

Статистичний аналіз величини елементарних соціально-економічних ознак об'єкту

Необхідним початковим етап вивчення природи функціонування та розвитку об'єкта в економіці являється аналіз спостережень чи даних його елементарних ознак, що передбачає аналіз закономірної та випадкової мінливості величини ознаки, сформованої в показник, виявлення закономірної мінливості на фоні випадкової. Сучасні методи математичної статистики та їх комп'ютерна реалізація у вигляді спеціальних пакетів дозволяють комплексно провести дослідження, оцінити параметри наявних закономірностей, перевірити різні гіпотези про ці закономірності.

Для того щоб об'єктивно дослідити величину елементарної ознаки об'єкта, рівні її значень, потрібно аналізувати сукупність значень в динаміці, по-перше, на даному об'єкті, по-друге, в однорідній групі об'єктів, по-третє, у всій сукупності об'єктів, якщо це можливо і доцільно. Переконані, що дане дослідження величини ознаки можливе, якщо використати засоби описової статистики. Корисність методів описової статистики ґрунтується на тому факті, що декілька простих та досить інформативних статистичних показників та візуальних засобів виключають монотонний перегляд всіх значень величини ознаки.

До інструменту описової статистики входять:

1. Показники положення чи стану описують положення значень величини ознаки на числовій осі. Це вибіркове середнє, вибіркова медіана, мода, найбільше та найменше значення величини ознаки, перцентилі, зокрема, квартилі.
2. Показники розкиду описують степінь розкиду значень величини відносно свого центру та характеризують степінь їх мінливості. До цих показників відносяться дисперсія сукупності значень (вибіркової чи генеральної), стандартне відхилення, розмах сукупності, міжквартильний розмах. Ці показники характеризують кучність значень величини відносно свого центру.
3. Показники форми розподілу значень величини ознаки – коефіцієнт асиметрії і ексцесу. Відомо, що до показників асиметрії відносяться: коефіцієнт асиметрії, стандартизований коефіцієнт асиметрії, положення вибіркової медіани відносно вибіркового середнього, положення вибірових квартилів і так далі. Важливість коефіцієнта асиметрії обумовлена тим, що він описує симетрію розподілу значень величини відносно свого центру. Плосковершинність (гольчатість) розподілу

значень величини ознаки відображує коефіцієнт ексцесу. Якщо значення стандартизованого коефіцієнту ексцесу менше двох, то асиметрії не має.

4. Графічні засоби, що описують закон розподілу значень величини ознаки. Приведені засоби дають уявлення про закон розподілу значень. До їх складу відносяться гістограми, кумуляти, блокові діаграми, таблиці частот.

Вважаємо, що комплекс перерахованих засобів описової статистики є основним, достатнім для комплексного дослідження різних величин ознак об'єкту в економіці. Звичайно, даний комплекс можливо розширяти, але одержання принципової інформації про величину здійснюється завдяки приведеному комплексу засобів. На практиці традиційно використовується вибіркове середнє, медіана, дисперсія і стандартне відхилення. Проте для одержання точніших та достовірніших висновків настійно рекомендують уважно вивчати і інші характеристики, а також звертати увагу на умови одержання вибірових сукупностей.

Особливу увагу слід звертати на наявність у сукупності значень викидів – грубих (помилкових), надто відмінних від основного масиву значень. Саме вони спотворюють вибіркові характеристики, зокрема, середнє, дисперсію, стандартне відхилення, коефіцієнти асиметрії та ексцесу. Викиди виявляють або за допомогою правила 3-х сигм, або за допомогою гістограми з достатньо великою кількістю інтервалів групування, або за допомогою детальної блокової діаграми. Присутність викидів також приводить до багатьох помилкових висновків, тому потрібно ретельно вивчати причину викидів. Традиційно статистичні характеристики чутливі до відхилень від умов застосування методу. Інтенсивно розвиваються сучасні статистичні методи, які стійкі до викидів, але вони ще не одержали широкого розповсюдження на практиці, за виключенням рангових процедур для найстандартніших задач. Причину тут бачимо у обчислювальній складності цих методів, що утруднює їх застосування у відсутності спеціальних комп'ютерних програм.

Описова статистика формує перше уявлення про теоретичні та вибіркові характеристики випадкових величин елементарних ознак, дозволяє досліджувати тенденції змін значень величини ознаки з врахуванням реальних умов функціонування об'єкту. В складних ситуаціях вона представляє ефективні засоби «побачити всю картину». Детальне вивчення кожного окремого випадку само по собі не являється задачею статистики, але вимірювання величин ознак у розв'язуванні практичних задач управління об'єктом в економіці потребує виявлення та ідентифікації особливостей, які в

цілому характерні для розглядуваних випадків та складають цілі описової статистики ознак. Досягнення цілі статистики – опису інформації, що містять великі сукупності значень величин невеликим числом показників, що виражають найфундаментальніші властивості сукупностей, можливо здійснити за допомогою раціонально побудованого комплексу засобів описової статистики чи, як їх ще іноді називають – методів статистичного аналізу. Отже потрібно детально проаналізувати кожний засіб чи інструмент описової статистики, щоб відмітити важливість його включення до комплексу та створення єдиного інструменту дослідження значень величини ознаки об'єкту.

