

З. П. Бараник, д-р екон. наук, проф.,
заступник завідувача кафедрою статистики,
С. С. Ващаєв, канд. екон. наук, доц.,
заступник декана ФІСіТ
В. В. Майба, канд. екон. наук,
ст. викл. кафедри статистики,
ДВНЗ «КНЕУ імені Вадима Гетьмана»

ПРИКЛАДНА СТАТИСТИКА: ІСТОРІЯ СТАНОВЛЕННЯ ТА ПЕРСПЕКТИВИ РОЗВИТКУ

АНОТАЦІЯ. Визначено основні етапи становлення статистичної науки, розкрито ключові положення різних наукових шкіл та їх вклад у розвиток статистичного методу. Обґрунтовано необхідність на визначено напрямки розвитку статистики з урахуванням сучасних потреб науки і практики управління соціально-економічними явищами та процесами.

АННОТАЦИЯ. Определено основные этапы становления статистической науки, раскрыто ключевые принципы разных научных школ и их вклад в развитие метода статистики. Обосновано необходимость и определены направления развития статистики с учетом современных потребностей науки и практики управления социально-экономическими явлениями и процессами.

ANNOTATION. The basic stages of development of statistics science was certain; key positions of different scientific schools and their contribution in to development of statistical method was exposed. Proved the expedience and offered the ways of development of statistics method according to the modern necessities of science and management.

КЛЮЧОВІ СЛОВА. Описова статистика, політична арифметика, закономірність, стохастичність, нормальний розподіл, непараметрична статистика, нечислові дані, багатовимірний аналіз, розпізнавання образів.

Вступ

Виникнення будь-якої науки насамперед пов'язане з відповідними практичними потребами людей та необхідністю вирішення тих проблем, які виникають у процесі їх життєдіяльності. Становлення статистичної науки було зумовлене потребою у певних відомостях, об'єктивній інформації про явища та процеси соціально-економічного життя, складність та швидкість перебігу яких неупинно зростали. На початковому етапі свого становлення статистика передбачала просте накопичення первинних даних, їх упорядкування та система-

тизацію. Однак подальший розвиток суспільства вимагав більш потужних засобів пізнання, здатних пояснити сутність тих явищ і процесів, які оточували людину, зрозуміти механізм їх формування та закономірності розвитку. Статистика поступово розширювала свій інструментарій і перетворювалася на самостійний, науково обґрунтований метод пізнання. Еволюція суспільних відносин та їх базису зокрема, представленого способом виробництва, зумовила необхідність формування відповідного рівня надбудови, здатної ефективно виконувати управлінські функції та забезпечувати розвиток суспільства. При цьому статистика розглядалася як засіб виявлення невідомого, спосіб передбачення деяких подій за заданих умов, і використовувалася як універсальний метод, інструмент для обробки результатів будь-яких спостережень.

Постановка задачі. Сьогодні статистичні методи успішно застосовуються в різних сферах економічного життя та практично в усіх галузях науки. Практична значущість статистики для економіки, менеджменту, технічних досліджень, медицини і психології актуалізує необхідність проведення подальших досліджень у напрямі розвитку її методології [1—10]. Однак, оцінка значущості вже отриманих здобутків та визначення на цій основі перспективних напрямів розвитку наукового методу статистики неможливі без осмислення історичного шляху її становлення, визнання здобутків та виявлення невирішених завдань. Адже, як справедливо зауважив видатний учений-хімік Л. О. Чугаєв, «знання нашого минулого в кожній галузі науки є не лише нашою природною національною потребою, а й має величезне виховне значення для майбутніх діячів науки» [20]. Вивчення історичних шляхів, особливостей та суперечностей у становленні, розвитку і застосуванні статистичної науки дає можливість не лише пізнати, що досягнуто в минулому, а й дізнатися, за яких умов і завдяки чому воно досягнуто і який вплив здійснює на подальший розвиток науки.

Отже, мета даного дослідження полягає у розкритті процесу формування статистичних ідей, понять, теоретичних концепцій, методів, які зароджувалися і розвивалися у складній взаємодії практичної діяльності людей і логіки власного розвитку самої статистичної науки. Це дозволить визначити перспективні напрями розвитку статистичної методології, з урахуванням сучасних потреб управління.

Результати дослідження

Описова школа статистики. Практичні потреби завжди виконували роль могутнього каталізатора розвитку будь-якої науки, зокрема, й статистики. Так, із розвитком державності, практика управління господарською системою країни потребувала простої реєстрації фактів та обліку даних. Ці потреби сприяли зародженню описової школи статистики, основним завданням якої був систематизований опис стану держави за географічними, правовими та економічними характеристиками. Основи описової школи були закладені в працях німецького професора медицини і політики Германа Конрінга, який започаткував нову академічну дисципліну, що на німецькій мові отримала назву «*Staatenkunde*» (державознавство). Пізнання держави за Г. Конрінгом передбачало з'ясування чотирьох причин:

1) причина матеріальна — тобто країна і люди, або територія і населення;

2) причина кінцева, тобто наскільки жителі країни живуть щасливим життям (рівень забезпеченості матеріальними та нематеріальними благами);

3) причина формальна — форма правління та устрій;

4) причина діюча — з одного боку це керівні постаті (правителі) — головна причина, а з іншого — підлегла причина, до якої належать головним чином гроші (фінансові ресурси країни), військо та флот [20].

У міру розширення кількості описуваних держав та кола соціально-економічних явищ, які ставали об'єктом вивчення статистики, значного збільшення обсягів статистичного матеріалу, зокрема, кількісного (числового) характеру, у рамках описової школи стали виникати нові напрями і методи статистичної обробки даних: таблична статистика (Й. П. Анхерсен, Ф. Д. Лівшиць, І. К. Кирилов), лінійна статистика або графічний метод (В. Плейфейр, А. Кроне, А. Геррі, А. Кетле), порівняльна статистика (А. Ф. Бюшінг) та ін. [11, 16].

Школа політичних арифметиків. Паралельно з описовою школою на початку XVII ст. в Англії зародився та розвивався інший напрям статистичної науки — школа політичних арифметиків. Погляди політичних арифметиків докорінно відрізнялися від поглядів представників описової школи тим, що головним завданням нового на той час наукового методу (статистики) вважали виявлення на основі масових спостережень закономірностей і взаємозв'язків у соціально-економічних явищах. Термін «полі-

тична арифметика» в науковий обіг ввів один із засновників даної школи, відомий класик політекономії Уільям Петті. Характеризуючи позитивні сторони даного напрямку він вказував на можливості вираження думок мовою чисел, ваг і мір замість того, щоб використовувати слова лише в порівняльному і найвищому ступені та вдаватись до абстрактних аргументів [17].

