

*Опря А. Т., доктор економічних наук, професор
Полтавська державна аграрна академія*

НАУКОВА КОНЦЕПЦІЯ СТАТИСТИЧНОЇ МЕТОДОЛОГІЇ: МЕТОДИ, ПОКАЗНИКИ, КРИТЕРІЇ НАДІЙНОСТІ

Рецензент – доктор економічних наук, професор В. К. Савчук

Розглянуто концептуальні підходи статистичної методології в дослідженні взаємозв'язків економічних явищ і процесів. Обґрунтована необхідність застосування статистичних прийомів у поглибленому дослідженні причинно-наслідкових зв'язків із позицій системного підходу їх використання: методи, статистичні показники, оцінки надійності.

Зроблена спроба поєднати емпіричний аспект дослідницької роботи з науковою методологією, виходячи з концептуальних положень статистичної науки.

Ключові слова: методологія, математико-статистичні методи, одиниці спостереження, статистичні показники, критерії вірогідності, економічні явища.

Постановка проблеми. Розглядаючи методику досліджень як науку про методи, під останніми ми розуміємо систему прийомів вивчення сутності явищ і процесів. Базуючись на класичному визначенні поняття «метод» (спосіб теоретичного дослідження або практичного здійснення будь-чого), у дослідженні нами враховувалися дві взаємопоєднані його сторони: формально-логічна та змістовно-генетична. Мається на увазі, що способи і прийоми проведення дослідження здійснюються у нерозривному зв'язку з теоретичними узагальненнями з використанням системи законів, категорій і принципів діалектики розвитку досліджуваних процесів та явищ (у даному випадку – економічних). Під поняттям «методологія» розуміють сукупність пізнавальних принципів дослідження, методів і способів його організації, що використовуються у будь-якій науці [7].

Навчальна література з теоретичної і прикладної статистики містить недостатньо матеріалу з питань статистичної методології, а фрагментарний виклад їх не завжди забезпечує вирішення питань, передбачених завданнями дослідження, зокрема економічних явищ. Особливістю вивчення останніх є те, що в дослідженнях не слід обмежуватися лише емпіричним аналізом. Останній являє собою початкову стадію науки і з розвитком наукового пізнання такий аналіз по-

ступається з методологією і теорією, що дають змогу дослідити складний механізм взаємозалежностей явищ і процесів [5].

Виходячи з зазначеного вище, нами зроблена спроба поєднати емпіричний аспект дослідницької роботи з науковою методологією, спираючись на концептуальні положення статистичної науки щодо методів, статистичних показників і оцінок надійності інформаційного поля у форматі початкових і заключних стадій дослідження (накопичення, представлення, аналіз, інтерпретація).

У дослідженнях, як правило, застосовують такі основні методи наукового пізнання, як експериментальний, науково-конструктивний і статистичний [4]. Використання останнього базується на масових статистичних даних як на етапах підготовки дослідження (експерименту), так і в процесі його проведення. Зауважимо: масовість первинного статистичного матеріалу забезпечує вірогідність результатів аналітичних розрахунків, об'єктивність висновків і науковість методологічних підходів у дослідницькій роботі.

У цілому використання статистичних прийомів у дослідженні включає використання (окремо і в комплексі) наступних методів статистики: масового спостереження; статистичних групвань із використанням узагальнюючих і аналітичних показників (середніх величин, показників варіації та ін.); статистико-математичного аналізу причинно-наслідкових зв'язків із використанням статистичних рядів розподілу, дисперсійного і кореляційно-регресійного аналізу, графічних побудов, теоретичного узагальнення.

Одним із недоліків організації дослідницької роботи в галузі економіки слід назвати наявність методологічних помилок у її здійсненні, в результаті чого маємо наступні недоліки: залишається нерозкритою ціла низка факторів причинно-наслідкових моделей економічних явищ і процесів; збідненим виявляється прикладний аспект дослідження проблеми; практичні висновки обмежуються аналізом (інколи лише співставленням) рівнів або їх співвідношень. У дослідженнях прикладного аспекту нерідко можна

спостерігати утилітарний кількісний підхід, коли конкретні показники економічних явищ розглядаються поза зв'язком з їх економічною природою чи політекономічним змістом. Слід пам'ятати, що економічний показник – це діалектична єдність якісної та кількісної сторін досліджуваного явища. Тому для його характеристики потрібен поглиблений аналіз із широким залученням традиційних і сучасних методів статистики, зокрема математичної статистики. Отже, постає питання про необхідність подальшого вдосконалення методики статистико-економічного аналізу з метою забезпечення системного дослідження тих чи інших явищ і процесів суспільного життя.

У даній статті зроблена спроба комплексного підходу у дослідженні певної проблеми, коли методологія статистики розглядається у форматі трьох напрямів, а саме: основні статистичні прийоми кількісного аналізу причинно-наслідкових зв'язків; об'єктивність економічних показників; критерії вірогідності показників інформаційного поля досліджень. Третій із зазначених напрямів слід вважати особливо важливим у дослідницькій роботі, оскільки на цьому етапі знаходить підтвердження об'єктивність вихідної, проміжної й результативної інформації.

У літературі такий підхід наразі ще не одержав достатньо повного висвітлення. Більше того, до цього часу не здійснено системної структуризації названих вище напрямів практичної реалізації статистичної методології в економічних дослідженнях.

Аналітичний огляд основних публікацій з даної проблеми. Теоретико-методологічним і прикладним аспектам дослідження проблемних питань статистики свої праці присвятило чимало вчених дальнього зарубіжжя, зокрема: Д. Вайнберг, В. Венсель, М. Езекиєл, М. Кендел, К. Пірсон, Ч. Спірмен, Р. Фішер, К. Фокс, Р. Фріш, А. Хьютсон, Е. Юл та інші.

Щодо розвитку статистичної науки на теренах колишнього Радянського Союзу, значну наукову спадщину зі складних і актуальних проблем теорії та методології залишили вчені-статистики: А. Я. Боярський, І. Г. Венецький, А. М. Гатаулін, М. К. Дружинін, Л. О. Казинець, Г. С. Кільдишев, Я. І. Лукомський, В. С. Немчинов, С. С. Сергеев, Є. С. Слущкий, Н. С. Четвериков, О. О. Чупров, Б. С. Ястремський та інші.

