

УДК 656.072

П.Ф. Горбачов, О.В. Макарічев, Є.В. Любий*Харківський національний автомобільно-дорожній університет, Україна***ІНТЕРВАЛЬНЕ ОЦІНЮВАННЯ ТОЧНОСТІ РОЗРАХУНКУ КОЕФІЦІЄНТУ СПІВВІДНОШЕННЯ ПЛАТНИХ І ПІЛЬГОВИХ ПАСАЖИРІВ**

В рамках даного дослідження авторами запропоновано методика проведення обстеження пасажиропотоків для виявлення кількості платних і пільгових пасажирів, відмінність якої полягає у використанні спеціальної процедури формування вибірки. Запропонований авторами підхід щодо підтвердження точності визначення коефіцієнтів співвідношення платних і пільгових пасажирів дозволяє встановити довірчий інтервал потрапляння розрахункового значення коефіцієнту з ймовірністю 0,95.

Ключові слова: громадський транспорт, коефіцієнт співвідношення, інтервальна оцінка точності.

Постановка проблеми

Будь-яке обстеження співвідношення кількості безплатних і платних пасажирів обов'язково буде вибіркоvim. Це обумовлене безперервним характером роботи міського пасажирського транспорту та відсутністю автоматизованих засобів обліку пасажиропотоків. Таке становище викликає труднощі у визначенні обсягу генеральної сукупності, так як вона постійно зростає, це необхідно враховувати при виборі методу проведення розрахунків [1].

Аналіз останніх досліджень і публікацій

Генеральною сукупністю зветься безліч усіх одиниць статистичної сукупності. Багато завдань статистичного аналізу пов'язані з описом великих сукупностей об'єктів. Найчастіше на практиці з тих або інших причин неможливо розглянути всі елементи таких сукупностей. У цьому випадку обмежуються вивченням лише деякої частини генеральної сукупності. Ця частина називається вибірковою сукупністю або вибіркою. Отримані при вивченні вибірки результати прагнуть поширити на всю генеральну сукупність. Для цього вибірка повинна бути не будь-якою довільною частиною генеральної сукупності, а такою її частиною, що досить правильно відображає основні параметри цієї сукупності. Таким чином, вибірка повинна бути репрезентативною [2, 3].

Домогтися репрезентативності вибірки, тобто того, щоб вона правильно відображала основні властивості, властиві генеральній сукупності можливо за рахунок випадкового відбору одиниць у вибірку сукупність. Випадковість ніяким чином не отожднюється зі стихійністю або довільністю відбору – навпаки, випадковість означає те, що всі

об'єкти генеральної сукупності повинні мати рівні шанси потрапити до вибірки. Найбільш простим є випадковий відбір, наприклад, за допомогою звичайного жеребкування. Для випадкового відбору часто використовуються так звані таблиці випадкових чисел. Для більш великих, але досить однорідних сукупностей використовується механічний відбір [2].

Для неоднорідних сукупностей з певною структурою частіше застосовується типовий відбір. Існують й інші методи, у тому числі – комбінації різних способів відбору на декількох етапах побудови вибіркової сукупності. Однак ніяка, навіть самим ретельним чином сформована вибірка, не може дати точного знання про генеральну сукупність. Таким чином, у вибіркових результатах завжди присутні помилки. Ці помилки можна розділити на два класи: випадкові та систематичні.

До випадкових помилок відносяться відхилення вибіркових характеристик від генеральних (наприклад, відхилення середнього значення ознаки у вибірці від середнього значення цієї ознаки в генеральній сукупності). Випадкові помилки обумовлені самою природою вибіркового методу, і тому вони неминучі. Однак величина випадкової помилки піддається обчисленню (оцінці).

Систематичні помилки, навпаки, не носять випадкового характеру, вони пов'язані з відхиленням структури вибірки від реальної структури генеральної сукупності. Систематичні помилки з'являються тоді, коли порушується основне правило випадкового відбору – забезпечення для всіх об'єктів рівних шансів потрапити у вибірку. Необхідно будувати вибірку так, щоб усунути систематичні помилки.