Першою науковою працею, з якої бере початок школа політичних арифметиків і з часу видання якої починається історія сучасної статистичної науки, була книга Джона Граунта «Природні й політичні спостереження, згадані у додатковому переліку й зроблені на основі бюлетенів смертності Джоном Граунтом, громадянином Лондона, що стосуються управління релігії, торгівлі, зростання, повітря, хвороб та різних змін названого міста», яка вийшла у світ 1662 р. Тим самим Дж. Граунт зробив перші спроби об'єктивного, наукового вивчення природного руху населення, зокрема смертності, заклавши фундамент особливої складової статистичної науки — демографії. У. Петті не тільки широко застосовував у своїх працях чисельний метод, а й визначив основним завданням статистики дослідження закономірностей основних соціально-економічних явищ свого часу. Він перший з-поміж багатьох учених, хто у дослідженнях широко використовував середні величини [17, 20].

Отже, в центрі уваги як представників описової школи, так і політичних арифметиків було населення країн та матеріально-речові елементи національного багатства — земля, її надра, армія, флот, особисте майно. Багато в чому статистика завдячує насамперед Людині, як об'єкту вивчення. Статистика виявила, що узагальнення масових явищ дозволяє не лише пояснювати їх поведінку, але й прогнозувати їх. Однією з таких спроб був розрахунок середньої майбутньої тривалості життя, відповіді на безліч питань типу «що необхідно для того, щоб...». Первинне вивчення людини велося з позицій демографії та економіки.

Важливий вклад у розвиток статистики зробив К. Гаус, який у 1794 р. розробив метод найменших квадратів (МНК), один з найпопулярніших і до цього часу статистичних методів, і застосував його при розрахунку орбіти астероїда Церера — для оцінки помилок астрономічних спостережень. Метод найменших квадратів, завдяки широкій сфері застосування, посідає виняткове місце серед методів математичної статистики. Задачею МНК є оцінка закономірностей, які спостерігаються на тлі випадкових коливань, та її використання для подальших розрахунків, зокрема, для прогнозів [16].

Становлення стохастичної статистики. Розвиток статистичної науки у ХІХ ст. насамперед пов'язаний з ім'ям ученого, який зробив найважливіший крок на шляху до становлення статистики як самостійної науки. Ним був бельгієць Адольф Жак Кетле, вклад якого у становлення та розвиток статистичної науки був настільки значний, що набув багатьох прихильників та сприяв формуванню самостійної статистичної школи [14].

У своїх роботах А. Кетле застосовував до політичних і моральних наук метод, який ґрунтується на спостереженні й обчисленні і який використовується природничими науками. Для цього він створює особливу науку — соціальну фізику, змістом якої є встановлення закономірностей суспільного життя за допомогою статистичних методів, що ґрунтуються на теорії ймовірностей. Ю. Янсон говорив, що соціальна фізика є першою спробою побудувати статистику на міцних індуктивних і наукових основах. Адже спостереження та реєстрація фактів ще не є наукою. Кожна наука прагне не констатувати факт, але пояснити причини його виникнення та зміни. Успіх наукових праць А. Кетле пояснюється тим, що йому вдалося на основі масових фактичних даних довести наявність закономірностей у суспільних явищах, при чому на прикладах найгостріших для тогочасного суспільства проблем (народжуваності та смертності). Кетле також розробляв статистичні методи управління якістю, сертифікації і класифікації продукції, які є актуальними і сьогодні.

Хоча при визначенні сутності статистики А. Кетле значною мірою наслідував своїх попередників — представників описової школи та політичних арифметиків, однак стосовно теорії статистичної науки він мав власні погляди, які є основоположними і до сьогодні. Зокрема, завдання статистики А. Кетле вбачав у оцінюванні зібраних нею даних та у висновках, які з них можна зробити. Він вказував на необхідність розумного використання статистичних даних — отримання на їх основі корисних висновків про усталеність того чи того соціального ладу, з'ясування його внутрішньої будови. Саме А. Кетле вперше чітко і зрозуміло виразив думку, що предметом статистики є масові явища, тобто явища, в яких окремий елемент є індивідуальним у тому значенні, що поруч з деякими загальними для всієї конкретної сукупності властивостей спостерігаються відмінності між ними, зумовлені дією другорядних причин [20].

А. Кетле вважав, що можна виміряти будь-яку ознаку людини, при цьому виокремити в ній детерміновану компоненту та випад-

кову (індивідуальну) складову: $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, де y_{ijk} — значення ознаки y для k -ї одиниці, яка входить до i -ї групи за однією ознакою та до j -ї за іншою; μ_{ij} — характеристика типу (середня величина). Багато уваги приділяв А. Кетле статистичному вивченню населення, зокрема: сприяв організації та проведенню переписів, вивчав структуру та динаміку чисельності населення. Він розробив концепцію зміни чисельності населення згідно з рівнянням логістичної кривої [14]:

$$P(t) = \frac{K}{1 + e^{a-rt}},$$

де $P(t)$ — чисельність населення в момент часу t ; e — основа натурального логарифму; K, a, r — параметри формули логістичної кривої.

Попри те, що вчення А. Кетле про закономірності суспільного життя здобуло широке визнання, наприкінці 70-х років XIX ст. його праці піддавалися значній критиці. Причиною стало те, що послідовники А. Кетле надавали метафізичного забарвлення його вченню (А. Вагнер). При цьому центром критики була сталість статистичних чисел і пов'язана з нею проблема свободи статистичних висновків. У зв'язку з цим деякі положення теорії А. Кетле визнавалися ненауковими і була відкинута можливість застосування математики, й насамперед теорії ймовірності у статистичних дослідженнях. У результаті цього майже до кінця XIX століття увага статистиків концентрувалась виключно на організаційних і технічних питаннях статистичного спостереження масових явищ.