Серед вітчизняних учених методологічні аспекти статистичної науки знайшли відображення у працях: В. Ю. Андрієнка, О. А. Бугуцького, С. С. Герасименка, А. В. Головача, А. М. Єріної,

В. Б. Захожая, О. І. Кулинич, Р. М. Моторіна, Н. О. Парфенцевої, Й. С. Пасхавера, М. В. Птухи, В. П. Трофімова, Є. В. Чекотовського, В. Г. Швеця та інших.

Водночас, розглядаючи теоретико-методологічні засади формування й розвитку статистичної науки в контексті системного підходу щодо використання статистичних і математико-статистичних методів у дослідницькій роботі, актуалізується питання необхідності (доцільності) подальшої розробки даного напрямку. Науковий вклад у статистику названих вище вчених полягає в тому, що в їхніх працях статистичні методи розглядаються, зазвичай, із позицій автономного (ізолюваного) їх використання в дослідженнях. У їхніх працях пріоритети надаються питанням важливості статистичних методів у дослідженні, але знайти хоча б орієнтири на системний підхід у використанні – це виняток. Окремі розробки зорієнтовані на уявлення про функції статистичних методів із позиції їх додаткової ролі у виконанні завдань дослідження без акцентів на можливість вивчення глибинних зв'язків і відношень, що визначають закономірності, пропорції, тенденції, властивості й специфіку явищ. Такий широкий спектр завдань неможливо вирішити без системного підходу в процесі використання прийомів і методів статистики. До того ж взаємозв'язки економічних явищ відрізняються складністю їх структурних побудов, що утворюють причинно-наслідкову залежність, і кількісно виміряти дію факторів впливу у такій залежності можливо лише за умов комплексного підходу в застосуванні сучасних і традиційних методів статистики. Вирішенням цих питань зумовлюється актуальність розробки теоретико-методологічних і прикладних напрямів, як гаранту надійності інформаційного забезпечення на всіх етапах дослідницької роботи, починаючи з характеристик вірогідності вихідної інформації й закінчуючи оцінкою надійності результатів дослідження.

Мета дослідження – висвітлити можливості статистичної науки у поглибленому кількісному аналізі соціально-економічних явищ і процесів.

Завдання – розкрити методичні особливості та специфіку комплексного використання статистичних і математико-статистичних прийомів у контексті забезпечення надійності інформаційного поля дослідження з метою досягнення вірогідності, а отже, об'єктивності його результатів.

Результати дослідження. Оскільки методологічною основою наукового дослідження, як правило, виступають статистичні й математико-

статистичні методи, зупинимося більш детально на їх використанні та науковій результативності (мова йтиме про основні).

Масові явища суспільного життя складаються з одиниць, що різняться між собою як кількісно, так і якісно. Ці різниці можуть зумовлюватись істотними і неістотними (випадковими) факторами, що утворюють складну й неоднорідну сукупність одиниць спостереження (вихідна інформація). Тому одне з головних завдань статистики полягає в тому, щоб розділити таку складну сукупність одиниць на однорідні всередині, але істотно відмінні між собою сукупності, а також своєчасно виявити за варіацією кількісних одиниць якісні їх переходи. Таке завдання статистика вирішує за допомогою методу статистичних групувань.

Використання названого методу у дослідженнях економічних явищ і процесів базується на концепції, що статистичне групування – це процес утворення необхідних у тому чи іншому відношенні груп, який здійснюється за наявною статистичною інформацією. Це єдиний аналітико-синтетичний процес, у результаті якого утворюються відповідні групи. У ньому знаходить свій прояв математичний «Закон великих чисел», адже у показниках, обчислених на достатньо великій кількості одиниць спостереження, взаємонейтралізується випадковість і лишається те істотне, що необхідне дослідникові.

Метод статичних групувань, окрім характеристик взаємодії всієї різноманітності факторів варіації (зокрема, економічних показників) доводить їх істотність у дослідженні тих чи інших соціально-економічних явищ. У такому разі потрібно звертати увагу на суть методу, а саме на те, що складне явище (а в економічних моделях воно наявне) розглядається не як єдине ціле, а в ньому виділяються окремі групи одиниць, показники яких дають кількісну оцінку всій досліджуваній сукупності. Доречно нагадати, що сам метод статистичних групувань було розроблено стосовно вивчення суспільних явищ, і тому він знаходить широке застосування в аналізі саме соціально-економічних процесів, що відбуваються у суспільстві. Завдяки даному методу статистика вважається одним із наймогутніших знарядь соціального пізнання. З допомогою прийомів групувань вдається вирішувати у наукових економічних дослідженнях наступні завдання: виділення соціально-економічних типів явищ; характеристика структури досліджуваного явища; вивчення зв'язків і взаємозалежностей між явищами. Відповідно до цих завдань статистич-

ною наукою розроблено типологічні, структурні й аналітичні види групувань.

Підкреслимо, науковому статистичному групуванню передують теоретико-економічний аналіз досліджуваного явища. Водночас використання сучасних статистичних прийомів дає можливість кількісно оцінити ступінь однорідності виділених груп, здійснити відбір істотних ознак, удосконалити методіку визначення величини інтервалів групувань.

У дослідженнях науковці нерідко допускають помилки в розв'язанні питання про утворення груп (окрім групувань за атрибутивною ознакою, де обмежуються кількістю значень ознаки). Стосовно цього існують різні методичні підходи [2]. Визначення числа груп, запропоноване Стерджессом (такі рекомендації містять майже всі підручники і посібники з теорії статистики), яке полягає у розрахунку формули $n_{int} = 1 + 3,322 \lg n$, звільняє від можливості економічного мислення; сам прийом слід визнати формальним, а в окремих випадках – небезпечним. Слід визнати найвдалішими рекомендації В. П. Левінського, який пропонує своєрідні нормативи числа інтервалів (груп), зумовлені обсягами досліджуваної сукупності. Перевага цих рекомендацій, порівняно з пропонованими Стерджессом, полягає у тому, що за такого підходу не так жорстко пов'язано число груп із чисельністю одиниць спостереження: у такому разі дослідник має можливість певного вибору числа груп залежно від характеру сукупності. В економічних дослідженнях найбільш поширений обсяг сукупності 100 – 500 одиниць. За Стерджессом, за такої чисельності одиниць кількість груп дорівнюватиме 7–10, а на думку В. П. Левінського – 10–17.

