УДК 656.13.05+629.114.5+519.23

Т. Є. Василенко, канд. екон. наук, доц., Є. О. Корольов, канд. фіз.-мат. наук, доц., М. Є. Корольов, канд. фіз.-мат. наук, доц., А. І. Мельник

Автомобільно-дорожній інститут ДВНЗ «Донецький національній технічний університет», м. Горлівка

ПОДАННЯ ТА ПЕРВИННА ОБРОБКА СТАТИСТИЧНИХ ДАНИХ У БАГАТОВИМІРНОМУ ШКАЛЮВАННІ. КЛАСИЧНА МОДЕЛЬ БАГАТОВИМІРНОГО ШКАЛЮВАННЯ ТОРГЕРСОНА

Сформульовано основні групи показників, які впливають на якість транспортного обслуговування пасажирських перевезень. Запропоновано класичну модель багатомірного шкалювання Торгерсона, за допомогою якої дано оцінку якості транспортного обслуговування. Виділено об’єкти, на яких виконувалася обробка статистичних даних. Використовуючи модель Торгерсона, побудувано симетричну матрицю з подвійним центруванням, із подальшим застосуванням методу головних факторів побудовано факторний простір зі спостережуваними об'єктами.

Постановка наукової проблеми та задачі, що вирішується

Підвищення якості роботи міського пасажирського транспорту має найважливіше соціальне й економічне значення. У місті Горлівка пасажири, як і в Україні в цілому, переміщуються в основному громадським міським транспортом. При цьому автобусний транспорт відіграє основну роль у задоволенні попиту на перевезення пасажирів у містах.

Встановлюється наступна номенклатура основних груп показників якості, за якими характеризуються споживчі властивості пасажирських перевезень:

– комфортність поїздки;
– безпека транспортного обслуговування;
– якість організації транспортних послуг.

Надалі в статті за допомогою цих груп показників проводиться оцінка якості транспортного обслуговування пасажирів міським транспортом [1, 2].

Мета роботи

Метою роботи є застосування класичної моделі багатомірного шкалювання Торгерсона при обчисленні показників підвищення якості транспортного обслуговування пасажирів автобусним транспортом.

Основна частина

Дамо оцінку якості транспортного обслуговування пасажирів міським транспортом із застосуванням моделі багатомірного шкалювання Торгерсона. Обробку статистичних даних будемо проводити за такими об’єктами: автобуси моделей Богдан та ПАЗ, різних років випуску.

Використовуючи дані таблиці 1 та застосовуючи модель Торгерсона, необхідно побудувати симетричну матрицю з подвійним центруванням, із подальшим застосуванням методу головних факторів [3, 4].

Етапи розв'язання задачі:

1. Стандартизація вихідних даних.
2. Побудова матриці відмінностей за стандартизованими даними.
3. Застосування класичної моделі багатовимірного шкалювання Торгерсона: обчислення «середніх» характеристик, перетворення матриці відмінностей в матрицю з подвійним центруванням.

	Ознака		
Об'єкт	комфорт	безпека	якість
ПАЗ, 2012 рік випуску	11,8	429,6	63,4
Богдан, 1999 рік випуску	2	88,8	10
ПАЗ, 2010 рік випуску	2	124,6	11,5
Богдан, 2010 рік випуску	3,1	17,1	26,8
Середнє значення	4,6	165,025	27,95

Розв'язок:

1. Стандартизуємо вихідні дані:

$$z_{ij} = v_{ij} / \bar{v}_{j}, \tag{1}$$

$$z_{ij} = \begin{pmatrix} 2,57 & 2,6 & 2,27 \\ 0,33 & 0,54 & 0,36 \\ 0,43 & 0,76 & 0,41 \\ 0,67 & 0,10 & 0,96 \end{pmatrix}.$$

2. Будуємо матрицю мір відмінностей профілей.