Відродження наукових досліджень теорії статичної науки із широким застосуванням теорії ймовірностей розпочалося з появою праць видатного німецького вченого Вільгельма Лексіса та його послідовників. Учення В. Лексіса про сталість статистичних чисел увійшло в історію статистичної науки під назвою «теорія дисперсії», яка викликала бурхливі дискусії та мала значний вплив на розвиток статистики у XX ст. В. Лексіс спростовував погляди А. Кетле щодо сталості статистичних чисел і можливості описати фактичні розподіли за допомогою кривої нормального розподілу. Він висунув положення, що відхилення від середньої можуть мати або характер кривої нормального розподілу, або бути вище чи нижче. У своїх працях він продовжує дослідження А. Кетле щодо застосування теорії ймовірностей до вивчення на-

селення, зокрема смертності. В. Лексіс зробив першу спробу класифікації відносних величин. Однак запропоновану ним класифікацію не можна визнати цілком задовільною, адже вона досить неоднозначна і нетипова.

В. Лексіс був першим з тих, хто зробив спробу оцінити однорідність статистичних даних, вивчаючи сталість динамічних рядів. Він запропонував за характером зміни рівнів статистичних рядів розрізняти такі їх основні типи: еволютарні — у яких спостерігається тенденція до зміни рівнів у певному напрямі; ундуляторні — мають нерегулярний, хвилеподібний характер зміни рівнів; періодичні — рівні таких рядів мають правильне регулярне піднесення і спади; осциляторні — ряди, які характеризуються безладними коливаннями рівнів.

Запропонована В. Лексісом теорія однорідності даних базувалася на судженні — якщо різниця між рівнями ряду випадкова, то в його основі лежить деяке стале число — імовірність, яка базується на багаторічних даних. Її оцінкою слугує середня частка (p), а дисперсія частки визначається як $r^2 = pq/n$, де $q = 1 - p$, а n — середньорічна кількість спостережень. Якщо ряд стабільний, то дисперсія частки по роках має бути рівною r^2 [16]:

$$R^2 = \frac{\sum (p_i - p)^2}{(m-1)},$$

де p_i — частка i -го року; m — кількість рівнів (років) ряду динаміки.

За міру стійкості слугувало співвідношення цих двох дисперсій, яке отримало назву — коефіцієнт дивергенції або число Лексіса:

$$Q^2 = \frac{R^2}{r^2}.$$

Слід зазначити, що теорія сталості статистичних рядів, запропонована В. Лексісом, мала певні вади, зокрема, не було чітко визначено формулу розрахунку коефіцієнта дисперсії, не зроблено узагальнення щодо схем сталості згідно з теоремами Я. Бернуллі, С. Пуассона, незрозумілою залишалася й сама наукова концепція коефіцієнта дисперсії. Спроби дати відповіді на ці та інші питання зроблені його послідовниками, зокрема В. Борткевичем та О. Чупровим.

Саме Владислав Борткевич був найпомітнішим послідовником В. Лексіса. Він був досить різнобічним ученим: у своїх працях

приділяв увагу теоретичним і прикладним питанням теорії ймовірностей та теорії дисперсії, вибіркового та індексного методів, застосування статистичних методів у дослідженнях населення, страхування, доходів тощо. В. Борткевич сформулював і відкрив «закон малих чисел» для випадків рідкісних явищ (частота яких близька до нуля), який увійшов у теорію статистики і впродовж кількох десятиріч здобув загальне визнання. Він стверджував, що чим рідше відбувається досліджуване явище, тим краще статистичні висновки узгоджуються з відповідними математичними функціями [20].

В. Борткевич також зробив значний внесок у розвиток теорії індексного методу. Він розглядав індекси не лише як узагальнюючі показники динаміки, а й як засіб вивчення впливу факторів, тим самим наголошуючи, що індекси окрім синтетичної функції виконують ще й аналітичну.

Тож можна стверджувати, що ХІХ століття — час народження сучасної статистики, яка стала не просто кількісними даними, прийомами їх систематизації, опису і аналізу, вирішення конкретного (ситуативного, одиничного, випадкового) завдання, а й посіла своє власне місце серед інших суспільних наук. Статистика породжувала саму себе, з'явилася метанаука (наука про науку). Почали формуватися загальні уявлення про предмет статистики — статистичну закономірність, про об'єкт — статистичну сукупність, про одиницю сукупності, її ознаки. Ці уявлення склали категорійний апарат теорії статистики. Все те, що так довго складалося, нарешті, було названо своїми іменами, і від цього темпи розвитку статистичної науки різко зросли. Розвиток статистичної методології стимулював вихід статистики за рамки соціально-економічних явищ (Гальтон, Пірсон, Фішер) [11, 16, 17, 20].

Наступним кроком у розвитку теоретичних основ статистики стало з'ясування її взаємин з іншими науками, серед яких найближчою була визнана теорія ймовірності. Пізнання соціально-економічних явищ обумовлювало зв'язок статистики з політекономією та іншими суспільними науками. Минув той період, який стримував її розвиток з причин невизначеності щодо суті статистики — чим вона є: наукою з власним предметом дослідження або тільки методом пізнання, інструментарієм інших наук. Вирішенню цієї дилеми сприяв дуалістичний погляд на статистику, який послідовно розвивав А. Чупров. Він відносив статистику як до номографічних наук, що вивчають загальні закономірності розвитку масових явищ, так і до ідеографічних наук, що пояснюють закономірності конкретних явищ у певних умовах місця і

часу. Статистичний метод то повністю ототожнювався з індуктивним, то оголошувався особливим, самостійним.

К. Джині відзначав, що «статистична індукція є окремим випадком узагальнювальної неповної індукції, в якій відмінність між вивченими подіями і подіями невивченими (а отже, між вивченими подіями і комплексом подій) залежить від випадку». Дедуктивно-гіпотетичний метод був покладений в основу перевірки статистичних гіпотез, проведення типологічних групувань [16].

Поєднання індукції і дедукції лежить в основі стохастичної теорії статистики, яка обережно і послідовно порівнює апріорні (ті, що передують досліду) дані теорії імовірності та апостеріорні (емпіричні) статистичні характеристики. Будь-яке емпіричне дослідження розглядається як вибірка, результати якої носять випадковий характер, а кінцева мета — отримання надійних, невивпадкових величин. Вихідними поняттями стохастичної теорії статистики є імовірність, випадкова величина, стохастичний зв'язок між випадковими змінними. Поняття стохастичного зв'язку передбачає, що дві змінні знаходяться в стохастичній залежності за умови, коли умовні закони розподілу значень результативної ознаки змінюються з переходом від одних значень до інших пояснюючої змінної. Це поняття лягло в основу теорії кореляції. З'явилася можливість розглядати функціональну залежність як граничний випадок стохастичного зв'язку зі зникаюче малим розсіюванням. Стохастична теорія статистики склала ідейну основу математичної статистики — теорії перевірки гіпотез і теорії оцінювання апріорних величин за емпіричними даними.