Вирішуючи питання стосовно вибору кількості груп по досліджуваному масиву інформації, слід пам'ятати, що дана кількість знаходиться у пропорційній залежності від зміни групувальної ознаки: чим вона більша, тим більше необхідно утворювати груп. Виділені групи повинні бути достатньо заповнені одиницями спостереження. Наявність незаповнених інтервалів або незначна кількість у них одиниць спостереження – це результат невдалого вибору інтервалів. Наявність мало наповнених груп (до 3 одиниці) допустиме лише по краях групування, де концентруються найвищі й найнижчі показники відносно середнього рівня. Зазначена методична особливість насамперед стосується виду структурних групувань. Кількість груп тут не повинна бути досить великою чи досить малою. У першому випадку виникає ризик загубитися у дрібницях, у друго-

му – не виявити досить важливі властивості досліджуваної сукупності.

Отже, спираючись на викладене вище, можна впевнено стверджувати, що практичному застосуванню методу статичних групувань передують ретельний теоретичний аналіз факторів, виявлення головного напрямку розвитку досліджуваного явища, а також виділення зі складної сукупності окремих груп одиниць, які належать до різних типів. Разом із тим слід зазначити, що попередній теоретичний підхід при здійсненні групувань не є догмою й не означає, що даний метод відіграє пасивну технічну роль. Його використання дає змогу одержати кількісну характеристику стану досліджуваного явища, виявити якісні його перетворення, перевірити наукові гіпотези відносно розвитку явища.

Поскілки досліджувані ознаки статистичної сукупності завжди варіюють, виникає потреба у визначенні центрального цифрового її значення, яке кількісно відображало б кількісно той типовий рівень, яким наділені окремі елементи досліджуваної сукупності й навколо якого мають тенденцію зосереджуватися конкретні спостережувані значення ознаки. Статистична методологія розглядає кілька різних таких показників центральної тенденції ознаки. Найбільш поширеною є середня арифметична величина, хоча в конкретних випадках розрахунків використовують і інші види середніх, зокрема: геометричну, гармонійну, квадратів (кубів тощо), хронологічну, прогресивну, моду, медіану [2].

Середня величина вважається найважливішою характеристикою варіаційного ряду, поскілки вона відображає об'єктивне існування певних умов, які проявляються в кожній одиниці досліджуваної сукупності. Свою основну пізнавальну функцію середня величина може виконувати лише за умов правильного її обчислення: вона вважається науково обґрунтованою, якщо методологічні принципи її розрахунку відповідають наступним вимогам.

Розрахунок середніх повинен здійснюватися за однорідними, одноякісними явищами. Існує досить поширена помилка – визначення середньої для неоднорідної сукупності, коли середня розглядається як міра дії основних причин, а відхилення від неї – як випадкових. Даний підхід деформує дійсну структуру статистичної сукупності й зумовлює неможливість викриття причин, котрі її викривляють. Обчислена в такий спосіб середня втрачає свій реальний зміст, перетворюється на фікцію.

Водночас із зазначеним вище відмітимо:

з-поміж усереднювальних ознак можуть зустрічатися й досить відмінні одна від одної (що цілком виправдано). Зрозуміло, що в процесі розвитку явищ виникає нове і відмирає старе, а тому значення ознак у представників нового й старого досить відмінні від середніх рівнів. Такі відмінності цікавлять статистику не менше (а інколи й більше), ніж середні величини ознак, адже нове і прогресивне в суспільстві підтримується, а виявлення старого, віджилого, повинно спонукати на подолання його залишків у досліджуваному явищі. У такому випадку не обмежуються розрахунком середніх показників, а доповнюють їх описом окремих явищ, які мають значну варіацію відносно цієї середньої величини.

Наукова обґрунтованість середніх вимагає правильного вибору явищ, тобто, досліджуваних одиниць сукупності, за якими розраховується середній розмір ознаки. Відповідно до цієї вимоги обчислення, наприклад, середньорічного показника надою від однієї корови, треба було б розраховувати відносно чисельності поголів'я дійних корів. У практиці даний середній показник розраховується відносно фуражного поголів'я. Такий підхід створює небажані «стимули» до зменшення чисельності недейних корів. За аналогічним принципом розраховується показник середньої врожайності. За його обчислення валовий збір ділиться не на збиральну чи фактично зібрану площу, а на весняну продуктивну площу. Остання включає площі, на яких навесні може статися загибель посівів або вони залишаться незібраними.

Середній показник повинен обчислюватися за всіма явищами, а це означає, що він спирається на всі одиниці спостереження, в іншому разі він не буде репрезентувати всю сукупність. У випадках, коли середня розраховується по частині досліджуваної сукупності, вона повинна бути представлена її типовим фрагментом. Статистична методологія передбачає такі випадки – згідно з науковими рекомендаціями, типовість середньої може бути забезпечена чисельністю одиниць не менше ніж 25–30 [6]. Якщо середня розраховується за меншою чисельністю, то досліджувані ознаки повинні відрізнятися одна від одної несуттєво.

Статистична наука враховує й інші не менше важливі властивості середніх величин, зокрема: зручність в обчисленні; незалежність від випадкових коливань; простота алгебраїчних перетворень; відсутність абстрактного математичного характеру; незалежність від суб'єктивної оцінки; чітка визначеність.

У процесі дослідження (чи аналізу) соціально-економічних явищ із використанням середньої слід пам'ятати, що вона дає узагальнюючу характеристику лише за однією ознакою, тоді як природа таких явищ досить різноманітна (а отже, складна) й кожна з них має багато властивостей та ознак. Тому аналізуючи такі явища не обмежуються обчисленням однієї ізольованої середньої, щоб всебічно розкрити явище, використовується ціла система середніх величин. У разі вибору шляху поглибленого аналізу, систему середніх показників застосовують у комплексі з іншими методами статистики.

Зауважимо, що поглиблене дослідження соціально-економічних явищ неможливе без використання методів математичної статистики. Застосування їх у аграрно-економічних дослідженнях стало нагальною необхідністю і потребою.