Виникає питання: як сильно відхиляється ця оцінка від істинного значення? Зокрема, чи не

можна вказати таку величину помилки, що «практично вірогідно» (тобто з ймовірністю, близької до одиниці) гарантувала б, що вибіркова оцінка не відрізняється від невідомого значення більш, ніж на величину цієї помилки? Тобто, чи не можна вказати навколо вибіркового значення такий інтервал, який би із заданою (досить високою) ймовірністю, – довірчою ймовірністю – «покривав» би істинне значення цього параметра? Цей інтервал у математичній статистиці називається довірчим інтервалом, його величина залежить як від довірчої ймовірності (тобто надійності оцінювання), так і від обсягу вибірки [2].

Виявляється, що всі ці вибіркові середні при достатньо великих вибірках розподілені нормально навколо генеральної середньої при достатньо великій кількості повторень вибірки одного й того числа об'єктів з генеральної сукупності. При цьому неминучий розкид вибірових середніх навколо генеральної середньої (тобто стандартне відхилення вибірових середніх) називається стандартною помилкою вибірки.

На практиці навколо вибіркового значення будується такий інтервал, який би із заданою (досить високою) ймовірністю – довірчою ймовірністю – «покривав» би істинне значення цього параметра в генеральній сукупності. Цей інтервал у математичній статистиці називається довірчим інтервалом. Довірча ймовірність P – це ступінь упевненості в тому, що довірчий інтервал дійсно буде містити істинне (невідоме) значення параметра в генеральній сукупності. Довірча ймовірність виражається відсотком вибірок даного обсягу, які дають довірчі інтервали, що містять значення в генеральній сукупності [2].

Формулювання мети статті

Метою даного дослідження є розробка підходу щодо оцінювання точності визначення коефіцієнтів співвідношення платних і пільгових пасажирів, на основі якого існує можливість встановлення довірчого інтервалу потрапляння розрахункового значення коефіцієнту з ймовірністю 0,95.

Виклад основного матеріалу

В загальному випадку коефіцієнт співвідношення кількості платних і пільгових пасажирів визначається за формулою

$$K = \frac{Q_{\text{пільг}}}{Q_{\text{пл}}}, \quad (1)$$

де K – фактичний коефіцієнт співвідношення кількості платних і пільгових пасажирів;

$Q_{\text{пільг}}$ – кількість пасажирів, які мають право та користувалися безоплатним проїздом в ГТ, пас.;

$Q_{\text{пл}}$ – кількість перевезених пасажирів, які відносяться до платної категорії, пас.

Методика збору інформації для розрахунку коефіцієнту співвідношення кількості платних і пільгових пасажирів базується на багаторічному досвіді кафедри транспортних систем і логістики Харківського національного автомобільно-дорожнього університету з проведення різноманітних обстежень пасажиропотоків. Цей досвід свідчить, що при випадковому та масовому відборі обліковців розраховувати на об'єктивність результатів не доводиться, особливо при наявності фінансових інтересів у інших суб'єктів процесу – водіїв та кондукторів. Результати, які можуть вважатися об'єктивними, можливо отримати лише за умови проведення обстеження силами незначної (не більше 20 осіб) групи професійно підготовлених обліковців з мінімальним контактом між обліковцями та водіями або кондукторами [1].

Це диктує необхідність переходу від супроводження обліковцями транспортного засобу протягом всього рейсу до випадкових вимірів кількості пасажирів, що знаходяться в його салоні. При цьому немає необхідності розділяти платних пасажирів на категорії за формами оплати проїзду – одноразовими талонами, чи абонементним квиткам, так як у залежності (1) вони не розділяються між собою. Це дає змогу більше уваги приділити основному призначенню обстеження – визначенню кількості пільгових і платних пасажирів.

