

МЕТОДИЧНІ ОСНОВИ КОМПЛЕКСНОГО ОЦІНЮВАННЯ РІВНЯ ІНФОРМАТИЗАЦІЇ
РЕГІОНІВ УКРАЇНИ

Анотація

У статті запропоновано та обгрунтовано переваги методики комплексної оцінки рівня інформатизації регіонів, яка базується на обчисленні відповідного інтегрального індексу, кластеризації регіонів за його складовими, виявленні перспектив у зближенні регіонів за інтенсивністю наростання в них процесів інформатизації.

Annotation

Advantages of method of complex estimation of level of informatization of regions, which is based on the calculation of the proper integral index, clusterization of regions, after his constituents, are offered in the article and grounded, exposure of prospects in rapprochement of regions after intensity of growth in them of processes of informatization.

Постановка проблеми. Інформатизація всіх сфер діяльності сучасної людини і бурхливий розвиток інформаційно-комунікаційних технологій відіграють виняткове значення для соціально-економічного зростання регіонів. На жаль, стан регіональної інформатизації є незадовільним і не відповідає сучасним потребам суспільства та держави, світовим тенденціям. Має місце значна нерівномірність у поширенні комп'ютерної техніки, баз даних, телекомунікаційних мереж, програмних продуктів між регіонами. Такий стан справ вимагає вироблення ефективних механізмів усунення диспропорцій у інформатизації регіонів та застосуванні математичного апарату для їх оцінки.

Аналіз останніх досліджень і публікацій. Значний внесок у висвітлення питань інформатизації та інформаційного суспільства здійснили Т. Алачова, Л. Балабанова, О. Височан, В. Брижко, В. Цимбалюк, Н. Георгіаді, А. Загородній, В. Гавловський, О. Кузьмін, В. Пожуєв, К. Беляков, С. Петренко, Г. Красноступ, В. Шквір, М. Швець, а також зарубіжні дослідники С. Бороненкова, В. Годін, Є. Духонін, Д. Ісаєв, І. Корнеєв та ін. Різні аспекти інформатизації та інформаційного розвитку регіонів були в центрі уваги таких науковців як І. Вахович, С. Пиріг, Н. Завієна та ін.

Водночас недостатньо висвітленими в сучасній економічній науковій літературі залишаються питання щодо впливу інформатизації на економічний розвиток регіону та механізмів його виміру, оцінки рівня інформатизації регіонів.

Мета статті – пошук нових методичних підходів до комплексної оцінки рівня інформатизації регіонів України.

Виклад основного матеріалу. У наш час надзвичайно важливо мати механізм комплексної оцінки рівня інформатизації регіонів України. Адже, наскільки своєчасно і якісно будуть проведені такі дослідження, залежатиме можливість посилити контроль за виконанням регіональних програм інформатизації, ефективність інформатизації регіональної економіки, нівелювання міжрегіональної диференціації регіонів за рівнем інформатизації та в кінцевому результаті інтенсивність соціально-економічного розвитку регіону.

Методика оцінки рівня інформатизації регіонів, на нашу думку, повинна включати чотири основних етапи (рис. 1). Так, перший етап передбачає визначення мети, об'єкту та предмету проведення такого аналізу. Його метою є оцінка рівня інформатизації регіонів, дослідження нерівномірності розвитку цього процесу у регіонах та визначення періоду подолання розриву між регіонами за рівнями розвитку їх інформатизації. Об'єкт аналізу – процеси інформатизації, що обумовлюють функціонування й розвиток регіональних соціально-економічних підсистем, а предмет – визначення рівня інформатизації регіонів України та періоду, за який відбудеться нівелювання міжрегіональної диференціації за цим показником.