Характеристику положення значень величини надають середня, медіана, мода – це різні способи вибору єдиного числа, яке найкраще описує всі значення величини в сукупності. Такий представлений одним числом показник називають типовим значенням або центром, мірою центральної тенденції зміни значень величини. Визначення демонструє безсумнівну важливість даного показника.

Середнє найкорисніше в якості узагальнюючого показника положення при відсутності викидів, коли набір значень представляється порівняно однорідною групою з елементами випадковості.

Середньозважене найкраще інтерпретувати як середнє, що використовується в ситуаціях, коли розставляється пріоритетність серед рівнів значень. Найважливіші рівні значень величини ознаки можуть вносити більший внесок в обчислення значення середнього зваженого. Зважені середні використовують для того, щоб скоректувати недостатки репрезентативності вибірки по відношенню до генеральної сукупності.

Медіана розміщується в центрі сукупності значень величини і дає уявлення про всю послідовність значень. Щоб знайти медіану, значення розміщуються в зростаючому порядку, а потім визначають центральне значення. У половини всіх спостережень значення величини ознаки більше медіани, у половини – менше. Якщо в сукупності даних немає одного центрального значення, то усереднюють ті два значення, які розположені посередині послідовності значень величини. Медіану можливо визначити в термінах рангу (коли це необхідно для метричної величини, і як єдиний спосіб – для неметричної величини). Ранги зв'язують числа $1, 2, \dots, n$ із значеннями величини таким чином, щоб найменше значення мало ранг 1, наступне значення величини – ранг 2 і так далі до найбільшого значення, що має ранг n . Якщо маємо сукупність значень метричної величини, то медіана являється середньою значень. Якщо сукупність значень величини, вимірюваної за допомогою

порядкових шкал, тобто значення виражаються ординальними числами, то значення розміщене в середині ряду значень є медіаною кількісних значень чи предікатних значень величини ознаки. Медіана – типове усталене до наявності викидів значення метричної і неметричної величини ознаки.

Якщо значення величини в сукупності розподілені симетрично, то значення медіани і середнього близькі між собою, оскільки симетричні розподіли мають чітко виражену середню точку. Але навіть при «практично нормальному» розподілу середнє і медіана можуть дещо відрізнятися один від одного, оскільки кожна з цих характеристик визначається по-своєму, і крім того, в реальних значеннях величини ознаки майже завжди присутня випадковість. Наявність викидів різко відображується на відмінності між середнім і медіаною. Якщо розподіл значень величини асиметричний, то медіана і середнє сильно відрізняються. Зазвичай середнє відносно медіани здвинуто в напрямку довшого хвоста чи в напрямку викидів, оскільки середнє реально враховує значення таких екстремальних спостережень, в той час як для медіани важливо лише по якій стороні від неї лежить те чи інше значення величини ознаки.

Мода визначається найпоширенішим значенням величини ознаки, тобто значенням, що найчастіше зустрічається в сукупності. Це єдина характеристика, яку можна визначити для неметричної величини, що вимірюється за допомогою шкал назв чи номінацій, оскільки неупорядковані категорії якісних описових ознак об'єкту не можуть бути просумовані, як це потрібно для середнього і не можуть бути проранжовані, як це потрібно для медіани. Моду знаходять і для величин, що шкалювані порядково, якщо ігнорувати упорядкованість категорій та виконати всі дії так як для сукупності величин номінальних ознак з неупорядкованими категоріями. Мода визначається для метричних величин за найвищою точкою на гістограмі, хоча при цьому може мати місце деяка невизначеність, наприклад, на гістограмі можуть бути два високі стовпці або визначення можуть залежити від того, яким чином побудована діаграма: зміна ширини стовпців може привести до умірених змін форми розподілу значень, в результаті чого мода може змінитися. Мода для сукупності значень метричних величин дуже неусталена характеристика. Вона зосереджує увагу лише на найважливішому значенні величини ознаки. При аналізі неметричних величин ознак акцентуємо увагу на модальній категорії ознаки.

Отже, рекомендації по виборі характеристик сукупності значень величин ознак об'єкту залежно від типу ознак сформулюємо так. У випадку ознак, що мають метричну величину, що дозволяє обчислювати всі три характеристики,

якщо розподіл значень величини близький до нормального, то різниця між трьома характеристиками буде невелика. У випадку асиметричного розподілу значень величини ці три характеристики можуть помітно відрізнятися. Середнє слід використовувати, коли сукупність значень величини має розподіл нормальний чи близький до нормального, оскільки в цьому випадку середнє являється самою простою характеристикою. Медіану доцільно використовувати при асиметричному розподілу, оскільки на нього не впливає невелика кількість значень високого рівня. У випадку сильної асиметрії розподілу медіана значно краще середнього характеризує більшість значень величини. Використання медіани надзвичайно доцільне наразі в сукупності значень величини ознаки викидів, оскільки вона усталена до їх впливів. Медіану корисно обчислювати для сукупності значень неметричних величин номінальних ознак, а також ознак вимірних за порядковими шкалами, оскільки важливо знати найпоширенішу категорію.