У багатофакторному аналізі соціально-економічних явищ і процесів практично не використовується досить ефективний спосіб математико-статистичної обробки даних – дисперсійний метод. Як і інші ймовірно-статистичні методи, він набагато розширює можливості економістів (практиків чи науковців), значно підвищуючи науковий рівень досліджень та їх результативність. Виходячи з сучасних вимог до рівня економічного аналізу, такий метод повинен зайняти в ньому одне з провідних місць, оскільки за його допомогою розв'язуються різнопланові завдання. По-перше, маючи самостійне значення у вивченні причинно-наслідкових зв'язків, метод забезпечує: 1) кількісний вимір сили впливу досліджуваних факторних ознак та їх сполучень на досліджувану результативну ознаку; 2) визначення вірогідності (достовірності) впливу та його довірчих меж; 3) аналітичний вимір змін окремих середніх і статистичну оцінку різниці між ними. По-друге, у поглибленому економічному аналізі метод дисперсійного аналізу може виконувати допоміжні функції. У цьому напрямі його використання відкриває широкі можливості щодо науково обґрунтованого підходу до застосування інших методів статистики у кількісному аналізі.

Розглядуваний метод, як й інші математико-статистичні прийоми, – це суто технічний засіб наукового пізнання. Тому його використання передбачає насамперед знання суті (економічної природи) досліджуваних явищ, розуміння природи виникнення й дії в них причинно-наслідкових зв'язків, а також вміння виділити найважливіші сторони їх залежностей та взаємозумовленостей.

Оскільки метою і завданням даної статті передбачався виклад не технічних, а методологічних особливостей використання даного методу в аналізі соціально-економічних явищ, зупинимося на них детальніше.

Використовуючи дисперсійний метод у форматі виконання ним самостійних функцій на останньому етапі моделювання вирішують питання про надійність (вірогідність) одержаних результатів розрахунку. Тут може мати місце той факт, коли дія окремих факторів та їх сполучень виявляється не вірогідною, тобто фактичні значення критерію надійності Фішера нижчі від їх стандартного математичного (табличного) значення. Окремі аналітики у такому разі ведуть мову не про недоведеність вірогідності впливу, а про доведеність неіснування (відсутності) впливу факторних ознак на результативний показник. Помилковість такого твердження очевидна, поскільки сама природа економічного явища свідчить про наявність залежності, а природа (математична) критерію – про недоведеність у конкретному випадку наявності вірогідного впливу. У такому разі слід іти шляхом поглиблення дослідження: з'ясувати питання щодо відповідності досліджуваної сукупності законам розподілу (Гаусса, Стюдента, Пірсона та ін.); розширити обсяг одиниць досліджуваної сукупності, що забезпечить можливість спрацювати «Закону великих чисел»; перевірити характер розподілу на асиметрію та ексцес тощо.

Математична природа методу дисперсійного аналізу відкриває можливість об'єктивної оцінки результатів статистичних групувань (особливо складних – комбінаційних їх видів), визначення істотності різниць середніх, надійності коефіцієнтів кореляції, оцінки лінійної регресії.

Прикрим, на жаль, слід визнати той факт, що даний метод у підручниках і навчальних посібниках викладено не для економічних напрямів підготовки, – а в такому випадку використовувати алгоритми в обчисленні дисперсійних комплексів для економічних моделей неможливо (даний недолік усунено у навчальних виданнях за нашим авторством).

Використовуючи метод дисперсійного аналізу в дослідженні, слід пам'ятати про суттєву особливість (можливо, недолік) даного прийому математичної обробки інформаційного поля, зокрема це залежність результатів аналізу від рівня показників сформованих груп (градацій) за досліджуваними факторами. Тобто, дисперсійні моделі, сформовані за одних рівнів факторних ознак, можуть давати вірогідний вплив, за інших

– його відсутність. Наукова методологія в такому разі передбачає різні способи формування моделей. Дисперсійні комплекси можуть бути побудовані за принципами рівномірного, нерівномірного, пропорційного, ортогонального розподілу. Потрібно також орієнтуватись у методичних підходах до побудови груп із різною чисельністю одиниць спостереження. Саме науковий підхід у застосуванні даного методу в дослідженні соціально-економічних явищ і процесів забезпечує йому великі переваги з-поміж інших статистичних методів багатфакторного кількісного аналізу.

Надаючи важливого значення вищевикладеному методу, дослідникам варто звернути увагу й озброїтися таким ефективним прийомом аналізу причинно-наслідкових зв'язків як кореляційно-регресійний метод. Зауважимо, в літературі він зустрічається, як правило, під назвою «кореляційний метод» (або «кореляційний аналіз»), що слід вважати неправомірним із точки зору вирішуваних ним аналітичних завдань. Кореляційний метод (кореляційний аналіз) вирішує наступні завдання: оцінку параметрів нормально розподіленої досліджуваної (генеральної) сукупності (генеральних середніх, дисперсій парних коефіцієнтів кореляції), множинних і окремих коефіцієнтів кореляції; перевірку істотності оцінюваних параметрів та одержання інтервальних оцінок для визначення істотних серед них; виявлення структури взаємозалежності ознак. Регресійний метод (регресійний аналіз) передбачає вирішення таких завдань: ступінь відокремленого і спільного впливу факторів на результативну ознаку, а також кількісну оцінку цього впливу.

Зазначимо, метод кореляційно-регресійного аналізу знайшов достатнє висвітлення у навчальній і науковій літературі, однак автори теоретичних і практичних сторін його використання інколи не враховують методологічних особливостей даного математико-статистичного прийому. Що тут слід пам'ятати? По-перше, математична природа кореляційно-регресійного зв'язку орієнтує на те, що це суто математичний прийом, який не розкриває фізичної картини взаємозв'язків, а лише констатує їх наявність. По-друге, кореляція хоча й не виявляє причин зв'язку, але дає оцінку встановленим зв'язкам із позиції їх сили й тісноти, а також встановлює ступінь вірогідності міркувань про наявність зв'язків. Тому, аналізуючи соціально-економічних явище дослідник, керуючись методологічними принципами (правилами) кореляційно-регресійного аналізу, насамперед повинен вихо-

дити з економічного змісту досліджуваних залежностей; після цього може бути встановлений і кількісно вимірний їх причинно-наслідковий характер.

