

Використання статистичних методів обробки масивів даних для вимірювання диференціації рівня доходів населення

В статті проведено аналіз існуючої практики застосування статистичних методів обробки масивів даних для вимірювання диференціації рівня доходів населення. Встановлено, що існуюча на сьогодні методика вибіркового обстеження сімей хоча і відповідає міжнародним стандартам, але не відбиває реальної диференціації доходів населення. Запропоновано для вимірювання диференціації рівня доходів використовувати неперсоніфіковані данні про сукупний річний дохід громадян, що подаються до податкової адміністрації.

статистичні методи обробки, масиви даних, доходи населення, диференціація доходів населення, кількісні показники

Оцінка рівня диференціації доходів в розрізі регіонів викликає значні методологічні проблеми, основною з яких є неможливість отримання достовірного значення ВВП за регіонами. Це пов'язано з тим, що методи розрахунку цього показника (виробничий, за кінцевим споживанням) спрямовані на отримання результатів в цілому по країні, розрахунки проводяться за галузями економіки. Тому, єдино можливими для застосування залишаються два наступні методи визначення рівня диференціації доходів населення:

- за допомогою вибірових обстежень, що проводяться працівниками статистики.
- за даними про сукупний дохід, що подаються до органів державної податкової адміністрації.

Існуюча на сьогодні методика вибіркового обстеження сімей хоча і відповідає міжнародним стандартам, але не відбиває реальної картини рівня диференціації доходів населення. Перехід на існуючу методику було здійснено Держкомстатом України у 1998 році. Вона передбачає використання в розрахунках не доходів, а витрат домогосподарств, як джерела більш надійної інформації. Але, слід відмітити, що при багатоступінчастій імовірнісній виборці у вибірову сукупність об'єктів обстеження потрапляють сім'ї, для котрих отримання плати за участь в обстеженнях має сенс. Тобто, високодохідні групи населення відмовляються від проведення обстеження. Отже, домогосподарства, що мають високу ступінь доходу переважно не потрапляють до вибірки, що робить її нерепрезентативною. Крім того, організація обстеження на засадах випадкової вибірки унеможливує репрезентативність обстеження на рівні регіонів. Уникнути такої ситуації можна у разі залучення до обстеження дуже великої кількості сімей, що неможливо із-за великих втрат навіть для країн з розвиненою економікою.

До другого способу визначення рівня диференціації доходів населення можна віднести використання неперсоніфікованих даних про сукупний річний дохід громадян, що подаються до податкової адміністрації. Нині ці данні використовуються для обчислення прибуткового податку з громадян. Аналіз рівня диференціації рівня доходів, який можна здійснити за цими даними, не робиться і не оприлюднюється.

Слід зазначити, що публічний аналіз податкової статистики є рідкісним явищем не тільки на території колишнього СРСР, а й у світовій практиці. Виключення становить податкове законодавство Канади, яке щороку публікує результати своїх аналізів. В Україні математичною обробкою неперсоніфікованих даних податкової статистики протягом 1998-2001 року займався М.Лощинін [2, 4, 5]. В його роботі [4] “Эконометрия физических лиц Украины” наведено розрахунок доходів середнього класу в розрізі областей.

Висока диференціація цього показника (найбільший прошарок середнього класу має Київський регіон – 66%, найменший – Тернопільський – 37%) по території України свідчить про суттєву відмінність рівня диференціації доходів населення в розрізі регіонів. Слід відмітити, що по Кіровоградській області прошарок середнього класу склав 50%, що є менше ніж в середньому по країні (51,36%).

Використання кількісних показників можливо як при вивченні рівня диференціації доходів, так і заробітної плати, як найбільш суттєвої компоненти доходів. У процесі оцінки рівня життя застосовуються наступні кількісні показники:

1. Квантілі (градієнти) характеризують точки розподілу заробітної плати, що ділять його в певному співвідношенні. Прикладом квантілів є медіана (Me), яка поділяє ряд розподілу заробітної плати навпіл. З інших градієнтів найбільш часто використовуються квартилі (чверті) - Q , децілі (десяти частки) - d , півдецілі (двадцяті частки) і перцентілі (соті частки).

Квартільне відхилення характеризує міру розсіювання. Воно дорівнює половині інтервалу між квартилями

$$D = \frac{Q_3 - Q_1}{2}, \quad (1)$$

де Q_3 - верхній квартиль;

Q_1 - нижній квартиль.