Перцентилі узагальнюють інформацію про сукупність значень величин ознак, характеризуючи значення, що досягається заданим процентом загальної кількості спостережень, після того, як значення проранжуються по зростанню. Перцентилі – це характеристики сукупності значень величини ознак, які виражають ранги значень у вигляді процентів від 0 до 100%, а не у вигляді чисел від 1 до n , таким чином, найменшому значенню відповідає нульовий перцентиль, найбільшому – 100-й перцентиль, медіані – 50% перцентиль і так далі. Перцентилі іноді розглядають як показники, що характеризують розділення сукупності значень різних величин на окремі частини. Перцентилі використовують в двох цілях: щоб показати рівень величини ознаки при заданому перцентильному ранзі; щоб показати перцентильний ранг даного значення в сукупності. Перцентилі відіграють важливу роль опорних характеристик. Для того, щоб узагальнити основні риси розподілу значень величини, достатньо мати декілька значень перцентилів. Так, 50-й перцентиль – це медіана, оскільки він знаходиться посередині між найбільшим і найменшим значеннями величини ознаки. Представляють інтерес екстремуми величини ознаки – найбільші і найменші значення. Доповнюють набір базових характеристик сукупності значень величини квартилі, що визначаються як 25 і 75-й перцентилі.

Існує декілька методів визначення вибірових значень кварталів, різних за алгоритмом обчислення. Квартилі – це значення ранжованого ряду значень величини, що знаходяться на відстані однієї чверті в послідовності від найменшого до найбільшого значення. Очевидний факт, що дане визначення не

указує точно як обчислити квартилі. Джон Т'юкі, один із засновників практичного аналізу даних, визначає квартилі таким чином: обчислює ранг медіани за формулою $\frac{1+n}{2}$ і відкидає дробову частину; прибавляє до одержаного значення 1 та ділить на 2. Отримане значення представляє собою ранг нижнього квартилю; віднімає отримане значення від $(n+1)$, результатом буде ранг верхнього квартилю. Значення кварталів обчислюється, виходячи з цих рангів. Таким чином приведена схема обчислення кварталів реалізується формулами:

$$\text{ранг нижнього квартилю} = \frac{1 + \text{int}\left[\frac{1+n}{2}\right]}{2};$$

ранг верхнього квартилю $= n + 1 - \text{ранг нижнього квартилю}$, де *int* позначає функцію цілого значення величини.

В статистичному аналізі відомий потужний засіб проведення порівняльного аналізу за різними вибірками однієї і тієї ж ознаки – це блокова діаграма Т'юкі, або її ще називають «ящик з вусами». Блокова діаграма зображує значення п'яти базових статистичних показників, що характеризують положення значень величини ознаки. Вона містить п'ять базових показників множини значень величини однієї ознаки, що дозволяє швидко визначити характер розподілу. Блокова діаграма як і гістограма візуально демонструє розподіл, але при цьому є іншим способом графічного відображення. Блока діаграма не обтяжена деталями і тому охоплює статистичний опис множини значень величини ознаки в цілому, та дозволяє порівнювати декілька груп ознаки, що мають різні значення величини, не вдаючись в кількісне описання кожної групи. Зазначимо, що аналіз форми розподілу краще виконувати на гістограмі. Якщо відсутні викиди, то до п'яти базових показників входять найменше значення величини (0-й перцентиль), нижній квартиль (25-й перцентиль, на четверть відстані від найменшого значення), медіана (50-й перцентиль, середина), верхній квартиль (75-й перцентиль, на три четверті відстані від найменшого значення або на четверть відстані від найбільшого значення), найбільше значення (100-й перцентиль). Якщо ж множина значень величини ознаки має викиди, то вони зображуються світлими точками, тоді межі вусів встановлюють на 1,5 міжквартильного розмаху. Разом ці перераховані статистичні характеристики дають достатньо чітке уявлення про особливості ще не обробленої множини значень величини ознаки. Два значення – найбільше і найменше, характеризують розмах (діапазон) зміни значень величини, медіана показує

центр, два квартилі визначають межі «розміщеної в центрі половини множини значень величини», а положення медіани відносно квартилей і середнього дає грубе уявлення про наявність чи відсутність асиметрії. Велике значення величини розглядається як викид, якщо воно перевищує верхній квартиль плюс 1,5 міжквартильного розмаху. Мале значення величини розглядається як викид, якщо воно менше, ніж нижній квартиль мінус 1,5 міжквартильного розмаху. Так визначаються викиди за Д. Т'юкі.

Другий тип інструментів описової статистики направлений на вивчення різноманіття значень величини ознаки, тобто на мінливість значень величини. Очевидний факт, що коли б значення величини ознаки не змінювались, то не було б сенсу звертатися до методів статистичного аналізу. Ситуація, в якій присутня мінливість, часто містить ризик, оскільки навіть використання всієї доступної інформації не дозволяє точно передбачити, що відбудеться в майбутньому. Для об'єктивного аналізу ризику потрібно розуміти його природу і вміти виміряти мінливість, що являється його наслідком.