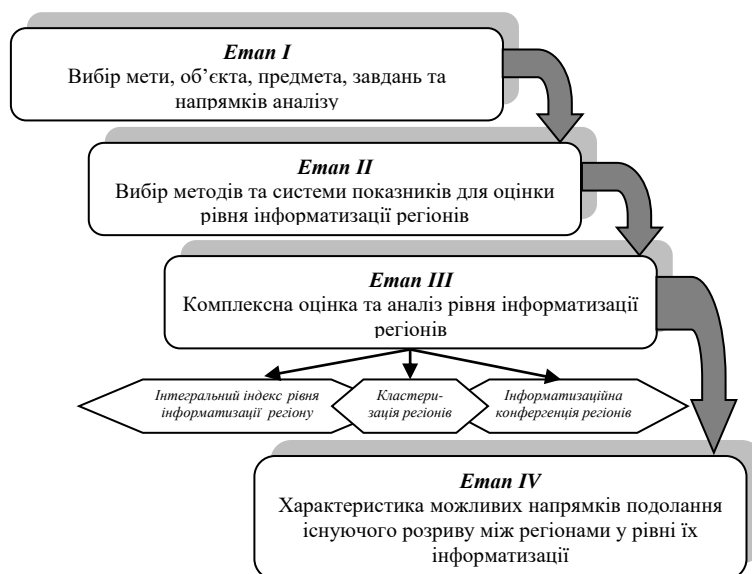


Рис.1. Етапи аналізу та оцінки рівня інформатизації регіонів

Відповідно до окресленої мети дослідження необхідним є вирішення наступних завдань:

1. Проведення оцінки рівня інформатизації регіонів України (інтегральний індекс рівня інформатизації регіону).
2. Групування регіонів у кластери за рівнем інформатизації.

3. Моделювання дивергентно-конвергентних процесів у розвитку інформатизації на рівні регіонів.
4. Визначення періоду зближення регіонів за рівнем інформатизації.

Комплексність дослідження на етапах II і III забезпечує наступна система показників:

x_1 - Кількість персональних комп'ютерів на підприємствах, од.; x_2 - Кількість підприємств, які використовували комп'ютери упродовж року, од.; x_3 - Середня кількість працівників на підприємствах, що використовували комп'ютер, осіб; x_4 - Кількість підприємств, які мали функціонуючу домашню сторінку у внутрішній комп'ютерній мережі (Інтранет), од.; x_5 - Кількість підприємств, які використовували внутрішню комп'ютерну мережу, од.; x_6 - Середня кількість працівників на підприємствах, що регулярно використовували приєднаний до Інтернет комп'ютер; x_7 - Кількість підприємств, які для електронного обміну інформацією використовували веб-сайт, од.; x_8 - Кількість підприємств, які використовували бездротовий доступ для своєї внутрішньої комп'ютерної мережі, од.; x_9 - Кількість підприємств, які використовували Інтернет для отримання послуг освіти, од.; x_{10} - Кількість підприємств, які мали розширену внутрішню комп'ютерну мережу, од.; x_{11} - Кількість підприємств, які мали доступ до Інтернет, од.; x_{12} - Кількість підприємств, які використовували Інтернет для отримання банківських та фінансових послуг, од.; x_{13} - Кількість підприємств, які для електронного обміну інформацією використовували автоматизований обмін даними, од.; x_{14} - Кількість підприємств, які використовували Інтернет для отримання інформації, од.; x_{15} - Кількість підприємств, що мали веб-сайт або домашню сторінку, од.; x_{16} - Кількість підприємств, які здійснювали автоматизований обмін даними, од.; x_{17} - середній обсяг реалізованих послуг підприємствами, що здійснюють діяльність у сфері надання послуг з ремонту і технічного обслуговування офісної та електронно-обчислювальної техніки; x_{18} - середня кількість ЕОМ у парку обчислювальної техніки; x_{19} - середня кількість інтернет-користувачів; x_{20} - середня частка інвестицій в основний капітал, що припадає на операції з нерухомим майном, оренда, інжиніринг та надання послуг підприємцям.

На їх основі формується матриця часткових показників X розміру $m \times n$:

$$X = (x_{ij}), \quad j = \overline{1, n}, \quad i = \overline{1, m} \quad (1)$$

де n - число показників; m - кількість регіонів; x_{ij} - значення j -го показника в i -му регіоні.