Помилковим слід визнати методичний підхід, коли дослідник із метою одержання корисної практичної інформації намагається розглянути (моделює) велику кількість факторів, до того ж окремі з них тісно пов'язані між собою. За таких обставин зміна одного з таких факторів зумовить зміну іншого (пов'язаного з ним) фактора, внаслідок чого важко відокремити чистий вплив одного з них і задовольнити математичну природу множинної регресійної моделі. Методологічно більш правильним буде відібрати основні більш важливі фактори. З позиції дії «Закону великих чисел» вважається, що чисельність одиниць спостереження у досліджуваній кореляційно-регресійній моделі повинна перевищувати кількість обраних факторів у 6–8 разів (краще – у 8–10 разів).

Суттєві методологічні особливості має кореляційно-регресійний аналіз часових рядів і рядів динаміки. До речі, у спеціальній літературі ці два статистичні поняття ототожнюються, що слід визнати помилковим. Поняття «динаміка» – більш широке, охоплює не тільки часові зміни, а й будь-які зміни економічних показників, зумовлені дією зовнішніх умов. Дане питання має специфічні особливості у прикладному напрямі його вирішення. (Детальний розгляд методичних підходів у побудові таких моделей міститься в наших виданнях підручників і навчальних посібників).

Існує ще й така специфічна особливість моделювання кореляційно-регресійних зв'язків, в основу алгоритмів аналізу яких покладено непараметричні критерії. Про них мова йтиме пізніше.

Крім розглянутих вище статистичних методів дослідження причинно-наслідкових зв'язків є й інші, не менш ефективні математико-статистичні конструкції аналізу соціально-економічних явищ, такі як методи багатомірного статистичного аналізу (метод головних компонент, кластерний аналіз, розпізнавання образів, факторний аналіз та ін.). Їх практичне використання потребує спеціальних математичних знань.

Слід зауважити, що висвітлені методологічні й методичні особливості розглянутих вище основних статистичних методів дають ключ до розуміння творчого підходу в їх використанні не тільки ізольовано (самостійно), а й у комплексі. Останнє набуває особливого і виключно важливого значення, адже за таких умов дослідник (чи аналітик-

практик) має можливість поглиблено розібратися в існуванні, природі й дії факторів у причинно-наслідкових моделях зв'язків соціально-економічних явищ і процесів, що їх утворюють.

Наукові принципи застосування методів статистики у певному розумінні мають спільну методологічну природу зі статистичними показниками з позицій об'єктивності їх використання. У теоретичному і практичному форматі цим (статистичним) показникам властиві пізнавальні, управлінські, стимулюючі й ідеологічні функції. Такий широкий спектр їх «обов'язків» у пізнанні кількісної та якісної сторін суспільного життя вимагає уважного відношення до природи утворення явищ; обчислювальних і логічних дій; критеріїв науковості; часових і просторових особливостей явищ; репрезентативності інформаційних потоків; емпіричних і логіко-теоретичних шляхів утворення; сутності абсолютних і відносних розмірів явищ.

З огляду на виконувани статистичними показниками функції, можна сформулювати вимоги (або правила) їх наукового застосування. Виходячи з того, що природа соціально-економічних явищ досить багатопланова і специфічна, виконання пізнавальних функцій показників повинно забезпечуватися науковими принципами їх розрахунку. Мова йде не лише про послідовність обчислювальних логічних і інших дій (тобто про алгоритми), а й про теоретичну обґрунтованість, а також належність до емпіричної бази. Процес утворення статистичного показника базується на знанні філософії, економічної теорії, конкретної економіки й окремих спеціальних наук. Щодо теоретичної обґрунтованості – цей критерій науковості визначає суть статистичного показника. Під критерієм належності до емпіричної бази розуміють утворення показника на базі наукової інформації. Поскілки будь-яке дослідження починається з накопичення інформації, висновки, які акумулюються у статистичних показниках, матимуть об'єктивну основу у разі обґрунтованості їх конкретними фактами.

Викладені загальнонаукові принципи орієнтують на наступні вимоги до побудови статистичних показників: повнота вихідної інформації, порівнюваність і вірогідність (чи точність) даних [6]. Вимога щодо повноти показників передбачає як просторове, так і часове охоплення розвитку економічних явищ і процесів. Порівнюваність орієнтує на узагальнення показників у часовому й просторовому полі й стосується як абсолютних, так і відносних їх розмірів. Науковість цієї позиції вимог зводиться до наступного: спільний

зміст фактів; ідентичність вимірних і розрахункових дій; однорідність кола об'єктів і території. Зазначимо, питання співставності даних вважається певною мірою проблемним, і їх вирішення повинно базуватися на наукових принципах алгоритмізації інформації. Вірогідність (достовірність) статистичних показників передбачає ступінь їх наближення до відображуваної реальності. Поняття «вірогідність» інколи ототожнюють із поняттям «точність», розуміючи під останнім повну відповідність реальному розміру. У більш вузькому розумінні поняття точності показника використовують у процесі дослідження явищ, які формуються під впливом як закономірностей, так і випадковостей. Тому тут поняття точності доповнюється поняттям надійності оцінки точності, тобто ступенем ймовірності відхилення одержаного показника від реального. Така різниця називається похибкою (помилкою), а сам показник вважається вірогідним.

Таким чином, резюмуючи сказане щодо блоку методологічних питань статистичних показників, зазначимо, що утворенню та побудові їх системи передують досить складний шлях, пов'язаний у всіх його проявах і відтінках із суспільною практикою, яку слід вважати вихідною базою статистичного пізнання і дослідження, початковим етапом формування інформаційної бази.

Особливе місце у статистичній методології дослідження займають питання критеріїв надійності, тобто статистичних оцінок розрахунків та їх результатів. Оскільки будь-яке дослідження базується на даних спостереження (вихідна інформація), їх необхідно перевіряти на надійність, тобто об'єктивність. З цією метою статистична наука пропонує низку критеріїв.

Первинна інформація повинна охоплювати масові дані, адже в іншому разі, як зазначалося вище, не знайде свого прояву математичний «Закон великих чисел». Упорядковані первинні дані у вигляді рядів розподілу (варіаційних чи атрибутивних) випробовують на відповідність їх характеру нормального розподілу (закону Гаусса-Лапласа). Зауважимо, зустріти в ідеальному вигляді такий розподіл в економічних явищах майже неможливо, проте наближення до нього повинно перебувати у полі зору дослідника. Таке наближення дають характеристики асиметрії й ексцесу розподілу, тобто скошеність чи вершинність кривої Гаусса. Величина коефіцієнта асиметрії (A_S) може коливатися в межах 0 ± 1 . Його значення з абсолютною величиною $A_S \neq 0,5$ характеризують значне зміщення. Якщо $A_S \leq 0,25$ – асиметрія вважається незначною. При $A_S = 0$ – розподіл вважається симетричним.