Забезпечити максимальну відповідність параметрів вибіркової сукупності генеральній сукупності можливо за рахунок власно випадкового методу відбору одиниць у вибірку сукупність. Але в умовах обмеженого фінансування робіт і терміну проведення обстеження пасажиропотоків на маршрутах міста з метою розрахунку коефіцієнту співвідношення пільгових і платних пасажирів, формування вибірки власно випадковим методом стає неможливим. В цьому випадку потрібно обирати декілька параметрів роботи маршрутів: рейс за маршрутним розкладом руху, напрям руху та ділянку, на яких збирається інформація. Це не є ефективним через те, що для переїзду на нову ділянку обстеження обліковці витрачають багато часу, а обстежувати пасажиропотоки під час переїзду неможливо, оскільки це означає порушення умов власно випадкового відбору.

Внаслідок вище наведеного, для реалізації принципів вибіркового методу пропонується методика проведення обстеження, яка передбачає спеціальну процедуру формування вибірки. Виходячи з обмежених ресурсів на проведення обстеження в якості кількості бригад обліковців приймається дві одиниці. Ці бригади працюють кожного дня обстеження по десять робочих годин за

змішаним графіком таким чином, щоб на лінії завжди знаходилась хоча би одна група обліковців. Метою роботи обліковців є створення псевдовипадкової вибірки максимально можливого обсягу. Для досягнення більшої випадковості процесу обліковці не знайомі з розкладом руху на маршрутах, які обстежують. Вибір напрямку руху не регламентується, обліковцям ставиться завдання максимального охоплення території міста.

Обстеження проводиться протягом одного тижня. З цієї вибірки, за допомогою генератора випадкових чисел, обирається 20 карток, дані з яких є основою для розрахунку потрібного обсягу вибіркової сукупності за залежністю [4, 5]

$$N_B = \frac{t^2 \cdot w \cdot (1-w)}{\Delta^2}, \quad (2)$$

де N_B – потрібний обсяг вибірки, од.;

t – параметр розподілу Стьюдента (для 95 % довірчої імовірності $t=1,96$);

w – питома вага однієї з категорій пасажирів, платних або пільгових;

Δ – максимально припустима помилка вибірки, $\Delta = 0,05$.

Кінцева вибірка формується на основі випадкового відбору карток з псевдовипадкової сукупності. Для цього розраховується частка карток, які необхідно виключити з псевдовипадкової сукупності ν

$$\nu = 1 - \frac{N_B}{N_{пв}}, \quad (3)$$

де $N_{пв}$ – обсяг псевдовипадкової сукупності, од.

Відбір до кінцевої сукупності здійснюється за допомогою генератора випадкових чисел при виконанні умови

$$\varepsilon > \nu, \quad (4)$$

де ε – випадкова величина, розподілена за рівномірним законом в межах $0;1$.

Кінцевим результатом розрахунків випадкової величини на основі вибіркового обстеження повинна бути оцінка точності її визначення, тобто ширини інтервалу навколо отриманого значення, в який із заданою ймовірністю потрапляє шукане значення генеральної сукупності. В даному випадку результатом обстеження є відношення двох випадкових величин, тобто кількості пільгових та платних пасажирів. Для отримання інтервальних оцінок відносних експериментальних даних про настання двох протилежних подій у серії з n експериментів необхідно використовувати

коефіцієнт, який в даному випадку визначається як відношення кількості пільгових пасажирів n_l до кількості пасажирів, що сплатили послуги міського пасажирського транспорту n_l

$$K_l = \frac{n_l}{n_l}, \quad (5)$$

де $n_l = n - n_l$.