У зв'язку із тим, що всі показники мають різні одиниці вимірювання, то необхідно їх стандартизувати. Найчастіше стандартизацію здійснюють шляхом їх нормалізації за середнім квадратичним відхиленням за формулою:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \text{ або } z_{ij} = \frac{x_{ij} - \frac{1}{m} \sum_{i=1}^m x_{ij}}{\left[\frac{1}{m} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2 \right]^{1/2}}, \quad (2)$$

де z_{ij} - нормоване значення вхідного показника; \bar{x}_j - середнє арифметичне вхідних показників j -ої ознаки; σ_j - середнє квадратичне відхилення значень j -ої ознаки.

У результаті таких перетворень отримаємо матрицю розміром $m \times n$ виду:

$$Z = (z_{ij}), \quad j = \overline{1, n}, i = \overline{1, m}. \quad (3)$$

У даній матриці сукупність значень показників по регіонах можна інтерпретувати як деяку точку P_i в n -вимірному векторному просторі, координати якої рівні z_{ij} ($i = \overline{1, m}$).

Для обчислення інтегральних індексів рівня інформатизації регіонів застосуємо метод таксономічного аналізу, який дозволяє впорядкувати елементи сукупності за відстанню до певної визначеної у просторі точки, що є еталоном розвитку [1, с. 8] та забезпечує комплексність дослідження при врахуванні впливу великого числа показників.

Еталон розвитку або еталонна точка $P_0 = (z_{01}, \dots, z_{0n})$ визначається вибором по кожному із показників матриці Z найбільшого значення серед регіонів, якщо ця ознака є стимулятором, і найменшого, коли ознака є дестимулятором.

Значення інтегрального індексу рівня інформатизації i -го регіону отримаємо із наступних формул:

$$f(x_i) = 1 - \frac{d_{i0}}{d_0}, \quad (4)$$

$$\text{де } d_0 = \frac{1}{m} \sum_{i=1}^m d_{i0} + a \left[\frac{1}{m} \sum_{i=1}^m (d_{i0} - \bar{d})^2 \right]^{\frac{1}{2}} \quad (5)$$

відстань від точки P_i до точки P_0 :

$$d_{i0} = \left[\sum_{j=1}^n (z_{ij} - z_{0j})^2 \right]^{1/2}, \quad (i = \overline{1, m}) \quad (6)$$

де a - деяке додатне число, яке вибирають таким чином, щоб усі значення $f(x_i)$ знаходилися між нулем і одиницею (найчастіше $a = 3$). Згідно запропонованої моделі обчислення функції переваги, значення кожного із індексів в ідеалі рівне 1. Тобто, чим значення функції переваги $f(x_i)$ в i -му регіоні ближче до 1, тим вищим є рівень інформатизації регіону.

Наступним кроком в оцінюванні рівня інформатизації регіонів є їх групування за цим показником або розбиття на кластери. У кожен кластер повинні потрапити регіони, що мають схожі характеристики. Схожість або відмінність між регіонами із застосуванням

кластерного аналізу встановлюється залежно від метричної відстані між ними. Якщо кожен регіон описується n ознаками, то він може бути представлений як точка в n -мірному просторі, і схожість його з іншими регіонами визначатиметься їх взаємним розташуванням, а отже, і відстанню між ними.

Таким чином для кластеризації регіонів має бути побудована матриця відстаней із застосуванням однієї із метрик відстаней: Евклідова, зважена Евклідова, Манхеттенська, Чебишева, Мінковського, Махалобіса D^2 та ін. Найчастіше в якості відстані між двома спостереженнями z_i і z_l використовують "зважену" евклідову відстань:

$$\rho_{il} = \sqrt{\sum_{k=1}^m w_k (z_{ik} - z_{lk})^2} \quad (7)$$

Вагу k -го показника можна визначити за правилом Фішберна [2]:

$$w_k = \frac{2(n-k+1)}{n(n+1)} \quad (8)$$

Якщо $w_k = 1$ для всіх $k = \overline{1, n}$, то отримуємо звичайну евклідову відстань.