Для кількісного виміру вершинності розподілу розраховують коефіцієнт гостровершинності – ексцес (E_X). Якщо $E_X=0,4$, крива вважається слабо ексцесивною. Для гостровершинних кривих $E_X>0$, для плосковершинних $E_X<0$. При помірній (нормальній) вершинності $E_X=0$.

Після перевірки надійності вихідної інформації дослідник, як правило, займається зведенням і групуванням даних спостереження. Згідно з вимогами статистичної науки, слід звернути увагу на досить важливий момент у практичному використанні результатів групувань – процес перевірки їх на вірогідність. У цьому зв'язку на початковому етапі здійснення статистичних групувань перевіряють «сумнівні» факторні ознаки на належність їх до статистичного ряду розподілу. З цією метою використовують Тау-критерій (τ). Так, перш ніж розрахувати величину рівномірного інтервалу, попередньо необхідно оцінити крайні показники ранжированого ряду розподілу на належність їх до останнього. Критерієм такої належності до досліджуваної сукупності виступає стандартизоване відхилення значень «сумнівних» показників (це, як правило, мінімальні, максимальні й близькі до них) від середнього їх рівня. Розмір стандартизованого відхилення не повинен перевищувати число 3, тобто

$$\tau = \frac{x_i - \bar{x}}{\sigma_x} \leq 3 \quad (\tau - \text{критерій належності; } x_i - \text{максимальне і мінімальне значення групувальної ознаки; } \sigma - \text{середнє квадратичне відхилення}).$$

На заключному етапі здійснення статистичних групувань одержані висновки по їх результатах необхідно також перевірити на надійність. Як зазначалося раніше, у нагоді тут допоміжні функції дисперсійного методу аналізу. Шляхом співвідношення факторної й залишкової дисперсій розраховують критерій Фішера (F -критерій) і порівнюють його з теоретичним значенням за стандартними математичними таблицями при відповідних порогах ймовірності (для економічних розрахунків – $P = 0,954$). Якщо розрахункове значення F -критерія дорівнює або перевищує теоретичне, результати статистичних групувань вважаються вірогідними, тобто надійними.

В аналітичній (дослідницькій) роботі досить часто мають справу з малою чисельністю одиниць спостереження. Так, економічні розрахунки можуть виконуватися на сукупності підприємств адміністративного району (10–25), зоотехнологічні дослідження – на невеликій кількості тварин (5–10 гол.), польові дослідження проводять на обмеженій кількості ділянок (4–5) і т. ін. У статистиці

такі чисельності називають малими вибірками. Мала сукупність одиниць спостереження зумовлює певні обмеження щодо використання вибірових даних для характеристики всієї (генеральної) досліджуваної сукупності. Враховуючи те, що малі вибірки можуть бути сформовані як залежні (спостереження однієї вибірки залежать від спостережень другої) і як незалежні (одиниці спостереження не протиставляються одна одній), оцінці підлягають середні показники. У залежних вибірках оцінюють середню різницю, в незалежних – різницю середніх. Таку статистичну оцінку визначають шляхом розрахунку критерія Стюдента (t -критерій), який являє собою нормоване відхилення середніх відносно стандартної помилки (σ). Співставленням розрахункового і стандартного (за математичними таблицями) значення t -критерію робиться висновок про істотність різниць вибірових і генеральних середніх.

Повернемося до великих вибірок, на матеріалах яких використовують розглянуті вище кількісні математико-статистичні прийоми дослідження, зокрема кореляційно-регресійний і дисперсійний методи. Що передбачає наукова методологія щодо об'єктивності отриманих результатів на базі їх алгоритмів?

У побудові кореляційних моделей необхідно врахувати наступний важливий момент методологічного характеру: у кореляційну модель не можна включати групу факторів, лінійна залежність яких дорівнює постійній величині або близька до неї. У таких випадках система нормальних рівнянь (при визначенні коефіцієнтів регресії) або не має рішення, або воно одержане в результаті випадках коливань досліджуваних ознак. У подібних випадках критерієм надійності результатів моделювання є коефіцієнт кореляції між ознаками-факторами, величина якого не повинна перевищувати 0,8 (із певним довірчим рівнем). Такий випадок зв'язків називають мультиколінеарністю, і включати у кореляційну модель у даному разі можна лише один із досліджуваних факторів.

Щодо оцінки вірогідності характеристик кореляції і регресії, необхідною складовою є обчислення ймовірних помилок та оцінка параметрів досліджуваної сукупності. Наукова методологія статистики розглядає оцінки кореляційно-регресійних залежностей із позиції трьох груп критеріїв: коефіцієнтів простої, множинної і часткової кореляції; критеріїв вірогідності параметрів рівнянь; помилок коефіцієнтів рівнянь регресії. Найбільш поширеними способами оцінки кореляційно-регресійних моделей є критерій

Стьюдента (*t*-критерій) і Фішера (*F*-критерій).

У кореляційно-регресійному аналізі не завжди враховується така методична особливість коефіцієнта парної кореляції (*r*) – залежність від обсягу чисельності одиниць сукупності. Тому для оцінки його надійності необхідно вивірити величину коефіцієнта на надійність. Даний параметр вважається істотним, якщо він перевищує (у модулі) середню квадратичну помилку втричі і більше $|r|:\sigma \geq 3$. Для малочисельних сукупностей істотність коефіцієнта кореляції оцінюється за

критерієм Стьюдента $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ (*n* – число парних спостережень).

Істотність коефіцієнта множинної кореляції (*R*) перевіряється також по *t*-критерію за наступним співвідношенням $t=R:S_R$. $S_R = \frac{1-R^2}{\sqrt{n-1}}$ (S_R –

середня квадратична помилка). Одержані розрахункові характеристики порівнюють із табличними їх значеннями за стандартними математичними таблицями (при відповідному рівні ймовірності).