Під профілем розуміється простий набір кількісних ознак об'єкта [5]. Набір кількісно певних даних по j-им ознакам для i-го об'єкта. Рядки – профіль оцінок об'єкта. Іншими словами, визначено відстані між профілями.

$$\Delta = \begin{pmatrix} 0 & 3,593755 & 3,37678 & 3,39704 \\ 3,593755 & 0 & 0,247761 & 0,816552 \\ 3,37678 & 0,247761 & 0 & 0,883844 \\ 3,39704 & 0,816551 & 0,883844 & 0 \end{pmatrix}.$$

Визначаємо значення елементів матриці з подвійним центруванням $\Delta^* = \left(\delta_{ij}^*\right)$:

$$\delta_{ij}^* = -\frac{1}{2}\left(\delta_{ij}^2 - \delta_{\cdot i}^2 - \delta_{\cdot j}^2 + \delta_{\cdot\cdot}^2\right), \tag{2}$$

де $\delta_{\cdot i}^2 = \frac{1}{j}\sum_j \delta_{ij}^2$ – середня для характеристик відмінностей в j-х стовбцях i-ої строки, що зведені в квадрат;

$\delta_{i}^2 = \frac{1}{j}\delta_{ij}^2$ – середня для характеристик відмінностей в i-х строках j-го стовпця, що зведені в квадрат;

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

Матриця відмінностей		$\delta^2_{\cdot i}$
$\Delta = \begin{pmatrix} 0 & 3,594 & 3,377 & 3,397 \\ 3,594 & 0 & 0,248 & 0,817 \\ 3,377 & 0,248 & 0 & 0,884 \\ 3,397 & 0,817 & 0,884 & 0 \end{pmatrix}$		8,964 3,411 3,061 3,247
$\delta^2_{\cdot j}$	8,964 3,411 3,061 3,247	$\delta^2_{\cdot i} = \delta^2_{\cdot j}$

$$\delta^2_{\cdot\cdot} = 4,671.$$

Для матриці з подвійним центруванням застосовано класичну модель головних чинників із кореляційною матрицею:

$$R_h = \Delta^*. \tag{3}$$

Матриця відмінностей		серед. стр.
$\Delta^* = \begin{pmatrix} 6,629 & -2,605 & -2,024 & -2,000 \\ -2,605 & 1,075 & 0,870 & 0,660 \\ -2,024 & 0,870 & 0,726 & 0,428 \\ -2,000 & 0,660 & 0,428 & 0,912 \end{pmatrix}$		0 0 0 0
серед. стол.	0 0 0 0	

Зауваження 1: для перевірки матриці з подвійним центруванням – середні значення елементів кожного її рядка та кожного стовбця дорівнюють нулю.

Зауваження 2: теоретично метод Торгерсона базується на жорстких гіпотетичних припущеннях:

– у деякому певному шкальному просторі X відстані між об’єктами, що спостерігаються, відповідають величинам, які характеризують їх відмінності, тобто

$$\delta_{ij} = d_{ij} \tag{4}$$

– самі відстані між об'єктами в теоретичному просторі досить точно описуються метрикою Евкліда:

$$\delta_{ij} = d_{ij} = \left(\sum_k x_{ik} - x_{jk} \right)^{1/2}; \tag{5}$$

– у шкальному просторі X середні значення координат стимулів по кожній осі дорівнюють нулю, нуль – вихідна точка відліку:

$$\frac{1}{i}\sum_i x_{ik} = \frac{1}{j}\sum_j x_{jk}; \tag{6}$$

– алгоритм Торгерсона мінімізує міру відповідності:

$$F = \sum_{i,j} \left(\delta^*_{ij} - \sum_k x_{ik} x_{jk} \right)^2 \to \min, \tag{7}$$

тобто сума квадратів різниць характеристик відмінностей об'єктів і відстаней між об'єктами в деякому теоретично певному нормованому шкальному просторі X – повинна бути мінімальною, тобто витримується вимога методу найменших квадратів.

Далі для матриці з подвійним центруванням застосовано метод головних факторів (алгоритм Хоттеллінга).

