	1
69	14	69	1
68	15	68	1
67	16	67	1
60	17	60	1
58	18	58	1
57	19	57	1
55	20	55	1
46	21	46	1
44	22	44 (min)	1

Вибірковий розподіл належить до понять теорії статистичного висновку. Із деякої генеральної сукупності виводиться вибірка об'єму (n) і вводяться записи для кожної із вибірок за деякими оцінками вибіркового середнього ($\bar{x}_b$). Так можна виводити сотні вибірок і будувати розподіл частот вибірових середніх у вигляді рисунка розподілу ймовірностей в генеральній сукупності.

Випадкова величина (X) може приймати значення від 0 до 9 з ймовірністю 0,10.

Із цієї генеральної сукупності виводимо кількість випадкових вибірок об'ємом $n = 2$. Для кожної вибірки вираховується середнє, рівне $(3+2) : 2 = 2,5$. Далі виводимо наступні вибіркові середні і представляємо їх графічно.

Використовуючи теорему статистичного висновку можна проводити вибірки із великої сукупності. Середня цієї сукупності позначається μ , а дисперсія δ^2 . Наприклад, із генеральної сукупності беруться вибірки об'єму n . Проводиться обчислення середніх значень всіх вибіркових середніх μ яке дорівнюватиме середньому значенню генеральної сукупності. Далі необхідно знайти дисперсію δ^2 вибіркових значень.

Результати досліджень та їх обговорення. Отримані рекомендації щодо застосування математико-статистичних методів у економічних показниках якості.

Застосовуючи апарат багатовимірного статистичного аналізу та метод головних компонент, можна встановити, які коефіцієнти в більшій степені впливають на шукану характеристику та в який бік – в бік зростання чи спадання. Оцінки параметрів визначалися за власними значеннями кореляційної матриці, розрахованої за вхідними даними.

Висновки та перспективи. Якщо проводити спостереження $X_1, X_2, \dots, X_s$ які відображають велику кількість характеристик, то без застосування методів багатовимірного статистичного аналізу, отримати якісні оцінки та прогноз неможливо. Отримані оцінки дозволяють забезпечити процес ідентифікації математичної моделі та оптимізувати матеріальні витрати.

Список літератури

1. Гирко В.Л. G-оценка сингулярных собственных чисел матриц / В. Л. Гирко, И. В. Степахо // Докл. АН УССР. – 1990. – №8, серия А. – С. 14–17.
2. Girko, V. L. Theory of Linear Algebraic Equations with Random Coefficients / V. L. Girko. – New York, 1996. – 302 с.
3. Girko, V. L. An Introduction to Statistical Analysis of Random Arrays. VSP, 1998. – P. 5–11.

4. Girko, V. L. Theory of stochastic canonical equations. Kluwer Publishers (Netherlands), v. I , v. II , 2001. – 316 p.

References

1. Girko, V. L., Stepahno, I. V. (1990). G–ozenka singulyarnykh sobstvennykh chisel matriz. [G-estimate of matrices singular eigenvalues]. Doklady AN USSR, 8(A), 14 – 17.

2. Girko, V. L. (1996). Theory of Linear Algebraic Equations with Random Coefficients. New York, 302.

3. Girko, V. L. (1998). An Introduction to Statistical Analysis of Random Arrays. VSP, 5 - 11.

4. Girko, V.L. (2001). Theory of stochastic canonical equations. Kluwer Publishers (Netherlands), I , II , 316.

MATHEMATIC-STATISTICAL ANALYSIS FOR THE IDENTIFICATION OF THE MATHEMATICAL MODEL IN APPLIED PROBLEMS WITH THE PURPOSE OF COST SAVINGS

I. Stepakhno

Abstract. The method of statistical analysis has been provided, considered the notion of identification, the influence coefficients of model parameters on determined values of maneuvering and economic characteristics were calculated,

Keywords: *multi-statistical analysis, probability, mathematical model, stochastic matrix*