У межах цього напряму особливу роль у розвитку статистичної науки відіграли праці та наукові погляди видатного українського вченого-статистика, математика, теоретика економічної науки Євгена Євгеновича Слуцького. Будучи випускником Київського комерційного інституту (тепер ДВНЗ «Київський національний університет імені Вадима Гетьмана») значну частину свого творчого життя Є. Слуцький проводить саме у його стінах, зокрема у 1920-х роках завідує кафедрою статистики [17, 18, 20].

Першим вагомим науковим дослідженням Є. Слуцького є його дипломна робота на тему «Теорія граничної корисності», виконана, як писав пізніше сам Слуцький, «під кутом зору математика, який вивчає політичну економію» [17]. У концепції дослідження наголошується на необхідності тісного взаємозв'язку між економічним та психологічним у поясненні корисності. Саме ж дослідження містить глибокий системний аналіз розробок граничної корисності К. Менгера, Ф. Візера, Л. Вальраса, В. Парето, Е. Бем-Баверка та ін. Не відкидаючи цілком суб'єктивного трактування корисності західними

теоретиками, він вніс елементи об'єктивного аналізу зазначеної проблеми через порівняння відносних рівнів цінностей різних благ та виявлення залежності зміни функції корисності від руху цін та грошових доходів населення. Серцевиною теорії корисності стало відоме й до сьогодні рівняння Слуцького:

$$\frac{dx^*}{dp_n} = \left(\frac{dx^*}{dp_n} \right) - \frac{dx^*}{dM} x_n^*,$$

де dx^* — зміна попиту на товари; dp_n — зміна ціни на n -й товар; dM — зміна грошових доходів споживачів.

Дане рівняння, яке характеризує кількісну залежність між зміною цін на окремі товари та зміною доходів споживачів, з одного боку, і структурою споживацького попиту — з іншого, отримало загальне визнання й увійшло до теорії рівноваги споживача.

Значний вклад зробив Є. Слуцький у вивчення випадкових процесів — йому вперше вдалося відкрити псевдоперіодичні хвилі, що виникають у рядах динаміки, рівні яких корельовані та встановити похибку коефіцієнта кореляції між такими рядами. Ці положення викладено у фундаментальній науковій праці під назвою «Теорія кореляції й елементи вчення про криві розподілу», яка вийшла у світ 1912 р. Доповідаючи її основні наукові результати на 34-му засіданні Товариства економістів при Київському комерційному інституті Є. Слуцький акцентує увагу на таких моментах [18]:

по-перше, метод кореляції остаточно перетворює статистику на точну науку, так як дає: а) точну математичну міру залежності статистичних величин; б) імовірнісну оцінку похибки цієї міри, що вказує на надійність або ненадійність висновків, які можуть бути зроблені на основі даного матеріалу; в) форму залежності, що дає змогу за однією величиною знайти імовірне значення іншої; г) границі більш або менш імовірних коливань цієї останньої;

по-друге, теорія кореляції багатьох змінних є абсолютно незмінним засобом вивчення залежності між кількома величинами;

по-третє, теорія кореляції має бути введеною в коло обов'язкових предметів освіти, а статистика і метод кореляції — в практику науково-статистичних закладів.

Зазначимо, що погляди Є. Слуцького дещо випереджали час, унаслідок чого їх масове практичне застосування в соціально-економічних дослідженнях відбулося вже наприкінці ХХ ст., коли з'явилася можливість використання потужної обчислювальної техніки.

Параметрична та непараметрична статистика. В процесі еволюції ринкових економічних відносин все повніше розкривалася роль статистики на всіх рівнях економічної діяльності — підприємства, регіону, галузі, національної економіки тощо. Статистичне моделювання і прогнозування, підхід до статистики як методу ухвалення рішення в умовах невизначеності посилювали зацікавленість статистичною наукою та підвищували її популярність.

Початком сучасного етапу розвитку прикладної статистичної науки можна вважати 1900 р., коли англієць К. Пірсон заснував журнал «*Biometrika*». Перша третина ХХ ст. пройшла під знаком параметричної статистики. Вивчалися методи, засновані на аналізі даних з параметричних сімейств розподілів, які описуються кривими сімейства Пірсона. Найпопулярнішим був нормальний (гаусівський) розподіл. Для перевірки гіпотез використовувалися критерії Пірсона, Ст'юдента, Фішера. Були запропоновані метод максимальної правдоподібності, дисперсійний аналіз, сформульовані основні ідеї планування експерименту. Так, розроблену в першій третині ХХ ст. статистичну теорію прийнято називати параметричною статистикою, оскільки об'єктом її вивчення були вибірки з розподілів, які описуються одним або невеликою кількістю параметрів. Найзагальнішим є сімейство кривих Пірсона, які задаються чотирма параметрами. Як правило, досить складно визначити ті характеристики, за якими конкретний розподіл результатів спостереження повинен входити в те або те параметричне сімейство кривих. При цьому добре відомі виключення: якщо імовірна модель передбачає підсумовування незалежних випадкових величин, то суму доречно описувати нормальним розподілом; якщо ж у моделі розглядається добуток таких величин, то підсумок описується логарифмічно нормальним розподілом. Однак у більшості реальних випадків подібних моделей немає, і наближення реального розподілу за допомогою кривих з сімейства Пірсона — чисто формальний прийом [16, 20].

Саме з таких міркувань академік С. Бернштейн критикував параметричну статистику. Зауважуючи, що ця теорія залишається основою викладання статистичних методів і продовжує використовуватися переважною кількістю дослідників, які не бажають вникати та розвивати нові її напрями.