Цей показник дає можливість визначити той інтервал, в якому знаходиться половина всіх членів ряду. Якщо K - точка на шкалі, що лежить посередині між першим і третім квартилями, то $K \pm D$ містить половину значень;

2. Квартильний коефіцієнт диференціації. Квартильний коефіцієнт характеризує відношення третього квартиля до першого:

$$K(Q) = \frac{Q_3}{Q_1}; \quad (2)$$

3. Деціальний коефіцієнт диференціації визначається відношенням дев'ятого деціля до першого:

$$K(d) = \frac{d_9}{d_1}. \quad (3)$$

Простота, наочність, а також відносний характер коефіцієнтів сприяли їх широкому використанню. Але вони не завжди дозволяють однозначно відповісти на питання про ступінь відхилень в диференціації двох рядів розподілу. Можливі також випадки, коли квартильний коефіцієнт показує меншу диференціацію одного ряду порівняно з іншим, в той час як деціальний, навпаки. В таких випадках не можна зробити правильний висновок відносно різниці в диференціації рядів. Тому, природно, виникає бажання знайти будь-який показник нерівномірності розподілу. Цим показником може бути середня абсолютна похибка по Джині;

4. Похибка по Джині (або коефіцієнт Джині) розраховується за формулою:

$$G^{\partial} = \frac{\sum_{i,j} |x_i - x_j|}{N^2}, \quad (4)$$

де i, j – відповідні елементи двох рядів розподілу;

N – кількість цих елементів.

Оскільки визначення цього показника трудомістке, то застосовується коефіцієнт концентрації Лоренца – K_l ;

5. Коефіцієнт концентрації Лоренца розраховується за формулою:

$$K_l = 1 - 2 \int_0^1 \eta_l^{(1)}(t) f(t) dt, \quad (5)$$

де $\eta_l^{(1)}(t)$ - перший момент розподілу, який називається кривою Лоренца.

Крива Лоренца характеризує кумулятивний розподіл, в якому кумують не тільки кількість робітників, а й їх доход. Вона характеризує співвідношення між доходами та працівниками у відсотках. Якби всі доходи розподілялись рівномірно, тобто 10% працівників одержували 10% доходу, а 50% - половину, то такий розподіл мав би вигляд прямої з кутом нахилу в 45° . Нерівномірний розподіл характеризується кривою, яка відхиляється від прямої тим більше, чим більше диференціація. Площа, що знаходиться між діагональною прямою і кривою Лоренца характеризує ступень диференціації. Чим більше ця площа, тим більша диференціація. Якщо всю площу квадрата прийняти за 1, то половина її, що відсікається діагональною прямою, становить $\frac{1}{2}$ (рис. 1).

Якщо припустити, що площа, яка огинається кривою Лоренца буде S_2 , тоді площа між прямою і кривою (позначеною S_1) буде визначатися формулою:

$$S_1 = S_{ABC} - S_2 = \frac{1}{2} - S_2 \quad (6)$$

або

$$2S_1 = 1 - 2S_2. \quad (7)$$

Одержана рівність є по суті, коефіцієнтом концентрації K_l :

$$K_l = 1 - 2S_2. \quad (8)$$

Чим більша диференціація розподілу, тим далі знаходиться крива від діагоналі прямої, тим більша величина віднімається від одиниці і тим менше коефіцієнт концентрації. Для рівномірного розподілу він дорівнює одиниці.

Коефіцієнт концентрації або показник ступеня диференціації в рівнянні кривої Лоренца розраховується не по двох точках (як квантільні коефіцієнти розподілу), а по всіх точках ряду розподілу. Саме в цьому і є його велика перевага над іншими показниками диференціації;

6. Показником, розрахованим на основі кумулятивних варіантів і кумулятивних частот, є так званий медіальний коефіцієнт, запропонований в якості вимірника диференціації П.П. Масловим. В ньому, крім медіани (Me), використовується ще медіана величина (Md), що поділяє фонд заробітної плати або доходів навпіл. В розподілі, що характеризується відповідною диференціацією, ці показники не співпадають. Чим більша різниця між ними, тим більша й диференціація. Відношення цієї різниці до медіани є “медіальним коефіцієнтом”, тобто:

$$K_{Md} = \frac{Md - Me}{Me}. \quad (9)$$

Для логарифмічно-нормального розподілу цей коефіцієнт дорівнює: $K_{Md} = \frac{\sigma_x^2}{\bar{x}^2}$,

тобто квадрату коефіцієнта варіації.