Отримані значення зручно представити у вигляді матриці відстаней:

$$R = (\rho_{ij}) \quad j = \overline{1, m}, \quad i = \overline{1, m} \quad (9)$$

Матриця R симетрична: $\rho_{ii} = 0$, а $\rho_{il} = \rho_{li}$.

Використовуючи матрицю відстаней, можна реалізувати агломеративну ієрархічну процедуру кластерного аналізу. Найбільш відомими та широко застосовуваними методами формування кластерів є: одиничного зв'язку (метод близького сусіда); повного зв'язку (метод далекого сусіда); середнього зв'язку; метод Уорда. Простим у реалізації є метод одиничного зв'язку (передбачає приєднання одиниці сукупності до кластера, якщо вона близька (знаходиться на одному рівні схожості) хоча б до одного представника цього кластера) [3, с. 94-119], зокрема із використанням можливостей пакету Statistica 8.0.

Додаткову характеристику якості розбиття сукупності об'єктів на кластери дає статистична перевірка гіпотез [4]. При цьому, припускається, що в середині утворених кластерів вони підлягають n -мірному нормальному розподілу з центром \bar{n}_j та однаковою коваріаційною матрицею W . Нульова гіпотеза формулюється наступним чином: в генеральній сукупності центри кластерів не відрізняються між собою ($H_0 : \bar{n}_i = \bar{n}_j$). Отже, відповідні їх елементи насправді належать одній і тій же генеральній сукупності, і розподіл на класи не є обґрунтованим. Альтернативною є гіпотеза, що елементи кластерів належать різним генеральним сукупностям.

Перевірку справедливості нульової гіпотези здійснюють за допомогою F -критерію Фішера. Спостережене (емпіричне) значення критерію розраховується за формулою

$$F_{EMP} = \frac{d_{ij}^{(M)}(m-n+1)k_i k_j}{mn(k_i + k_j)} \quad (10)$$

де $d_{ij}^{(M)}$ - відстань Махаланобіса між центрами кластерів; k_i та k_j - кількості елементів в кластерах; m - кількість об'єктів досліджуваної сукупності; n - кількість відібраних ознак.

Розраховане значення порівнюється з критичним, знайденим при рівні значущості α та ступенях вільності $V_1 = n$ у, та $V_2 = m$. Якщо $F_{EMP} < F_{KP}(\alpha, V_1, V_2)$, нульова гіпотеза приймається.

Аналогічно можна перевірити більш загальну гіпотезу про рівність середніх значень одразу всіх кластерів. Нульова гіпотеза в такому випадку матиме вигляд $H_0: \bar{n}_i = \bar{n}_j = \dots = \bar{n}_k$. Для перевірки її справедливості можна використати критерій Уїлкінса. Спостережене значення критерію обчислюється за виразом:

$$\Lambda = \frac{\det W}{\det C} \quad (11)$$

де C - сумарна матриця коваріацій вихідної сукупності ознак W та зважених коваріацій центрів кластерів,

$$C = W + \frac{\sum_{j=1}^k k_j (\bar{n}_j - \bar{n})^T (\bar{n}_j - \bar{n})}{m} \quad (12)$$

де \bar{n} - середнє значення (центр) всієї вихідної сукупності об'єктів. Перша складова матриці C характеризує розсіювання всієї вихідної сукупності об'єктів, а друге - міжкластерне розсіювання.

Оскільки Λ -розподіл Уїлкінса має досить складний вигляд, для перевірки нульової гіпотези використовують F -критерій Фішера. Емпіричне значення критерію розраховується за формулою

$$F_{EMP} = \frac{(1-\Lambda)^{1/n_\Lambda} (n^* n_\Lambda + 1 - \frac{nk}{2})}{\Lambda^{1/k_\Lambda} nk} \quad (13)$$

де $n^* = m - (n - k + 1) \setminus 2$

$$n_\Lambda = \begin{cases} \frac{(kn)^2 + 4}{k^2 + n^2 - 5}, & \text{при } k^2 + n^2 - 5 \neq 0 \\ 1, & \text{при } k^2 + n^2 - 5 = 0 \end{cases} \quad (14)$$

Спостережене значення критерію порівнюється з критичним, знайденим при рівні значущості α та ступенях вільності $V_1 = nk$ у, та $V_2 = n \cdot n_\Lambda - n + \frac{k}{2}$. Якщо $F_{EMP} < F_{KP}(\alpha, V_1, V_2)$, нульова гіпотеза приймається.