Мінливість є мірою відмінності окремих значень величини ознаки. В той час, коли статистичні характеристики центру: середнє значення, медіана, мода вказують на типове для величини значення, мінливість показує, наскільки близько до цього центру зазвичай розміщені окремі значення в множині. Якщо всі значення величини однакові, то мінливість дорівнює нулю. Чим більше розкид значень величини, тим більша її мінливість, мінливість ознаки об'єкту.

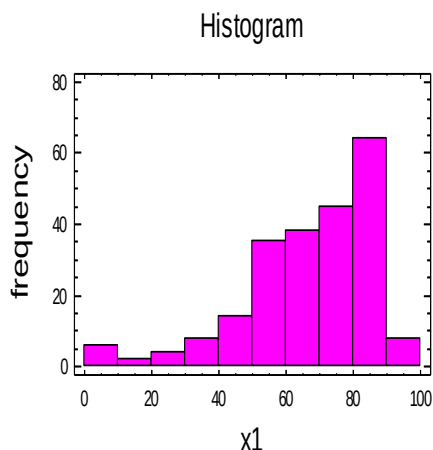
Мінливість визначається як степінь відмінностей між окремими значеннями величини ознаки. Подібний сенс мають також такі поняття, як різноманіття, невизначеність, розсіяння та розкид. Існують чотири основні характеристики описання степені мінливості множини значень величини: 1. Розмах, простий в обчисленні та надає дещо поверхневе уявлення про мінливість значень величини, і має обмежене використання. Ця характеристика описує межі змін значень величини в множині та є відстанню між найменшим і найбільшим значеннями. Великі відхилення рідкі. Тому у вибірці значень ознаки найменше і найбільше значення може не попасти, звідси розмах буде обчислений за іншими числами та буде заниженим, а отже розмах – неусталена та мало надійна статистична характеристика. 2. Стандартне відхилення використовують найчастіше. Цей статистичний показник описує, наскільки сильно значення величини ознаки, що спостерігається, зазвичай відрізняється від середнього значення. При піднесенні стандартного відхилення до квадрату маємо дисперсію. Стандартне відхилення відображує типову відстань між середнім значенням і окремими значеннями величини в множині. Стандартне

відхилення показує степінь випадковості в розміщенні окремих значень відносно їх загального середнього. Відхилення – це відстань між кожним значенням і середнім в множині значень величини ознаки. Стандартне відхилення являється деяким середнім з цих відхилень - «середнім квадратичним відхиленням», оскільки звичане середнє арифметичні завжди тут дорівнює нулю із-за різних знаків окремих відхилень. Стандартне відхилення є надзвичайно важливою характеристикою в статистиці, оскільки воно представляє собою основний інструмент визначення степені випадковості у дослідженні величини ознаки. Порівняння середньоквадратичних відхилень значень величини однієї і тієї ж ознаки в різних розподілах величини дозволяє встановити де варіація ознаки більша. Воно являється також показником однорідності даної сукупності значень величини ознаки. Невелике стандартне відхилення говорить про те, що середня добре відображує в собі представляєму сукупність значень величини. Тому кожне обчислення середньої в розподілу значень повинне супроводжуватися обчисленням стандартного відхилення. Коли сукупність значень величини ознаки має приблизно нормальний розподіл, стандартне відхилення набуває особливого сенсу. Ймовірність того, що випадкові відхилення значень величини від свого середнього будуть менші, ніж 2σ дорівнює 0,954, це є правило «двох сігм» при нормальному розподілі. При будь-яких інших розподілах використовується менше визначене відхилення випадкової величини від свого середнього, що по абсолютній величині практично не перевищує утроєного середнього квадратичного відхилення. На цьому правилі основана така наближена оцінка середнього квадратичного відхилення: з сукупності значень величини ознаки вибирають максимальне і мінімальне значення та їх різницю ділять на 6. Одержане число приймають як грубу оцінку середнього квадратичного відхилення. Відмітимо, що існує безліч скошених (чи інших, що відрізняються від нормального) розподілів, для яких не можливо вказати рівня довіри. В таких випадках використовують співвідношення, яке називається нерівністю Чебишева. Відповідно до цього правила, по меншій мірі $1 - \frac{1}{t^2}$ значень попадає в проміжок, що лежить в межах t стандартних відхилень від середнього значення. Наприклад, для $t = 3$ отримуємо, що по меншій мірі 88,9% значень сукупності будуть знаходитися в межах потроєного стандартного відхилення від середнього значення, тобто для довільного розподілу з гарантією 89% і вище справедливе відоме правило 3-х сігм. В силу чуттєвості до граничних значень розмах виявляється не зовсім корисним в якості тієї статистичної міри розкиду, яка характеризує сукупність