Таблиця 1 – Результати застосування алгоритму Хоттеллінга

Ознака	R_h				$S_i^{(1)}=\sum_j r_{ij}$	$a^{(1)}=\frac{S_i^{(1)}}{S_{\max}}$
X1	6,628967	–2,60537	–2,0239	–1,9997	–1,33E-15	–1
X2	–2,60537	1,075372	0,869928	0,660068	–4,44E-16	–0,3333
X3	–2,0239	0,869928	0,72587	0,428106	–4,44E-16	–0,3333
X4	–1,9997	0,660068	0,428106	0,911522	–4,44E-16	–0,3333

Перший цикл ітерації: зведення в квадрат кореляційної матриці.

Ознака	$R_h^2=R_h{}'R_h$				$S_i^{(2)}=\sum_j r_{ij}$	$P_i^{(2)}=R_hS_i^{(1)}$	$a^{(2)}=\frac{S_i^{(2)}}{S_{\max}}$	$d_i=\left\lvert a_i^{(2)}-a_i^{(1)}\right\rvert$
	X1	X2	X3	X4				
X1	58,82	–23,15	–18,008	–17,66	–1,42E-14	–5,88771E-15	–1,00	0
X2	–23,15	9,13	7,12	6,89	1,776E-15	2,31403E-15	0,393	0,726
X3	–18,008	7,12	5,56	5,32	2,665E-15	1,79759E-15	0,305	0,639
X4	–17,66	6,89	5,32	5,44	–2,66E-15	1,77609E-15	0,302	0,635

Оцінки S і P підтверджують правильність проведених обчислень. Таким чином, оцінки компонента першого власного вектора можна вважати достовірними.

Наступні чотири цикли ітерації не приводимо. Перейдемо до визначення навантажень першого головного фактора.

Ознака	$a_i^{(5)}=U_1$	$\beta_1=R_h\cdot a_i^{(5)}$	$A=\frac{U_1\sqrt{\lambda_1}}{\left(\sum_i U_{1i}^2\right)^{1/2}}$
X1	–1	–8,874	–2,575
X2	0,394	3,493	1,013
X3	0,306	2,717	0,788
X4	0,300	2,664	0,773
Власне число: $\max\beta_1=\lambda_1=8,87448$			вектор факторних навантажень

Підсумками першої ітерації будуть: перше власне число $\lambda_1=8,87448$; вектор факторних навантажень: $A=\begin{pmatrix}-2,575 & 1,013 & 0,788 & 0,773\end{pmatrix}$.

Залишається визначити матрицю парних кореляцій λ_1 та вирішити питання з необхідності виконання другої ітерації з пошуком другого власного числа та вектора факторних навантажень A_2:

$$R_h^+=A\cdot A' \tag{8}$$

$$R_h^+=\begin{pmatrix}-2,5746\\ 1,01342\\ 0,78827\\ 0,77293\end{pmatrix}\begin{pmatrix}-2,5746 & 1,01342 & 0,78827 & 0,77293\end{pmatrix}=\begin{pmatrix}6,629 & -2,609 & -2,029 & -1,990\\ -2,609 & 1,027 & 0,799 & 0,783\\ -2,029 & 0,799 & 0,621 & 0,609\\ -1,990 & 0,783 & 0,609 & 0,597\end{pmatrix}.$$

Різниця матриць показує залишкову кореляцію, що не пояснюється першим головним фактором, і допомогає відповісти на питання про доцільність виділення другого головного фактора [3, 4, 6]:

$$R_1 = R_h - R_h^+ = \begin{pmatrix} 0,000 & 0,004 & 0,006 & -0,010 \\ 0,004 & 0,048 & 0,071 & -0,123 \\ 0,006 & 0,071 & 0,105 & -0,181 \\ -0,010 & -0,123 & -0,181 & 0,314 \end{pmatrix}.$$

Матриця перших залишкових коефіцієнтів кореляції містить ще досить великі величини й цілком допускає оцінку другого головного фактора. Подальше виконання другої ітерації аналогічне першій, тільки розрахунки виробляються на даних матриці залишків R1.

Перейдемо до визначення навантажень першого головного фактора без приведення п’яти циклів ітерації.