З виразу (12) випливає, що величина Λ знаходиться в межах від 0 до 1. Аналіз формули (13) показує, що малі значення призводять до великих значень F_{EMP} , а отже, до висновку про обґрунтованість розбиття сукупності на кластери. І навпаки, близькі до 1 значення Λ свідчать про безпідставність такого поділу на однорідні групи. Дійсно, близькість Λ до нуля є наслідком високих значень визначника матриці C , яка характеризує високий ступінь міжгрупової дисперсії. Випадок близькості до нуля визначника матриці W також може бути причиною малих значень Λ . Це є наслідком мультиколінеарності вихідних ознак, а отже і їх невдалого відбору для аналізу.

Найчастіше для визначення оптимальної кількості кластерів у процесі класифікації використовують показник ентропії. Оптимальною буде кількість кластерів, яка приводить до найбільшої ентропії. Ентропія визначається за формулою [5, с. 160-174]:

$$H = -\sum_{i=1}^k \frac{N_i}{m} \log_2 \frac{N_i}{m} = \log_2 m - \frac{1}{m} \sum_{i=1}^k N_i \log_2 N_i, \quad (15)$$

де H – ентропія класифікації; m – кількість об'єктів у досліджуваній сукупності (регіонів); N_i – кількість об'єктів (регіонів), що потрапили в i -ий кластер; k – кількість кластерів.

Максимально можливе значення ентропії H_{\max} визначається за формулою (15) при значеннях N_i рівних між собою, тобто кількість об'єктів рівномірно розподілена в кластерах.

Відхилення ентропії від максимального значення визначається наступним чином:

$$\Delta H = \frac{(H_{\max} - H)}{H_{\max}} \cdot 100 \quad (16)$$

Вибір кількості кластерів може визначитися на основі аналізу спеціальних функціоналів якості, на основі порівняння розбивок на різну кількість класів, можливості зі змістовною інтерпретацією й інших критеріїв, проте остаточний вибір залишається за дослідником.

Слід зазначити, що процес інформатизації регіонів відбувається точково. Обчислювальна та комунікаційна техніка, телекомунікаційні мережі, бази і банки даних та знань, інформаційні технології (ІТ), системи інформаційно-аналітичних центрів різного рівня, виробництво технічних засобів інформатизації, системи науково-дослідних установ та підготовки висококваліфікованих фахівців спочатку виникають і проваджуються у великих містах, а потім дифузно поширюються в глиб регіону по районних центрах і т.д. Крім того,

цей процес інтенсивніше протікає у регіонах із високим рівнем соціально-економічного розвитку. Тим самим обумовлюючи дивергенцію регіонів за рівнем інформатизації. Зважаючи на те, що будь-який процес має межі зростання, то рано чи пізно настане конвергенція регіонів за їх рівнем інформатизаційного розвитку. Питання лише коли?

Існують різні емпіричні підходи для вимірювання зближення рівнів розвитку. Найвідомішим є метод перехресних регресій [6]:

$$\ln\left(\frac{y_t}{y_0}\right) = b \ln(y_0) + g'x \quad (17)$$

де y_t, y_0 - показники ВВП на душу населення за паритетом купівельної спроможності на початку та в кінці періоду, що розглядається;

x – вектор пояснювальних змінних.

Від’ємний коефіцієнт b інтерпретується як доказ умовного зближення.

Слід зазначити, що рівняння (17) зазвичай оцінюється у нелінійній формі, тобто замість коефіцієнта b обчислюється вираз $(-1 + e^{-\beta T})$, де T – період часу, що досліджується. Коефіцієнт β дає швидкість, з якою знижується початковий диференціал логарифма доходів на душу населення. Цей тип зближення часто називають умовним зближенням або „ β -зближенням”.