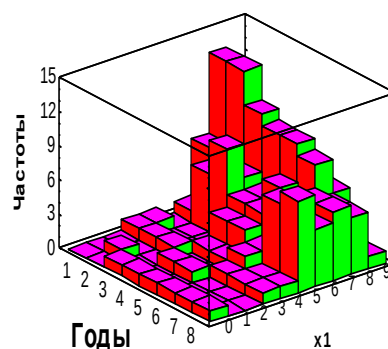
значень величини ознаки в цілому. Розмах не відображає типову мінливість значень, а скоріше фокусується всього лише на двох значеннях. Стандартне ж відхилення обчислюється по всім значенням сукупності, завдяки чому дана характеристика є об'єктивнішою і усталенішою, та дає чітке уявлення про всю картину значень в сукупності. 2. Коефіцієнт варіації вибирається в якості відносної (протилежно абсолютній) мірі мінливості. Цей показник використовується достатньо часто та показує наскільки сильно зазвичай відрізняється результат окремого спостереження ознаки від середнього значення у процентному відношенні до середнього, при цьому використовується відношення стандартного відхилення від середнього значення величини ознаки. Коефіцієнт варіації – безрозмірна статистична величина. Його доцільно використовувати для порівняння мінливості сукупностей значень величин ознак, що представлені в різних одиницях вимірювання. Слід відмітити, що коефіцієнт варіації може дорівнювати і навіть бути більшим за 100%. Зокрема, при показниковому законі розподілу теоретичний коефіцієнт варіації дорівнює 100%, оскільки для цього розподілу дорівнюють між собою $M(x) = \sigma$. Коефіцієнт варіації використовується для відбраковки маловаріабельних величин ознак, коли значення коефіцієнта менше 5%. 4. Міжквартильний розмах являється «робастною» (усталеною до наявності викидів) мірою мінливості.

Таким чином, виконавши теоретичний аналіз існуючих методів та засобів процесу проведення статистичного аналізу описання значень величин елементарних ознак об'єктів в економіці, одержані узагальнення відобразимо у вигляді їх концептуальної схеми, що приведена на Рис. 1.

Рекомендована концептуальна схема була реалізована для статистичного описання значень величин 48 ознак, що характеризують розвиток економіки 28 промислових підприємств Харківського регіону протягом 8 років з використанням спеціального статистичного пакету *Statgraphics Plus V5.1 International Professional*. За викладеними рекомендаціями описової статистики з системи показників спочатку слід виключити ті показники, у яких коефіцієнт варіації менше 5%, це виявились x_{11}, x_{12}, x_{15} . Отже, тепер в системі 45 показників. Згідно першого етапу схеми був виконаний візуальний аналіз розподілу значень величини ознак за допомогою гістограм – двохмірних та трьохмірних, що враховують динаміку та типи розвитку в однорідних групах підприємств, отриманих за допомогою кластерного аналізу, окремий розгляд якого здійснено в наступному розділі.



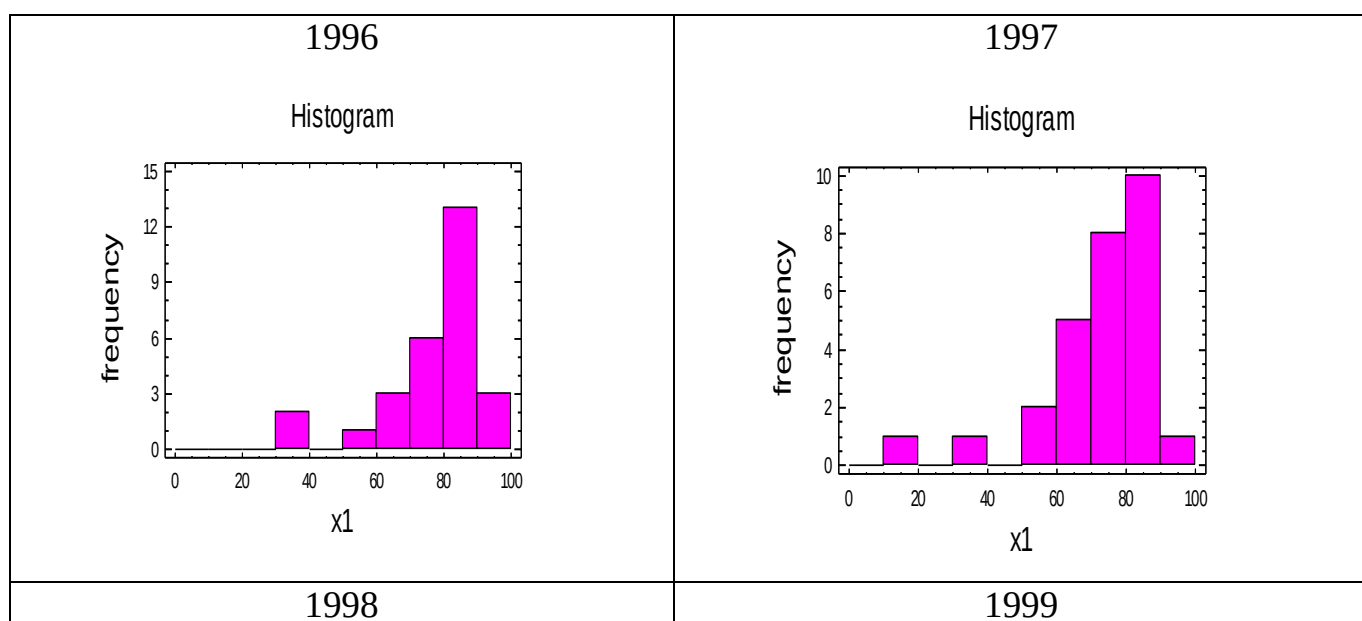
а)