Якщо коефіцієнт K_l представити через частотні характеристики, то (5) прийме наступний вигляд:

$$K_l = \frac{n_l}{n - n_l} = \frac{\nu_l}{1 - \nu_l}, \quad (6)$$

де ν_l – частота настання події l , $\nu_l = \frac{n_l}{n}$

Звідси можна визначити частоту ν_l

$$\begin{aligned} K_l = \frac{\nu_l}{1 - \nu_l} &\Leftrightarrow K_l \cdot (1 - \nu_l) = \nu_l \Leftrightarrow \\ &\Leftrightarrow K_l = (K_l + 1) \cdot \nu_l \Leftrightarrow \nu_l = \frac{K_l}{K_l + 1} \end{aligned} \quad (7)$$

Слід відзначити, що ймовірність того, що випадково обраний пасажир буде пільговиком можна визначити наступним чином:

$$p = \frac{L}{N}, \quad (8)$$

де N – загальна кількість пасажирів у всій генеральній сукупності (річний обсяг перевезень пасажирів у місті, оскільки K_l визначається на рік), осіб;

L – кількість пільговиків серед N , осіб.

Доцільно вважати, що обсяг N усіх пасажирів в генеральній сукупності є настільки великим, що події пов'язані з ідентифікацією наступного пасажирів асимптотично незалежні та однаково розподілені як індикатори пільговиків

$$I(L) = \begin{cases} 1, & \text{з ймовірністю } p \\ 0, & \text{з ймовірністю } 1 - p \end{cases} \quad (9)$$

Звідси частота появи пільговиків складатиме

$$\nu_n(L) = \frac{I_1 + I_2 + \dots + I_n}{n} \quad (10)$$

Відповідно до центральної граничної теореми теорії ймовірностей [6], при $n \rightarrow \infty$ частота появи пільговиків асимптотично нормальна з

математичним очікуванням, що дорівнює p та асимптотичною дисперсією рівній $\frac{p \cdot (1-p)}{n}$.

Практично це означає що [7]

$$P\left\{|v_n - p| < t_\gamma \cdot \sqrt{\frac{p \cdot (1-p)}{n}}\right\} = \gamma, \quad (11)$$

$$\text{де } \gamma = 2\Phi(t_\gamma) = 2 \int_0^{t_\gamma} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt - \text{подвоєне значення}$$

функції Лапласа в точці t_γ , яке позначає рівень довіри інтервалу нормально розподіленої випадкової величини з параметрами (0;1).

Для визначення інтервальної оцінки ймовірності p того, що випадково обраний пасажир буде пільговиком вирішується наступна нерівність

$$|v_n - p| < t_\gamma \cdot \sqrt{\frac{p \cdot (1-p)}{n}}. \quad (12)$$

Після зведення правої та лівої частини (12) в квадрат отримаємо

$$\begin{aligned} v_n^2 - 2 \cdot v_n \cdot p + p^2 &< t_\gamma^2 \cdot \frac{p \cdot (1-p)}{n} \Leftrightarrow \\ \Leftrightarrow p^2 \cdot \left(1 + \frac{t_\gamma^2}{n}\right) - 2 \cdot p \cdot \left(v_n + \frac{t_\gamma^2}{2 \cdot n}\right) + v_n^2 &< 0 \end{aligned} \quad (13)$$

Маючи (13) можна знайти менший та більший корінь відповідного квадратного рівняння

$$p_1 = \frac{v_n + \frac{t_\gamma^2}{2 \cdot n} - t_\gamma \cdot \sqrt{\frac{(1-v_n) \cdot v_n}{n} + \left(\frac{t_\gamma}{2 \cdot n}\right)^2}}{\left(1 + \frac{t_\gamma^2}{n}\right)}, \quad (14)$$

$$p_2 = \frac{v_n + \frac{t_\gamma^2}{2 \cdot n} + t_\gamma \cdot \sqrt{\frac{(1-v_n) \cdot v_n}{n} + \left(\frac{t_\gamma}{2 \cdot n}\right)^2}}{\left(1 + \frac{t_\gamma^2}{n}\right)}, \quad (15)$$

де p_1, p_2 – відповідно менший та більший корінь квадратного рівняння.