Підсумками другої ітерації будуть: перше власне число $\lambda_2 = 0,46725;$ вектор факторних навантажень $A = \begin{pmatrix} -0,017 & -0,220 & -0,323 & 0,560 \end{pmatrix}$. За результатами підсумкової таблиці побудовано двовимірну конфігурацію шкал, що представлено на рисунку 1.

Ознака	$a_i^{(5)} = U_2$	$\beta_2 = R_h \cdot a_i^{(5)}$	$A = \dfrac{U_2\sqrt{\lambda_2}}{\left(\sum_i U_{2i}^2\right)^{1/2}}$
X1	–0,031	–0,014	–0,017
X2	–0,392	–0,183	–0,220
X3	–0,577	–0,270	–0,323
X4	1	0,467	0,560
Власне число: $\max\beta_2 = \lambda_2 = 0,46725$			вектор факторних навантажень

Таблиця 2 – Підсумкова таблиця

	Головний чинник (факторні навантаження)	
	технічний стан	комфортність
ПАЗ, 2012 рік випуску	–2,575	–0,017
Богдан, 1999 рік випуску	1,013	–0,220
ПАЗ, 2010 рік випуску	0,788	–0,323
Богдан, 2010 рік випуску	0,773	0,560

На рисунку 1 за результатами підсумкової таблиці побудовано двовимірну конфігурацію шкал.

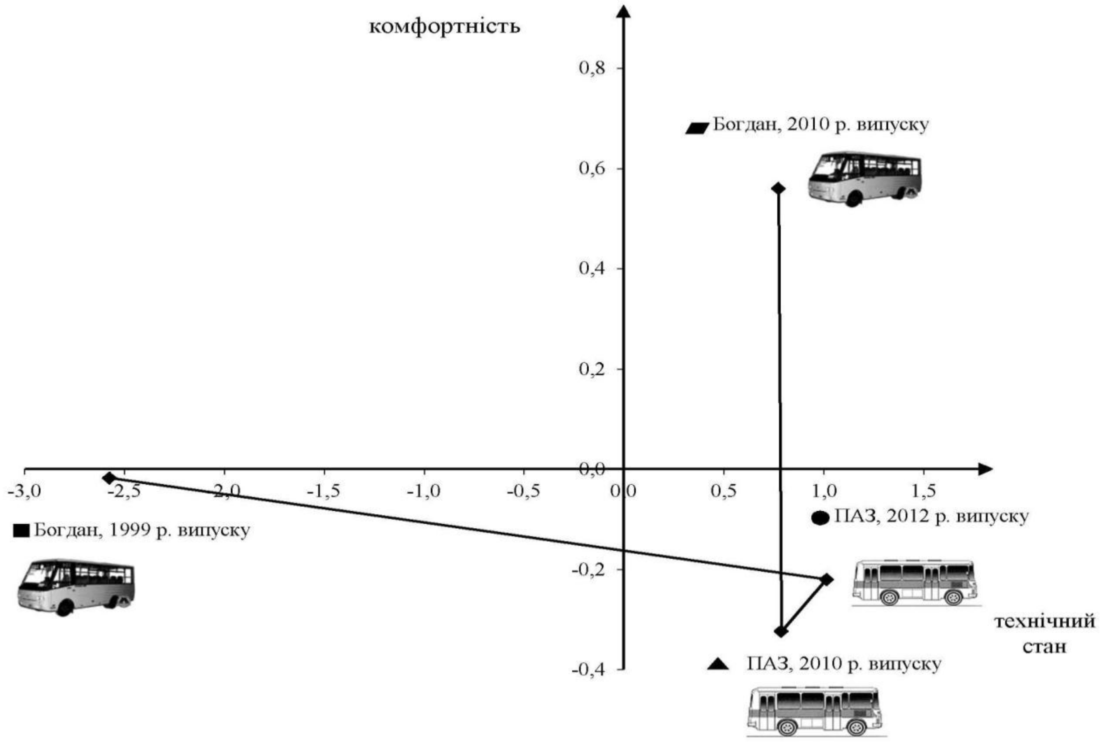

Рисунок 1 – Двовимірна конфігурація шкал

▰ – Богдан, 2010 р. випуску; ● – ПАЗ, 2012 р. випуску; ▲ – ПАЗ, 2010 р. випуску;

■ – Богдан, 1999 р. випуску

Висновки

У даній роботі було зроблено спробу застосувати один із методів багатовимірного статистичного аналізу щодо підвищення якості транспортного обслуговування пасажирів. Виділено головні (латентні) фактори, побудовано факторний простір зі спостережуваними об'єктами.