б)

Рис. 2. Двохмірна (а) та трьохмірна гістограми частот значень величини ознаки x_1

На Рис. 2. (а) зображена гістограма частот елементарної ознаки у всій сукупності підприємств протягом всього аналізованого періоду (роки замінені порядковим номером). Відмічаємо асиметрію в розподілі значень ознаки в напрямку високих значень. Трьохмірну гістограму (Рис. 2. (б)) можна представити як набір гістограм ознаки за кожний рік (всього 8 гістограм, що демонструють динаміку мінливості значень величини ознаки в часі). Оскільки просторовий графік достатньо складний для сприйняття, доцільно мати детальний аналіз розподілу значень ознаки в перетині кожного року (Рис. 3.)



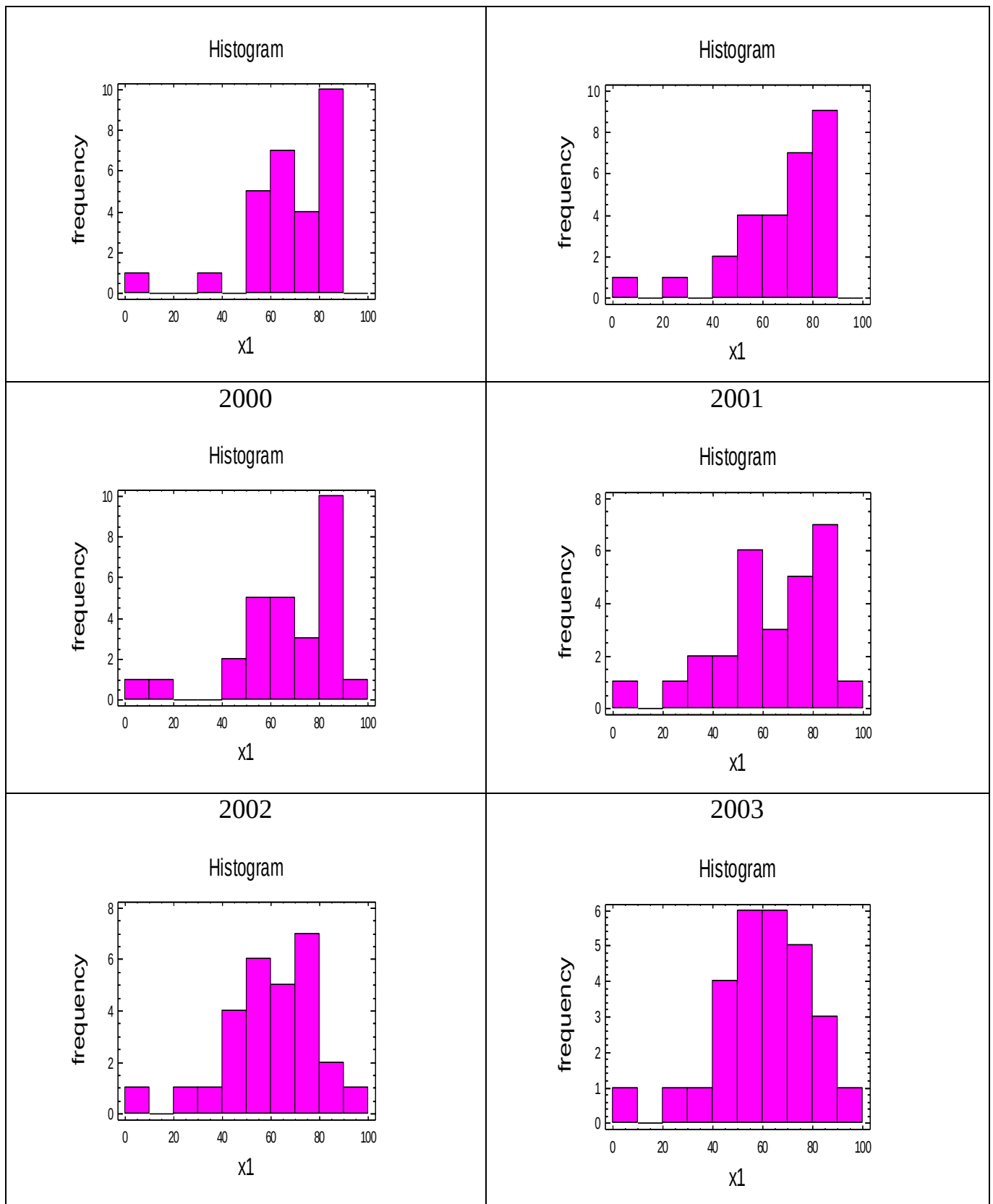


Рис. 3. Гістограми розподілу значень долі основних засобів та поза оборотних активів в балансі підприємств щорічно

Порівняльний аналіз даних гістограм засвідчує асиметрію в розподілі значень ознаки в перших шести роках, що дає змогу говорити про хаотичність розвитку підприємств і тільки в останні два роки спостерігаємо деяку узгодженість в розподілі.