Отже, і коефіцієнт концентрації і “медіальний коефіцієнт” становлять функцію дисперсії розподілу, що є показником розсіювання змінної величини, зокрема й заробітної плати;

7. Коефіцієнт концентрації доходів Джині (K_L) характеризує ступінь нерівності в розподілі доходів населення. Він розраховується за формулою:

$$K_L = \sum p_i q_{i+1} - \sum p_{i+1} q_i, \quad (10)$$

де p_i - частка населення, що має доход, не вищий, ніж його максимальний рівень в i -ій групі;

q_i - частка доходів i -ої групи в загальній сумі доходів населення, розраховується кумулятивною, аналогічно p_i , але не для показника чисельності населення, а для показника грошового доходу.

Враховуючи простоту розрахунків і те, що цей показник знайшов широке практичне застосування, у процесі досліджень розраховувався коефіцієнт концентрації доходів Джині (K_L) по Україні для сільського та міського населення за статистичними даними. В ході розрахунків, які проводилися з використанням табличного процесора MS Excel було отримано:

$$K_{L_{c2000}} = 4,721144 - 4,400606 = 0,32,$$

$$K_{L_{c2005}} = 4,53624798 - 4,22225886 = 0,31.$$

Отже, коефіцієнт концентрації доходів Джині для сільського населення в 2000 році складав 0,32, а в 2005 – 0,31. Аналогічний показник для міста відповідно - 0,22 ($K_{L_{m2000}} = 3,046784 - 2,822206$), а в 2005 - 0,25 ($K_{L_{m2005}} = 2,94122693 - 2,69380938$).

Коефіцієнт Джині змінюється в межах від 0 до 1. Чим більше його значення відхиляється від нуля і наближається до одиниці, тим більша ступінь диференціації доходів населення, тобто, більша частка доходів сконцентрована в руках окремих груп населення.

Згідно із здійсненими розрахунками, коефіцієнт концентрації доходів Джині в сільських домогосподарствах ближче до одиниці (0,32 – 2000 р.; 0,31- 2005р.), ніж міських (0,22 – 2000 р.; 0,25 – 2005р.). Тобто, диференціація доходів сільського населення в 2000 році була більша майже в 1,5 рази $\left(\frac{0,32}{0,22}\right)$ ніж міського, а в 2005 – 1,24 рази $\left(\frac{0,31}{0,25}\right)$. Тобто можна зробити висновок, що за період з 2000 по 2005 рр.

диференціація між містом і селом хоча і є суттєвою, але характеризується позитивною динамікою.

Більш наочно криві Лоренца для села і міста за 2000 рік, а також теоретичну криву рівномірного розподілу показує (рис.1).