Список літератури

1. Большаков А. М. Повышение качества обслуживания пассажиров и эффективности работы автобусов / А. М. Большаков, Е. А. Кравченко, С. Л. Черникова. – М.: Транспорт, 1981. – 258 с.

Bolshakov A. M. Povysheniye kachestva obsluzhivaniya passazhirov i effektivnosti raboty avtobusov (Quality Improvement in Passenger Service and Efficiency Improvement in Bus Performance) / A. M. Bolshakov, Ye. A. Kravchenko, S. L. Chernikova. – M.: Transport, 1981. – 258 s.

2. Дудниев Д. И. Организация перевозок пассажиров автомобильным транспортом / Д. И. Дудниев, М. И. Климова, А. А. Менн. – М.: Транспорт, 1974. – 295 с.

Dudniyev D. I. Organizatsiya perevozok passazhirov avtomobilnym transportom (Management of Passenger Transportation by Automobile Transport) / D. I. Dudniyev, M. I. Klimova, A. A. Menn. – M.: Transport, 1974. – 295 s.

3. Афифи А. Статистический анализ: подход с использованием ЭВМ / А. Афифи, С. Эйзен; пер. с анг. – М.: Мир, 1982. – 488 с.

Afifi A. Statisticheskiy analiz: podhod s ispolzovaniyem EVM (Statistical Analysis: with Computer-Based Approach) / A. Afifi, S. Eyzen; per. s ang. – M.: Mir, 1982. – 488 s.

4. Граберг А. Г. Финансы и статистика. Статистическое моделирование и прогнозирование: учебное пособие / А. Г. Граберг. – 1990. – 383 с.

Graberg A. G. Finansy i statistika. Statisticheskoye modelirovaniye i prognozirovaniye: uchebnoye posobiye (Finances and Statistics. Statistical Modeling and Prognosis: Study Guide) / A. G. Graberg. – 1990. – 383 s.

5. Ловецкий К. П. Математическое моделирование. Часть 1 / К. П. Ловецкий, Л. А. Севастьянов. – М.: РУДН, 2007. – 64 с.

Lovetskiy K. P. Matematicheskoye modelirovaniе. Chast 1 (Mathematical Modeling. Part 1) / K. P. Lovetskiy, L. A. Sevastyanov. – M.: RUDN, 2007. – 64 s.

6. Вальд А. Последовательный анализ / А. Вальд. – М.: Физматлит, 1960.

Vald A. Posledovatelnyy analiz (Sequential Analysis) / A. Vald. – M.: Fizmatlit, 1960.

7. Терехина А. Ю. Метрическое многомерное шкалирование / А. Ю. Терехина. – М.: НПУ, 1977.

Terekhina A. Yu. Metricheskoye mnogomernoye shkalirovaniye (Metric Multidimensional Scaling) / A. Yu. Terekhina. – M.: NPU, 1977.

8. Сошникова Л. А. Многомерный статистический анализ в экономике: учеб. пособие для вузов / Л. А. Сошникова, В. Н. Тамашевич, Г. Уебе, М. Шефер; под ред. проф. В. Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 541 с.

Soshnikova L. A. Mnogomernyy statisticheskiy analiz v ekonomike: ucheb. posobiye dlya vuzov (Multidimensional Statistical Analysis in Economics: Study Guide) / L. A. Soshnikova, V. N. Tamashevich, G. Uebe, M. Shefer; pod red. prof. V. N. Tamashevicha. – M.: YUNITI-DANA, 1999. – 541 s.

Рецензент: канд. техн. наук, доц. О. В. Толок, АДІ ДонНТУ.
Стаття надійшла до редакції: 01.07.2013