За другим етапом статистичного аналізу потрібно обчислити показники, що характеризують положення значень величини ознаки на числовій осі в кожній з 8-ми груп підприємств та в цілому в сукупності за 2003 рік.

Сумарная статистика для ознаки x_1

Таблиця 1

Статистичний показник	Номер групи підприємств (2003 г.)								Загальне по сукупності
	1	2	3	4	5	6	7	8	
Кількість об'єктів в групі	9	9	2	1	1	3	1	2	28
Середнє	75,04	51,45	52,22	64,88	94,02	45,77	1,85	65,76	59,80
Медіана	75,79	51,8	53,22	64,88	94,02	47,25	1,85	65,76	61,034
Мода			53,2	64,88	94,02		1,85	60,3	
Стандартне відхилення	5,49	8,30	0,02	0,00	0,00	16,79	0,00	7,71	18,70
Мінімум	66,70	37,65	53,20	64,88	94,02	28,28	1,85	60,30	1,85
Максимум	82,62	67,00	53,23	64,88	94,02	61,77	1,85	71,21	94,02
Нижня квартиль	71,23	46,00	53,20	64,88	94,02	28,28	1,85	60,30	50,45
Верхня квартиль	80,13	55,48	53,23	64,88	94,02	61,77	1,85	71,21	72,51
Міжквартильний розмах	8,90	9,48	0,03	0,00	0,00	33,49	0,00	10,91	22,06
Коефіцієнт варіації	7,3%	16,1%	0,0%	0,0%	0,0%	36,7%	0,0%	11,7%	31,3%

Порівнюючи значення середнього, медіани та моди в таблиці, витікає висновок про практично нормальний закон розподілу значень ознаки в кожній групі. Згідно значень величини стандартного відхилення в таблиці маємо, що ознака x_1 мінлива у всій сукупності підприємств, а також у групах: 6-й та 2-й. Перша група має найменший міжквартильний розмах, отже спостерігається скупченість значень величини, таку ж ситуацію маємо в 2-й групі. Розмита за значеннями величини ознаки 6-та група, звичайно, тут маємо найбільший коефіцієнт варіації. Для детального вивчення різних джерел варіації виконують дисперсійний аналіз (скорочено дана стандартна обчислювальна процедура

позначається ANOVA). Табл. 2. містить результати обчислень дисперсійного аналізу величини ознаки x_1 : сум квадратів, степенів свободи, середніх квадратів, F - відношень та p -значень.

Таблиця дисперсійного аналізу

Таблиця 2

Вид мінливості (джерело)	Сума квадратів	Степені свободи	Середні квадрати	F - відношення	Рівень значущості
Між групами	8023,01	7	1146,14	16,20	0,0000
В середині груп	1415,2	20	70,7599		
Загальна	9438,21	27			

F - статистика для однофакторного дисперсійного аналізу представляє відношення дисперсії з двох джерел: між групову варіацію та в середині груп. F - статистика показує у скільки разів мінливість між групами перевищує мінливість в середині груп, яка вважається випадковою. З Табл. 2. маємо, що міжгрупова дисперсія (по групах підприємств) в 16,2 рази перевищує варіацію в середині груп (випадкову). Таким чином, варіація значень величини ознаки по групах в 16,2 рази більше, ніж можливо було б чекати, виходячи тільки з варіації окремих значень ознаки на конкретному підприємстві. В останній колонці Табл. 2. приведена ймовірність отримання указанного чи більшого значення F -відношення (рівень значущості F -статистики), при справедливості нульової гіпотези ця ймовірність (рівень значущості) практично дорівнює нулю, тому між виділеними групами об'єктів наявні значимі відмінності. Отже, потрібно вияснити між якими групами спостерігаються істотні відмінності. Для цього корисно обчислити межі довірливих 95-них інтервалів на середні в кожній групі та порівняти їх. Результати цих обчислень приведені в Табл. 3.

Таблиця 3

Номер групи	Кількість об'єктів в групі	Середнє	Похибка середнього	Нижня межа	Верхня межа
1	9	75,0444	2,80396	70,9086	79,1803
2	9	51,4467	2,80396	47,3108	55,5825
3	2	53,215	5,94811	44,4415	61,9885
4	1	64,88	8,41189	52,4724	77,2876
5	1	94,02	8,41189	81,6124	106,428
6	3	45,7667	4,85661	38,6032	52,9302
7	1	1,85	8,41189	-10,5576	14,2576
8	2	65,755	5,94811	56,9815	74,5285
Загальне	28	59,8004			

На Рис. 5. зображені середні в кожному кластері з 95% процентним довірливим інтервалом.