У контексті нашого дослідження в якості пояснювальних змінних слід обрати ті показники, що безпосередньо впливають на процес інформатизації (x_1, x_2, \dots, x_{20}), залежної змінної y_t, y_0 - показники валового продукту регіонального на душу населення з урахуванням паритету купівельної спроможності початку та в кінці періоду, що розглядається.

Тоді економетрична модель для оцінки рівня зближення регіонів за рівнем інформатизації матиме наступний вигляд:

$$\ln\left(\frac{y_t}{y_0}\right) = a + b_1 \ln(y_0) + b_2 \ln(x_1) + b_3 \ln(x_2) + b_4 \ln(x_3) + \dots + b_{21} \ln(x_{20}) + u^* \quad (18)$$

де u^* - можливий залишок (стохастична складова); a – вільний член;

Дану систему рівнянь розв’язують методом найменших квадратів. Отримавши значення b_1 , можна обчислити швидкість щорічного зближення між регіонами за вкладом інформатизації у розвиток економіки регіону:

$$b_1 = (-1 + e^{-\beta T}) \Rightarrow \beta = -\ln(b_1 + 1) / T \quad (19)$$

де T – кількість років досліджуваного періоду.

Також можна обчислити кількість років необхідних для подолання половини розриву між регіонами [7, с. 201]:

$$100\%(1 - \beta)t = 50\% \Rightarrow (1 - \beta)t = 0,5 \Rightarrow t = \frac{\ln(0,5)}{\ln(1 - \beta)} \quad (20)$$

Аналогічним чином можна обчислити через скільки років розрив між регіонами за рівнем інформатизації становитиме 1% від наявного розриву [7, с. 201]:

$$100\%(1 - \beta)t = 1\% \Rightarrow (1 - \beta)t = 0,01 \Rightarrow t = \frac{\ln(0,01)}{\ln(1 - \beta)} \quad (21)$$

Проведенні обчислення на попередніх етапах є підґрунтям для окреслення можливих напрямків подолання існуючого розриву між регіонами у рівні їх інформатизації (етап IV).

Висновки. Комплексна система оцінювання рівня інформатизації регіонів України базується на індексній оцінці розвитку регіональної інформатизації, об'єднанні регіонів за її станом у споріднені групи методами кластерного аналізу, дослідженні та аналізі дивергентно-конвергентних процесів в інформатизації регіонів методом перехресних регресій із визначенням перспектив у зближення регіонів за цим показником у сучасних умовах розвитку України. Вона може бути корисною при прийнятті обґрунтованих рішень щодо визначення пріоритетів політики регіонального розвитку, окресленні стратегічних завдань інформатизації регіонів тощо.

Література

1. Плюта В. Сравнительный многомерный анализ в экономических исследованиях: методы таксономии и факторного анализа / Плюта В. - М. : Статистика, 1980. – 151 с.
2. Fishburn P. Utility Theory for Decision-Making. N.Y., Wiley, 1970.
3. Григорук П.М. Багатомірне економіко-статистичне моделювання: Навчальний посібник / П.М. Григорук. – Львів: «Новий світ - 2000», 2006. – 149 с.
4. Янковой А.Г. Многомерный статистический анализ в системе STATISTICA / А.Г. Янковой. - Одесса: Оптимум, 2001. Вып.1. - 216 с.
5. Колемаев В.А. Экономико-математическое моделирование / В.А. Колемаев. – М.: Юнити – Дана, 2005. – 295 с.
6. Данилюк В. Чи досягне Україна економічного рівня країн-членів Європейського Союзу? [Електрон. ресурс] / В.Данилюк. – Режим доступу: <http://newdemocracy.org.ua/files/ua-eu.pdf>
7. Красс М.С. Математика в экономике. Математические методы и модели: учебник. / Красс М.С., Чупрынов Б.П. – М.: Финансы и статистика, 2007. – 544 с.