Так, у першій третині ХХ ст., одночасно з параметричною статистикою, у роботах Спірмена і Кендалла з'явилися перші непараметричні методи. Для оцінки взаємозв'язку між ознаками порядкової (рангової) шкали К. Спірмен запропонував коефіцієнт рангової кореляції, який обчислюється за формулою [13]:

$$\rho = 1 - \frac{6 \sum d_j^2}{n(n^2 - 1)},$$

де $d_j = Rx_j - Ry_j$ — відхилення між рангами ознак x та y ; n — число рангів.

Однак, з огляду на те, що рангова кореляція не потребує додержання будь-яких математичних передумов щодо розподілу ознак, зокрема вимоги нормальності розподілу, рангові оцінки щільності доцільно використовувати для сукупностей невеликого обсягу.

Попри значну популярність у наукових дослідженнях, непараметрична статистика, яка не робить нереалістичних припущень щодо належності результатів спостереження будь-яким параметричним функціям розподілу, стала помітним методом прикладної статистики лише з другої третини ХХ ст. У 1930-х роках з'явилися роботи А. Колмогорова і Н. Смірнова, в яких пропонувалися та вивчалися статистичні критерії, названі на їх честь. Ці критерії засновані на використанні так званого емпіричного процесу — різниці між емпіричною і теоретичною функціями розподілу, помноженої на корінь квадратний з обсягу вибірки. Зокрема, в роботі А. Колмогорова (1933 р.) досліджено граничний розподіл супремума модуля емпіричного процесу, який має назву критерій Колмогорова. Н. Смірнов досліджував супремум і інфімум емпіричного процесу, а також інтеграл (за теоретичною функцією розподілу) квадрата емпіричного процесу.

У другій половині ХХ ст. розвиток непараметричної статистики відбувався швидкими темпами [12, 13]. Велику роль зіграли роботи Ф. Вілкоксона і його школи. Запропонований ними критерій застосовується для зіставлення показників, отриманих за однією вибіркою, але у різних умовах. Він дозволяє встановити не лише спрямованість змін, але і їх вираженість. За його допомогою визначають, чи є зрушення показників у якомусь одному напрямі більш інтенсивними, аніж в іншому. Розраховується критерій Вілкоксона за формулою:

$$z = \frac{R - \mu_R}{\sigma_R},$$

$$\text{де } \mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}, \quad \sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}.$$

За допомогою непараметричних методів можна вирішувати практично те ж коло статистичних завдань, що і за допомогою

параметричних методів. Проте для забезпечення широкого впровадження непараметричних методів необхідно провести ще цілий комплекс теоретичних і пілотних прикладних робіт. Значну роль відіграють непараметричні оцінки щільності, непараметричні методи регресії і розпізнавання образів. У вітчизняній статистиці непараметричні методи набули значної популярності після виходу в 1965 р. першого видання збірки статистичних таблиць Л. Большова та Н. Смірнова, в яких представлено значення основних непараметричних критеріїв.

Проте параметричні методи все ще більш популярніші ніж непараметричні, особливо серед тих дослідників, які мало знайомі зі статистичними методами. Значна кількість проведених експериментів засвідчує значну відмінність фактичних розподілів випадкових величин від нормальних розподілів, зокрема гаусівського. Проте теоретики продовжують будувати і вивчати статистичні моделі, засновані на «нормальності», а практики — застосовувати подібні методи і моделі. Іншими словами, «шукають під ліхтарем, а не там, де загубили».

Статистика нечислових даних. Сучасну прикладну статистику, залежно від особливостей об'єкта дослідження та використовуваних методів, можна поділити на такі чотири предметні області:

- 1) статистика числових даних (випадкових величин);
- 2) статистика часових рядів і випадкових процесів;
- 3) статистика об'єктів нечислової природи (непараметрична статистика);
- 4) багатовимірний статистичний аналіз.

Перші дві області наукових досліджень є класичними, вони були добре відомі ще в першій половині ХХ ст. Для сучасної науки та практики управління складними динамічними системами найцікавішими видаються дві останні. Так, статистика нечислових даних, або її ще іменують нечисловою статистикою, за оцінками багатьох учених стане центральною областю прикладної статистики у ХХІ ст., оскільки містить найзагальніші підходи і результати.

У чому ж її відмінність та позитивні риси? Початковим об'єктом статистичного дослідження є статистична сукупність, у математичній статистиці — це вибірка. У ймовірнісній теорії статистики вибірка — це сукупність незалежних, однаково розподілених випадкових елементів. У класичній статистиці елементи вибірки — це числа. У багатовимірному статистичному аналізі — вектори. А в нечисловій статистиці елементи вибірки — це

об'єкти нечислової природи, які не можна додавати та множити на числа. Іншими словами, об'єкти нечислової природи лежать на площинах, які не мають векторної структури.

Для класичної прикладної статистики характерною є дія додавання: при розрахунку вибірових характеристик розподілу (вибіркоче середнє арифметичне, вибіркова дисперсія і ін.), у регресійному аналізі і інших областях цієї наукової дисципліни постійно використовуються суми. Математичний апарат — закони великих чисел, центральна гранична теорема й інші теореми — спрямовані на вивчення сум. У нечисловій же статистиці не можна використовувати операцію додавання, оскільки елементи вибірки лежать у просторах, де додавання неможливе. Методи обробки нечислових даних засновані на принципово іншому математичному апараті — на застосуванні різних відстаней у просторах об'єктів нечислової природи.

Прикладами об'єктів нечислової природи є:

- значення якісних ознак, тобто результати кодування об'єктів за допомогою заданого переліку категорій (градацій, рангів);

- впорядкування (ранжування) експертами об'єктів сукупності — підприємств, регіонів, країн за рівнем їх ефективності, конкурентоспроможності та ін.;

- класифікація, тобто розподіл на групи (кластери) схожих між собою об'єктів;

- результати парних порівнянь або контролю якості продукції за альтернативною ознакою («якісна» — «бракована»), тобто послідовності з 0 і 1;

- множини (звичайні або нечіткі), складені експертами незалежно один від одного;

- словесна інформація, пропозиції, текстові дані;

- вектори, координати яких представлені сукупністю різнотипних ознак;

- відповіді на запитання експертної, маркетингової або соціологічної анкети, одна частина з яких носить кількісний характер (можливо, інтервальний), інша — зводиться до вибору однієї з кількох позицій, а решта — текстова інформація тощо.