У криволінійних зв'язках (параболічних, гіперболічних, експоненціальних та ін.) вірогідність кореляційного відношення (*η*) оцінюють за допомогою *F*-критерію:

$$F = \frac{\eta^2}{p-1} \div \frac{1-\eta^2}{n-p} \quad (P - \text{кількість параметрів рівнянь регресії}).$$

У практичних розрахунках здебільшого використовують прямолінійну функцію, хоча іноді (коли це потрібно) криволінійна функція точніше описує кореляційний зв'язок між досліджуваними ознаками. У наукових розробках з методологічної точки зору правомірним визнається підхід, якщо різниця у коефіцієнтах детермінації ($\eta^2 - r^2$) виявиться неістотною. Для оцінки вірогідності цих показників тісноти зв'язку пропонується критерій Стьюдента (за такого застосування його називають критерієм криволінійності):

$$t_{\eta^2, r^2} = \frac{\eta^2 - r^2}{\frac{2}{\sqrt{n}} \sqrt{(\eta^2 - r^2) - (\eta^2 - r^2)^2 (2 - \eta^2 - r^2)}}.$$

Зауважимо, коректність такого методологічного підходу в оцінці лінійності сумнівна, оскільки математична природа коефіцієнтів детермінації (η^2 і r^2) однакова – співвідношення факторної і загальної дисперсій.

Крім розглянутих вище параметричних критеріїв оцінки кореляційно-регресійних зв'язків існують і непараметричні критерії – це наближені оцінки тісноти зв'язку. Дослідники, на жаль, користуються (зазвичай) традиційними схемами обчислень, хоча потреба у використанні непараметричних критеріїв виникає, зокрема, коли має місце факт відсутності певних умов методологічного характеру, а саме: відповідність закону нормального розподілу; можливість кількісного виміру досліджуваних ознак; потреба у дослідженні атрибутивних ознак та ін.

До непараметричних критеріїв кореляційно-регресійних зв'язків належать: коефіцієнт кореляції рангів; критерій знаків (коефіцієнт Фехнера); коефіцієнт асоціації, коефіцієнт контингенції (схожості). Алгоритми щодо їх обчислень містяться у спеціальній літературі з математики і математичної статистики, а також у навчальних виданнях за нашим авторством. Існують й інші непараметричні критерії вивчення кореляційно-регресійних зв'язків, хоча вони дають досить наближені значення характеристик надійності тісноти зв'язку.

Для оцінки надійності як вихідної інформації дослідження, так і його результатів часто використовується математична характеристика – середнє квадратичне відхилення (σ). Дана статистична оцінка може виконувати самостійні і допоміжні функції. У першому випадку вона використовується для характеристики варіації (у натуральному вимірі ознаки), у другому – як основа для побудови інших статистичних характеристик: коефіцієнтів варіації; помилок репрезентативності характеристик розподілу; коефіцієнтів кореляції і регресії, структурних елементів дисперсійного аналізу; формул регресії. У разі практичного використання середнього квадратичного відхилення слід пам'ятати, що його величина залежить не тільки від ступеня варіації ознак, а й від абсолютних рівнів їх середнього розміру. Тому робити порівняння таких оцінок, розрахованих за варіаційними рядами з різномісними ознаками (як і з різними рівнями), безпосередньо не можна. Можливість такого порівняння забезпечує відношення середнього квадратичного відхилення до середнього рівня (*y* %) – коефіцієнт варіації ($V = \frac{\sigma}{x} 100$). Характеризуючи від-

носну міру варіації в рядах із різним рівнем середніх, дана статистична оцінка також зручна і для порівняння варіації різних явищ (наприклад, продуктивність праці та її оплата; прибутковість

підприємств і рівень їх спеціалізації тощо).

Коефіцієнт варіації є оцінкою надійності середнього розміру показника. При його значеннях до 5 % варіація вважається слабкою; 6–10 – помірною; 10–20 – значною; 21–50 – великою; понад 50 – дуже великою. Для невеликих сукупностей одиниць спостереження величина коефіцієнта варіації повинна бути не більше 33 % [1].

Зазначимо, розглянуті попередні питання критеріїв надійності стосувалися в основному величин параметрів досліджуваних сукупностей. Міркування тут велися виходячи з передбачень про однорідність сукупності та нормальність розподілу її одиниць, тобто, відповідність закону Гаусса-Лапласа. Це означало, що сукупності, які досліджуються, є типовими і різняться між собою лише за розмірами рівнів ознак. Проте окремі випадки вимагають перевірки гіпотези щодо характеру розподілу. Тобто, дослідник повинен вирішити такі завдання: визначити відповідність емпіричного розподілу тому чи іншому теоретичному виду розподілу – нормальному, біноміальному, поліноміальному і т.п.; визначити можливість належності двох і більше емпіричних розподілів до одного й того ж їх виду; з'ясувати наявність залежності в розподілі ознак однієї від іншої.

Перелічені вище завдання вирішуються шляхом застосування χ^2 -квадрат критерію (критерій згоди Пірсона- χ^2). Останній є об'єктивною оцінкою близькості фактичних розподілів до теоретичних. Підкреслимо, використовується даний критерій у випадках, коли необхідно встановити відповідність двох порівнюваних рядів розподілу – фактичного і теоретичного або двох фактичних.

Виходячи з того, що в дослідженні економічних явищ емпіричний розподіл, як правило, не відповідає нормальному, необхідно з'ясувати сильно чи слабо розходяться ці ряди. Йдеться про встановлення границі при недосягненні якої фактичний розподіл приймається за нормальний. Таку оцінку наближення емпіричного розподілу до нормального дає співвідношення фактичних і

теоретичних частот:
$$\chi^2 = \sum \frac{(n_i - n_t)^2}{n_t} \quad (n_i, n_t -$$

частоти емпіричного і теоретичного ряду). Порівнянням розрахованих значень критерію з табличними (за обраного рівня ймовірності) перевіряється статистична гіпотеза відповідності названих видів розподілу. Якщо обчислені параметри не перевищують теоретичні (граничні) значення, гіпотеза про відповідність нормальному закону розподілу приймається.

Зауважимо, розрахунок статистичних оцінок із використанням критерію згоди Пірсона дає змогу перевірити також належність кількох вибірових даних до однієї й тієї ж досліджуваної сукупності в разі з'ясування питання про однорідність вибірових даних. Розглянутий критерій може також бути використаний для розподілів, у яких розбіжності між очікуваними (теоретичними) і фактичними частотами не відповідають будь-якій очевидній структурі. Розраховують даний критерій і при з'ясуванні об'єктивності побудови ряду розподілу, що виключає можливість «підгонки» одиниць із метою відповідності їх одна одній.