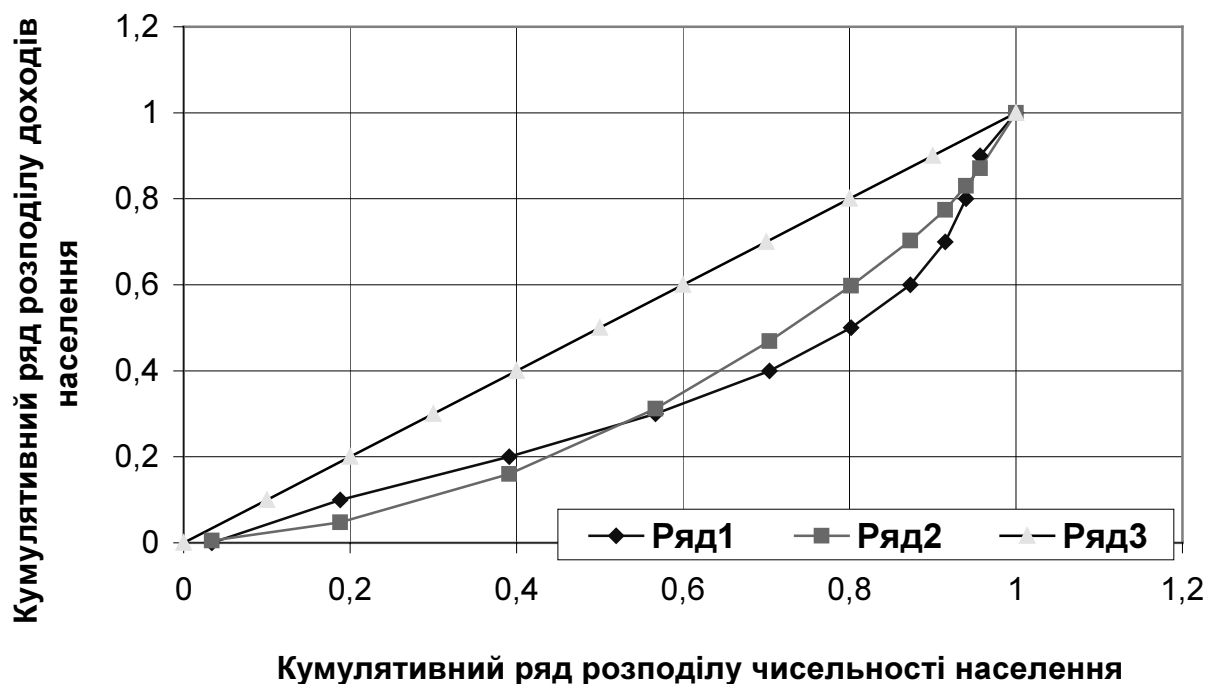


Рисунок - 1 Криві Лоренца для сільського (ряд 1) і міського населення (ряд 2) та теоретичний ряд рівномірного розподілу (ряд 3) 2000 р.

На рис. 1 видно, що лінія розподілу доходів сільського населення (ряд 1) більш прогнута вниз, ніж лінія розподілу доходів міського населення (ряд 2). Це свідчить про більшу ступінь диференціації доходів сільського населення України, ніж міського. Якщо ж врахувати тіньовий аспект, то диференціація стане ще більшою.

Отже, нині на сьогодні багатоступінчаста імовірна вибірка, що використовується органами статистики, не відображає рівень доходів найбагатшого прошарку населення. Визначити рівень їх доходу за рівнем витрат не можливо тому, що: по-перше – участь в оплачуваному обстеженні для них не має сенсу і вони ухиляються від нього; по-друге – підприємці малого та середнього бізнесу на особисте споживання витрачають не більше, ніж наймані робітники з високою заробітною платою, що пов'язано з можливістю частково покривати особисті витрати за рахунок фірм (відрядження, обслуговування автотранспорту, представницькі витрати, меблі, квартири тощо). Крім того, необхідно враховувати доходи дійсно “самих” багатих домогосподарств, рівень і спосіб життя яких відповідає хорошим західним стандартам (наявність нестандартної нерухомості й рахунків у банках за кордоном, подвійне громадянство тощо). Ця група складається з перших осіб фінансових та комерційних структур, “верхівки” мафіозних та корумпованих структур кримінального бізнесу. За оцінками експертів ця група складає від 1 до 3%. Навіть без математичних підрахунків зрозуміло, що реальна диференціація буде ще більша. За підрахунками вчених нині рівень тінізації доходів складає від 40-50%. Враховуючи, що лівову частку тіньових доходів мають саме високодохідні групи населення, реальний рівень диференціації доходів сільського населення буде ще більший. Таким чином, в сучасних умовах для вимірювання реального стану диференціації рівня доходів населення доцільно використовувати неперсоніфіковані дані про сукупний річний дохід громадян, що подаються до податкової адміністрації.

Список літератури

1. Абчук В.А. Экономико-математические методы: Элементарная математика и логика. Методы исследования операций. – СПб.: Союз, 1999. – 320 с.
2. Азаров Н., Бондаренко Г., Лекарь С., Лощинин М. Эконометрия физических лиц Украины // Економіст. —1999. - №12. - С. 37-53.
3. Волкова Г., Мигранова Л., Римашевская Н. Вопросы методики оценки дифференциации доходов населения//Вопросы статистики. 1997. - №2. - С.30-41.
4. М. Лощінін. Економетрія середнього класу // Економіст. —1999. - №12. - С. 62-75.
5. М.Лощінін. Людські молекули в трохи нагрітій банці // Економіст. - 1999. - №7. - С. 76-82.

В статье проведен анализ существующей практики применения статистических методов обработки массивов данных для измерения дифференциации уровня доходов населения. Установлено, что существующая на сегодня методика выборочного обследования семей хотя и отвечает международным стандартам, но не отражает реальной дифференциации доходов населения. Предложено для измерения дифференциации уровня доходов использовать неперсонифицированные данные о совокупном годовом доходе граждан, которые подаются в налоговую администрацию.

In the article the analysis of existent practice of application of statistical methods of treatment of arrays of data is conducted for measuring of differentiation of level of profits of population. It is set that the existing for today method of selective inspection of families answers international standards though, but does not reflect the real differentiation of profits of population. It is suggested for measuring of differentiation of level of profits to utilize неперсонифицированные information about the gross annual income of citizens which are given in tax administration.