Ще на початку 1970-х років під впливом прикладних досліджень у технічних, медичних і соціально-економічних науках активно розвивається статистика об'єктів нечислової природи, відома також як статистика нечислових даних або нечислова статистика. Провідна роль у формуванні цієї порівняно нової області прикладної статистики належить російським ученим. Велику роль відіграв заснований в 1973 р. науковий семінар «Експер-

тні оцінки і аналіз даних». Наукова спільнота почала цікавитися методами експертних оцінок ще у 1960-х роках. Як наслідок, почалося знайомство з конкретними математизованими теоріями, пов'язаними з цими методами [15]

На основі теорії експертних оцінок (у фінансах, соціології, психології, техніці, медицині) розвивалися конкретні напрями статистики об'єктів нечислової природи. Було встановлено зв'язки між конкретними видами таких об'єктів, розроблено для них імовірнісні моделі. На сьогодні статистика об'єктів нечислової природи з теоретичної точки зору є досить добре розвинутою: основні ідеї, підходи і методи розроблені і вивчені математично, доведено багато теорем. Проте вона залишається недостатньо апробованою на практиці. Це пов'язано як з її порівняно молодим віком, а також із загальновідомими особливостями організації науки наприкінці ХХ ст., коли були відсутні достатні стимули для того, щоб теорія активно впроваджувала результати своїх досліджень у практику.

Слід зазначити, що в статистиці об'єктів нечислової природи, як і в інших областях прикладної статистики і прикладної математики взагалі, одна і та ж математична схема може з успіхом застосовуватися і в технічних дослідженнях, і в менеджменті, і в економіці, і в геології, і в медицині, і в соціології, і для аналізу експертних оцінок, і в багатьох інших областях, а тому її краще всього формулювати і вивчати в найзагальнішому вигляді, для об'єктів довільної природи.

Багатовимірний статистичний аналіз. Явища та процеси соціально-економічного життя, як правило, характеризуються достатньо великою кількістю параметрів, різних як за своїм змістом, так і статистичною природою. Прикладом чого можуть бути дослідження, в яких оцінюється: економічний розвиток держав (рівень конкурентоспроможності), який характеризується як макроекономічними показниками (ВВП, ВНП, НД, індекс цін, платіжний баланс тощо), показниками забезпеченості ресурсами (природними, людськими, інтелектуальними та ін.), так і рівнем розвитку соціальної сфери; фінансовий стан суб'єктів господарювання (рівень прибутковості, дохідності, ліквідності, надійності, платоспроможності та ін.); рівень інвестиційної привабливості країн, регіонів, галузі. При цьому вихідні статистичні дані формують багатовимірні вектори, а вивчення взаємозв'язків між їх компонентами потребує відповідного інструментарію [2, 13, 19, 21, 22].

Таким інструментарієм є багатовимірний статистичний аналіз, який включає себе методи збору і обробки багатовимірних стати-

стичних даних, їх систематизації і обробки з метою виявлення характеру і структури взаємозв'язків між компонентами досліджуваного багатовимірного вектора (характеристики). За допомогою методів багатовимірного статистичного аналізу можна вирішувати наступні завдання:

— визначати взаємозв'язок між компонентами (показниками) багатовимірного вектора;

— оцінювати вплив показників багатовимірного вектора (факторів) на досліджувану результативну ознаку;

— здійснювати багатовимірну класифікацію одиниць сукупності за множиною ознак.

Перші два завдання вирішуються за допомогою *методів зниження розмірності* матриці вихідної інформації шляхом згортки її стовпців — переходу від вихідних ознак до їх лінійної комбінації. Такими методами є *метод компонентного і факторного аналізу*, за допомогою яких можна виявляти об'єктивно існуючі, але безпосередньо не спостережувані закономірності. Їх застосування є доречним насамперед при вирішенні задач, у яких наявність сильно корельованих ознак призводить до необ'єктивних результатів регресійного аналізу. Зокрема, коли знак параметра рівняння множинної регресії не відповідає апріорним припущенням щодо логіки взаємозв'язку між відповідною пояснюючою та залежною змінною.

Метод головних компонент вперше запропоновано К. Пірсоном, який використовував його у біометричних дослідженнях. Основне його призначення — виявити приховані (латентні) першопричини, які пояснюють кореляції між ознаками і змістовно інтерпретуються. Зв'язок між первинними ознаками і компонентами описується як лінійна комбінація [13]:

$$z_i = \sum_{j=1}^n a_{ij} F_j,$$

де z_i — стандартизовані значення i -ї ознаки з одиничними дисперсіями; a_{ij} — факторне навантаження i -ї компоненти на j -у ознаку (характеризує тісноту зв'язку між i -ю ознакою та j -ю компонентою).

Факторний аналіз вперше використано у дослідженнях в області психології Ч. Спірменом. Головна відмінність між цими двома методами полягає в наступному [19]:

— в моделі головних компонент критерієм оптимальності (щодо виокремлення компонент, які впливають на досліджувані параметри) є мінімум розбіжності між коваріаційною матрицею

вихідних ознак і тією, яку отримують у результаті оцінювання факторних навантажень;

— критерієм оптимальності у факторному аналізі є максимальна близькість вихідних кореляцій ознак до тих, які отримані в результаті оцінювання факторних навантажень.

У практиці управління соціально-економічними процесами ці методи застосовуються при ранжуванні та типології одиниць сукупності, при вивченні закономірностей динаміки, при вимірюванні взаємозв'язків тощо.

Не менш важливими як з точки зору пізнавальної діяльності, так і з позиції формування аналітичного забезпечення прийняття управлінських рішень є *методи багатовимірної класифікації*, які реалізуються за допомогою алгоритмів розпізнавання образів — кластерного та дискримінантного аналізу.