При цьому дискримінант даного квадратного рівняння буде дорівнювати

$$\begin{aligned} D &= \left(v_n + \frac{t_\gamma^2}{2 \cdot n}\right)^2 - \left(1 + \frac{t_\gamma^2}{n}\right) \cdot v_n^2 = \\ v_n^2 + \frac{v_n \cdot t_\gamma^2}{n} + \frac{t_\gamma^4}{(2 \cdot n)^2} - v_n^2 - \frac{v_n^2 \cdot t_\gamma^2}{n} &= . \quad (16) \\ &= t_\gamma^2 \cdot \left[\frac{v_n \cdot (1-v_n)}{n} + \left(\frac{t_\gamma}{2 \cdot n}\right)^2\right] > 0 \end{aligned}$$

Отриманий інтервал (p_1, p_2) залежить від обсягу вибірки n (кількість пасажирів, що потрапили під обстеження), частоти появи пільговиків v_n , рівня довіри γ (ймовірності того, що інтервал (p_1, p_2) покриє невідому ймовірність p) та $t_\gamma = \Phi^{-1}\left(\frac{\gamma}{2}\right)$ – квантіль відповідної $\frac{1+\gamma}{2}$ ймовірності нормального закону розподілу (визначається на основі таблиць математичної статистики, наприклад, при $\gamma = 0,95$ величина $t_\gamma = 1,96$ [6-9]).

Знаючи вираз для визначення коефіцієнту $K_1 = K = \frac{P}{1-p}$ можна знайти його інтервальну оцінку (K_1, K_2) для невідомого відношення кількості всіх пільговиків до кількості пасажирів, що не є такими

$$K_1 = \frac{p_1}{1-p_1}, \quad (17)$$

$$K_2 = \frac{p_2}{1-p_2}. \quad (18)$$

Звідси можна цілком стверджувати, що ймовірність того, що $K \in (K_1, K_2)$ практично дорівнює γ .

Висновки

Запропонований підхід для підтвердження точності визначення коефіцієнтів співвідношення платних і пільгових пасажирів дозволяє провести інтервальне оцінювання, на основі якого існує можливість встановлення довірчого інтервалу (ширини інтервалу) потрапляння розрахункового значення коефіцієнту з ймовірністю 0,95.

Література

1. Проведення аналізу та оптимізації транспортних систем міста Краматорськ [Текст]: Звіт про НДР (кінцевий) / Відділ транспорту, зв'язку та енергетики Краматорської міської ради, Харківський національний автомобільно-дорожній університет; № держ. реєстрації 0116U006995. – Харків, 2016. – 221 с.
2. Дружинин, Н.К. Выборочный метод и его применение в социально-экономических исследованиях [Текст] / Н.К. Дружинин. - М.: Статистика, 1970. – 104 с.

3. Венецкий, И.Г. Основные математико-статистические понятия и формулы в экономическом анализе [Текст] / И.Г. Венецкий, В.И. Венецкая. - М.: Статистика, 1975. - 264 с.
4. Минько, А.А. Статистический анализ в MS Excel [Текст] / А.А. Минько. - М.: Издательский дом «Вильямс», 2004. - 448 с.
5. Справочник по теории вероятностей и математической статистике [Текст] / под ред. В.С. Королюка. - К.: Наукова думка, 1978. - 580 с.
6. Гнеденко, Б.В. Курс теории вероятностей [Текст] / Б.В. Гнеденко. - Изд. 6-е, перераб. и доп. - М.: Наука. Гл. ред. физ.-мат. лит., 1988. - 447 с.
7. Ширяев, А.Н. Вероятность [Текст] / А.Н. Ширяев. - М.: Наука, 1980. - 574 с.
8. Вентцель, Е.С. Теория вероятностей и ее инженерные приложения [Текст] / Е.С. Вентцель, Л.А. Овчаров. - М.: Наука, 1988. - 480 с.
9. Севастьянов, Б.А. Курс теории вероятностей и математической статистики [Текст] / Б.А. Севастьянов. - М.: Наука, 1982. - 256 с.