Використовуючи розглянутий вище критерій слід пам'ятати про окремі обмеження (що не завжди враховується в дослідженнях), яких слід дотримуватися в його розрахунку. Важливо знати, що чисельність одиниць досліджуваної сукупності повинна бути не менше 50, а в кожній групі (інтервалі) мінімально допустима границя частот – 5 (в іншому разі їх об'єднують).

У практиці дослідницької роботи трапляються випадки, коли дослідник має справу з двома статистичними сукупностями з різною кількістю одиниць спостереження. У такому разі виникає необхідність перевірити статистичну гіпотезу про рівність двох дисперсій. Таку перевірку здійснюють шляхом розрахунку критерію Фішера (F). Якщо знайдена величина менше його табличного значення ($F_p < F_t$), відмінність між вибіровими дисперсіями вважається неістотною, а це свідчить про те, що досліджувані сукупності, з яких здійснено відбір, мають однакові статистичні характеристики.

В разі необхідності отримання оцінки істотності ряду дисперсій (більше двох) використовують інші критерії. У випадку за однакової чисельності вибірової сукупності розраховують критерій Кохрана, якщо вибірки неоднакові – критерій Бартлета. Перший із названих критеріїв являє собою відношення максимальної дисперсії (із порівнюваних) до суми всіх дисперсій:

$$K = \frac{\sigma_{\max}^2}{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}.$$
 Порівнянням розрахова-

ного критерію (K_p) із табличним його значенням (K_T) визначають надійність дисперсій. У випадку, якщо $K_p > K_T$, дисперсії вважаються неоднорідними, тобто їх відмінність істотна.

Для перевірки гіпотези про однорідність дисперсій більш потужним визнано критерій Бартлета (B). Теоретичною основою використання даного критерію є припущення про нормальність

розподілу ознак у досліджуваних сукупностях [3]. Суть даної математико-статистичної оцінки полягає в порівнянні середньої арифметичної (σ_a^2) і середньої геометричної (σ_z^2) дисперсій. Якщо дисперсії надійні, то середня арифметична і середня геометрична із дисперсій збігатимуться. У натуральній формі логарифмів даний критерій має вигляд:
$$B = \ln \frac{\sigma_a^2}{\sigma_z^2} \sum n_i .$$

Резюмуючи розгляд питань про критерії згоди, необхідно враховувати такі особливості використання їх в аналітичній роботі. По-перше, у разі порівняння емпіричного розподілу з тим чи іншим теоретичним завжди мається на увазі не вибіркова сукупність, а вся досліджувана. Вибірка тут характеризує досліджувану сукупність, а тому висновки про значимість чи невірогідність відмінностей у розподілах відносяться до всієї сукупності. По-друге, трактуючи поняття «значимість» і «невірогідність» потрібно пам'ятати, що відсутність значимих розбіжностей між емпіричним і теоретичним рядами розподілу ще не означає, що фактичний розподіл у досліджуваній сукупності в точності відповідає шуканому закону розподілу. Факт відсутності значимих розбіжностей дає можливість признати емпіричну сукупність як сукупність, розподілену за відповідним Законом, що не одне і те ж.

Якщо досліднику (експериментатору) невідома форма розподілу даних (що нерідко трапляється), з яких проводиться вибірка, використання розглянутих вище критеріїв може інколи призвести до суб'єктивної оцінки результатів спостереження. У зв'язку з цим математичною статистикою розроблені критерії оцінок вибірок із будь-якого виду розподілу. Ці критерії отримали назву «непараметричні». До таких критеріїв відносяться:

БІБЛІОГРАФІЯ

1. *Закс Л.* Статистическое оценивание : [монография] / М. : Статистика, 1976. – С. 81.
2. *Опря А. Т.* Статистика : [підруч.] / К. : Урожай, 2006. – 448 с.
3. О теории вероятностей и математической статистике (переписки А. А. Маркова и А. А. Чупрова) : [монография]. – М. : Наука, 1977. – 197 с.
4. *Попович И. В.* Методика экономических исследований в сельском хозяйстве : [монография] / М. : Экономика, 1977. – С. 4–8.

критерій Колмогорова (λ – лямбда), критерій Уайта, критерій Уілкоксона. Найбільш простий за своєю конструкцією критерій Колмогорова розраховується як відношення максимальної різниці (у модулі) нагромаджених частот емпіричного і фактичного рядів розподілу до квадратного кореня із чисельності вибірки:
$$\lambda = \frac{\max |n_i - n_t|}{\sqrt{n}} .$$

Чим більша розбіжність у рядах, тим більша величина λ -критерію. За повного збігу порівнюваних частот у рядах розподілу $\lambda=0$. Використовуючи його в дослідженні потрібно дотримуватися умови – достатнє число одиниць спостереження. До оцінки нечисленних вибірок критерій Колмогорова неприйнятний [1].

Висновок. Недостатня концентрація зусиль науковців у напрямі широкого використання статистичних і математико-статистичних прийомів у практику економічних розрахунків і неусвідомленість того, що регулюючі функції управління виробництвом потрібно розглядати в комплексі економічних і статистичних проблем, зумовили необхідність розглянути питання системного підходу в дослідженні економічних явищ і процесів у форматі поєднання емпіричного аналізу, теорії і методології статистики.

У методологічному плані механізм економічного дослідження причинно-наслідкових зв'язків повинен базуватися на наукових концепціях статистики з урахуванням економічної природи досліджуваних явищ і статистичної (математико-статистичної) природи методів, показників і критеріїв надійності інформаційного поля. Саме такий підхід зможе вирішити завдання поглибленого дослідження конструктивно складних економічних моделей причинно-наслідкових зв'язків.

5. *Рябушкин Т. В.* Теоретические концепции в отечественной статистике : [монография] / Т. В. Рябушкин, В. М. Симчера, Е. А. Машихин / М. : Наука, 1986. – 311 с.
6. *Суслов И. П.* Теория статистических показателей : [монография] / М. : Статистика, 1995. – С. 3–5.
7. *Философский словарь* / Под ред. М. А. Королева : [монография]. – М. : Издательство политической литературы, 1987. – С. 278.