Хоча вперше назва кластер-аналіз з'явилася ще 1939 р. у працях французького дослідника Р. Тріюнона, значного розвитку методи кластерного аналізу набули значно пізніше, зокрема, після виходу у світ книги Р. Сокела та П. Сніта «Початки числової таксономії» (1963 р.) [16]. Сутність *кластерного аналізу* зводиться до об'єднання однорідних одиниць сукупності у групи. Однорідність при цьому задається правилом обчислення певної метрики, що характеризує ступінь подібності (схожості) j -ї та k -ї одиниць сукупності. В якості метрики може використовуватися відстань між ними c_{jk} (для ознак метричної шкали) або коефіцієнт подібності r_{jk} (якщо значення ознак представлені двійковим кодом). Вибір метрики є ключовим моментом кластерного аналізу, від якого залежить кінцевий результат поділу сукупності на класи. На практиці найуживанішими є наступні метрики [19]:

Евклідова відстань:

$$c_{jk} = \sqrt{\sum_{i=1}^m (z_{ij} - z_{ik})^2},$$

де z_{ij} та z_{ik} — стандартизовані значення i -ї ознаки в j -ї та k -ї одиниць сукупності;

якщо вихідні ознаки x_i різновагомі, то використовується зважена Евклідова відстань з вагами w_{ij} :

$$c_{jk} = \sqrt{\sum_{i=1}^m (z_{ij} - z_{ik})^2 w_{ij}};$$

Манхеттенська відстань:

$$c_{jk} = \sum_{i=1}^m |z_{ij} - z_{ik}|$$

Для ознак номінальної шкали в якості метрики використовують коефіцієнти подібності:

коефіцієнт Роджерса-Танімото:

$$r_{ij} = n''_{jk} / (n'_j + n'_k - n''_{jk}),$$

де n''_{jk} — кількість одиниць, значення яких збігаються; n'_j та

n'_k — кількість одиниць у групах;

коефіцієнт подібності Хемінга:

$$r_{ij} = \frac{n_{ij}}{m},$$

де n_{ij} — одиниць та нулів, що збігаються за ознаками у j -ї та k -ї одиниць сукупності; m — кількість ознак.

Також використовують коефіцієнти Рао, Жаккара та коефіцієнт Дейка, розрахунок яких базується на порівнянні частот однакових і різних пар значень ознак [13].

За способом кластеризації розрізняють ієрархічні та ітераційні процедури об'єднання. Серед ієрархічних процедур найбільш вживана агломеративна (об'єднувальна), яка передбачає послідовне об'єднання двох найближчих одиниць сукупності. Результати ієрархічних процедур кластеризації оформлюються у вигляді деревовидних діаграм — дендрограм, що значно спрощує сприйняття та інтерпретацію результатів класифікації.

На відміну від кластерного аналізу *дискримінантний* не утворює нових класів, а допомагає виявити різницю між вже існуючими класами і віднести новий (нерозпізнаний) об'єкт до одного з них за принципом максимальної схожості. Дискримінантний аналіз набув популярності у багатьох сферах наукових досліджень — в економіці, менеджменті, соціології, медицині, психології та ін., що насамперед пояснюється колом завдань, які він вирішує. Зокрема, у сфері економічних відносин даний метод використовується: органами нагляду при розробленні систем оцінювання фінансової стійкості та ймовірності банкрутства підприємств, банків; при побудові скорингових систем оцінювання кредитоспроможності позичальників; в управлінні персоналом;

розробці рейтингових систем оцінювання; при визначенні ризику реалізації інвестиційних проектів тощо.

Інструментом, за допомогою якого новий об'єкт відносять до того або іншого класу, є дискримінантна функція. Вона являє собою лінійну комбінацію певної множини ознак, які називаються класифікаційними і на основі яких ідентифікуються класи. Особливість дискримінантної функції полягає в тому, що класи представляються шкалою найменувань, а класифікаційні ознаки x_i вимірюються метричною шкалою. Дискримінантна функція f_j визначається для кожного j -го класу [19]:

$$f_j = a_{0j} + a_{1j}\bar{x}_{1j} + a_{2j}\bar{x}_{2j} + \dots + a_{mj}\bar{x}_{mj}$$

де a_{ij} — коефіцієнт функції (змістовної інтерпретації не має); \bar{x}_{ij} — середнє значення i -ї ознаки в j -му класі.

Сама ж процедура інтерпретації ґрунтується на геометричній близькості h -ї одиниці до центрів виділених класів (значень f_j). Належність її до того чи іншого класу визначається на основі відстані Махаланобіса:

$$D^2 = (n - p) \sum_{i=p}^m \sum_{k=1}^m b_{ik} (x_{ijh} - \bar{x}_{ij})(x_{kjh} - \bar{x}_{kj}).$$

Прикладна статистика є «споживачем» теоретичних розробок і водночас стимулом для розвитку статистичної теорії. Нові завдання, пов'язані з необхідністю прийняття управлінських рішень в умовах невизначеності та підвищених ризиків, породжують необхідність розробки нових, ефективніших, адаптованих до сучасних вимог інформаційного суспільства методів статистичного аналізу даних. Важливими чинниками розвитку статистики на сучасному етапі є досягнення математики та стрімкий розвиток інформаційних технологій — комп'ютерної техніки та прикладних програмних продуктів (MS Excel, Statistica, SPSS, SAS, MatLab та ін.), які реалізують широкий спектр статистичних методів: від описувальних до методів багатовимірного моделювання та прогнозування. Це відкриває широкі можливості для впровадження в практику сучасного менеджменту аналітичних прийомів, які передбачають проведення значних обсягів трудомістких обчислень.

Однак, з іншого боку, наявність потужних технічних засобів статистичного аналізу даних, призводить до зниження рівня ста-

тистичної підготовки багатьох дослідників, які не вдаються до з'ясування математичного каркасу методу статистики. Застосування програмних засобів без розуміння математичної суті використовуваних алгоритмів сприяє розвитку у дослідника ілюзії простоти застосування багатовимірних статистичних методів, що може призвести до невірних або необґрунтованих результатів. Значущі практичні результати можуть бути отримані лише на основі професійних знань у предметній області, підкріплених володінням математичними методами і пакетами прикладних програм, у яких ці методи реалізовані.

Висновки. У процесі розвитку суспільства все повніше розкривалися роль і значення статистики на всіх рівнях соціально-економічного життя суспільства. Статистична наука пройшла шлях від «описового» методу до науково обґрунтованої системи знань, положень, категорій та потужного засобу пізнання масових явищ та процесів. Статистику стали розглядати як засіб, який забезпечує виявлення невідомого, передбачення подій за заданих умов, як універсальний метод пізнання. Використання методів статистичного моделювання і прогнозування сприяли використанню статистики як методу прийняття рішень в умовах невизначеності та підвищили інтерес до неї з боку менеджменту.