References

1. Analysis and optimization of the transport systems of the Kramatorsk City (2016): Report on research work (final). Department of Transport, Communications and Energy of Kramatorsk City Council, Kharkiv National Automobile and Highway University; State number registration 0116U006995. Kharkiv, 221.
2. Druzhinin, N.K. (1970). Selective method and its application in socio-economic studies, 104.
3. Venetsky, I.G., Venetskaja, V.I. (1975). Basic mathematical-statistical concepts and formulas in economic analysis, 264.
4. Minko, A.A. (2004). Statistical analysis in MS Excel, 448.

5. A Handbook on Probability Theory and Mathematical Statistics (1978) / ed. V.S., Coroluk, 580.
6. Gnedenko, B.V. (1988). Course of probability theory. Ed. 6-th, revised. and additional, 447.
7. Shiryaev, A.N. (1980). Probability, 574.
8. Venttsel, E.S., Ovcharov, L.A. (1988). Theory of Probability and its Engineering Applications, 480.
9. Sevastyanov, B.A. (1982). Course of the theory of probability and mathematical statistics, 256.

Рецензент: д-р техн. наук, проф. Ю.О. Давідіч, Харківський національний університет міського господарства імені О.М. Бекетова, Україна.

Автор: ГОРБАЧОВ Петро Федорович
доктор технічних наук, професор, завідувач
кафедри транспортних систем і логістики
Харківський національний автомобільно-дорожній
університет
E-mail – gorbachov.pf@gmail.com

Автор: МАКАРІЧЕВ Олександр Володимирович
д.ф.-м.н., доцент.
Харківський національний автомобільно-дорожній
університет
E-mail – amsol2904@gmail.com

Автор: ЛЮБИЙ Євген Володимирович
кандидат технічних наук, доцент
Харківський національний автомобільно-дорожній
університет
E-mail – lion_khadi@ukr.net

INTERVAL ESTIMATION OF THE CALCULATION ACCURACY FOR THE RATIO COEFFICIENT OF THE PAID AND REDUCED FARE PASSENGERS QUANTITY

P.F. Gorbachov, O.V. Makarichev, Y.V. Liubyi

Kharkiv National Automobile Highway University, Kharkiv

The majority of urban passenger transportation enterprises are unprofitable due to the fact that incomes resulted from the passenger transportation do not cover the costs to their implementation. However, the reasons for the loss of urban passenger transportation are different: on the one hand, the local authorities to restrain the growth of tariffs, and on the other hand, there is an increase in the number of categories of passengers exercising the right of preferential travel. The determination of the share of the preferential contingent that uses the services of urban electric transportation is a particularly acute problem due to the slowdown in development and the disappearance of this mode of transportation. There are two main ways to solve the problem. The first is to reduce the number of categories of citizens exercising the right of preferential travel. The second is to calculate the carriers' subsidies. Mostly, for urban electric transportation, the obtained "preferential-to paid contingent" ratio is carried out by an elementary determination of the arithmetic mean value from the measurement results, which can lead to the development of erroneous efforts. To fully assess the unknown parameter, which in this case is "preferential-to paid contingent" ratio, a point estimation is not enough. It is necessary to perform an interval estimation, which may be a confidence interval. In the research, the authors have proposed a technique for conducting passenger flow surveys to identify the quantity of the paid and the preferential passengers. Its difference from the existing ones is to use a special sampling procedure formed by the minimum number of properly trained accountants, who are unaware of the route schedules for the greater process randomness. The approach proposed by the authors to confirm the accuracy of determining "preferential-to paid contingent" ratio allows revealing the confidence interval of the calculated coefficient value with a probability of 0,95. The approach proposed to assess the accuracy of calculating the ratio of paid and preferential passengers has been checked during a passenger flow survey on the route network of the town of Kramatorsk, Donetsk region.

Keywords: public transport, ratio coefficient, accuracy interval estimation.