Сучасне суспільство з урахуванням своїх потреб постійно ставить перед статистикою нові завдання, або вимагає нових підходів до вирішення уже існуючих проблем. Цим самим сприяє розвитку статистичного методу та розширенню його предметної області, приведенню у відповідність статистичної інформації та методів її аналізу. На сьогодні актуальними та перспективними напрямками розвитку прикладної статистики є, насамперед, галузеві дослідження з використанням методів обробки нечислової інформації, методів моделювання динаміки, економетричних методів, методів багатовимірної класифікації, зокрема, що базуються на концепції статистики нечітких множин, методів оцінювання якості інформації тощо. Світова наукова спільнота давно дійшла висновку, що найефективнішими є міждисциплінарні дослідження (Interdisciplinary research), які розширюють класичні межі економічних проблем та, що найважливіше, сприяють удосконаленню існуючої методології, запровадженню нових підходів, засобів та інструментів аналізу. Прикладом чого слугують дослідження Нобелівських лауреатів у галузі економіки та критерії до їх оцінювання: більшість праць, які були відзначені преміями — це міждисциплінарні дослідження, які проводяться на межі економіки, соціології, політології, філософії, генетики, фізики, медицини, психології тощо [1, 4, 6, 8, 10]. При цьому вико-

нані вони із широким застосуванням методів математичної статистики та теорії ймовірностей, математичного моделювання та програмування, засобів комп'ютерної математики, інформаційних систем та технологій.

Література

1. *Christopher Pissarides*. Labour Market Adjustment: Microeconomic Foundations of Short-Run Neoclassical and Keynesian Dynamics, Cambridge: University Press, 1976 (re-issued in paperback, 2009)
2. *Christopher Pissarides*. The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer? *Econometrica*, 77 (September 2009) 1339—1369.
3. *Christopher Pissarides*. Why Do Firms Offer 'Employment Protection'? *Economica*, 77 (October 2010) 613—636.
4. *D. T. Mortensen*. «Wage Dispersion in the Search and Matching Model» *American Economic Review: Papers & Proceedings* (2010) 100: 2, 1—10.
5. *Diamond Peter*. «My Research Strategy.» Forthcoming *Eminent Economists II — Their Work and Life Philosophies*. New York: Cambridge University Press, New York.
6. *Diamond Petr*. Economic Theory and Tax and Pension Policies / The Economic Record published by The Economic Society of Australia June 9, 2011.
7. *Gruber, Jonathan, and David A. Wise*, (eds.). 2010. Social Security Programs and Retirement around the World: The Relationship to Youth Employment. National Bureau of Economic Research Conference Report. Chicago, IL: University of Chicago Press.
8. *Ostrom E*. People and Forests: Communities, Institutions, and Governance (MIT Press, Cambridge, MA), 2000.
9. *R. Lentz and D. T. Mortensen*. «Labor Market Models of Worker and Firm Heterogeneity,» *Annual Review of Economics* (2010) 2: 577—602.
10. The Nobel Prize: The First 100 Years», Agneta Wallin Levinovitz and Nils Ringertz, editors, Imperial College Press and World Scientific Publishing Co. Pte. Ltd., 2001.
11. Вопросы статистической методологии и статистико-экономического анализа: материалы Всесоюзного научного совещания. — М.: Статистика, 1980. — 208 с.
12. *Головач А. В.* Критерии математической статистики в экономических исследованиях / А. В. Головач, А. М. Ерина, В. П. Трофимов. — М.: Статистика, 1973. — 136 с.
13. *Єріна А. М.* Статистичне моделювання та прогнозування: Навч. посібник / Єріна А. М. — К.: КНЕУ, 2002. — 170 с.
14. *Кетле А.* Социальная физика, или Опыт исследования о развитии человеческих способностей / А. Кетле; пер. студ. Киев. коммерч. ин-та под ред. А. Русова // *Известия Киевского коммерческого института*. — 1913. — Кн. 19—21.

15. Орлов А. И. Нечисловая статистика / А. И. Орлов. — М.: МЗ-Пресс, 2004. — 513 с.

16. Плошко Б. Г. История статистики: учеб. пособие / Б. Г. Плошко, И. И. Елисеева. — М.: Финансы и статистика, 1990. — 295 с.

17. Слуцкий Е. Е. Сэр Вильям Петти. Краткий очерк его экономических воззрений с приложением нескольких важнейших отрывков из его призываний / Слуцкий Е. Е. — К., 1914. — 48 с.

18. Слуцкий Е. Е. Теория корреляции и элементы учения о кривых распределения / Е. Е. Слуцкий. — Известия Киевского коммерческого института. — К., 1912. — Кн. XVI. — IV. — 208 с.

19. Факторный, кластерный и дискриминантный анализ / [Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др.]; пер. с англ.; Под ред. И.С. Енюкова. — М.: Финансы и статистика, 1989. — 215 с.

20. Чекотовський Е. В. Історія статистичної науки: навч. посіб. / Е. В. Чекотовський. — К.: Знання, 2011. — 495 с.

21. Четыркин Е. М. Статистические методы прогнозирования: Изд. 2-ге, перераб. и доп. / Четыркин Е. М. — М.: Статистика, 1977. — 200 с.

22. Шеффе Г. Дисперсионный анализ / Г. Шеффе. — М.: Главная редакция физико-математической литературы изд-ва «Наука», 1980. — 512 с.

Стаття надійшла до редакції 13.05.2011 р.

УДК 519.86: 330.3

В. Д. Дербенцев, канд. екон. наук, доц.,
ДВНЗ «КНЕУ імені Вадима Гетьмана»

МОДЕЛЮВАННЯ ГРАНИЧНОЇ СХИЛЬНОСТІ ДО ІНВЕСТИЦІЙ ТА ДИНАМІКИ ВВП

АНОТАЦІЯ. Робота присвячена питанням моделювання динаміки валового внутрішнього продукту залежно від граничної схильності до інвестицій, частки податків та зарплати у структурі ВВП, норми амортизації та параметрів виробничої функції. Знайдено чисельну оцінку нижньої межі граничної схильності до інвестування, що гарантує необхідні умови для економічного зростання.

КЛЮЧОВІ СЛОВА. Модель динаміки ВВП, виробнича функція, гранична схильність до інвестицій.

АННОТАЦИЯ. Работа посвящена вопросам моделирования динамики валового внутреннего продукта в зависимости от предельной склонности к инвестициям, удельного веса налогов и зарплаты в структуре ВВП, нормы амортизации и параметров производственной функции. Найдена численная оценка нижней границы предельной склонности к ин-