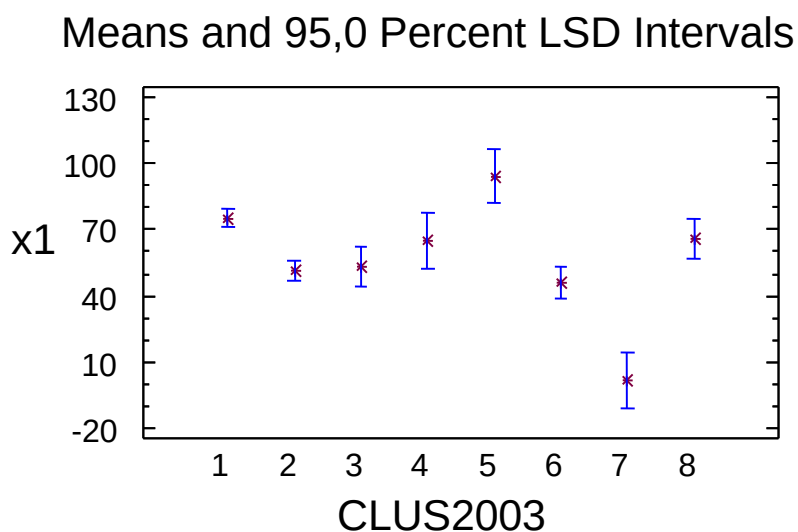


Рис. 5. Середні значення величини ознаки x_1 в кожному кластері (8 кластерів) з 95%-ними довірливими інтервалами

Чим більше об'єктів в групі, тим надійніше визначене середнє в групі, тим вужче довірливий інтервал (ширина якого обернено пропорційна кореню квадратному з числа об'єктів). Найширші інтервали (полоси невизначеності) для 4, 5, 7 груп, які містять по одному об'єкту (маємо найбільші розсіяння значень величини ознаки). Ширина довірливих інтервалів в 1-ій та 2-ій групах в 3 рази вужче, оскільки ці групи містять по 9 об'єктів. Довірливі інтервали для

деяких груп перекриваються, що відповідає відсутності значущих відмінностей між цими групами.

Враховуючи переваги блокових діаграм, що заключається в можливості сконцентрувати увагу на основних особливостях декількох множин значень величини ознаки одночасно, не відволікаючись на деталі. Отже маємо вісім самостійних наборів значень величини ознаки (кластерів). Для кожної групи підприємств обчислили п'ять базових статистичних величин та побудували блокову діаграму. Розмістив на одному малюнку побудовані в одному масштабі блокові діаграми, можна легко порівняти між собою особливості розподілу значень величини ознаки x_1 . На Рис. 6. зображені «ящики з вусами», побудовані по групах підприємств (часовий зріз: 2003 рік).

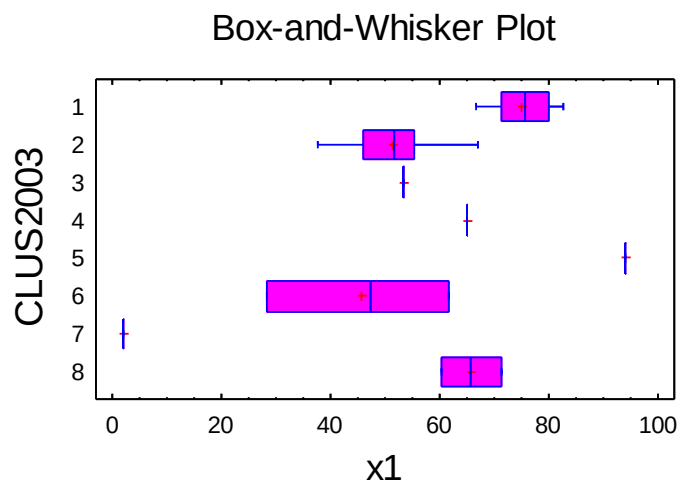


Рис. 6. Блокова діаграма значень величини ознаки x_1 по групах підприємств (в 2003 році)

Блокові діаграми 6-ої та 8-ої груп підприємств на мають «вусів», оскільки в цих групах мало об'єктів. Блокові діаграми 3-ої, 4-ої, 5-ої та 7-ої груп виявились виродженими, оскільки ці групи складаються з одного чи двох об'єктів. Результати аналізу множинних порівнянь по групах представлені в Табл. 4.

Порівняння груп об'єктів

Таблиця 4

Групи	Кількість об'єктів в групі	Середнє	Однорідні групи
7	1	1.85	X
6	3	45,7667	X
2	9	51,4467	X
3	2	53,215	XX
4	1	64,88	XXX
8	2	65,755	XX
1	9	75,0444	X
5	1	94,02	X
Парне порівняння	Різниця між середніми	Найменша суттєва різниця	
1-2	*23,5978	8,2717	
1-3	*21,8294	13,7171	
1-4	10,1644	18,4961	
1-5	*-18,9756	18,4961	
1-6	*29,2778	11,698	
1-7	*73,1944	18,4961	
1-8	9,28944	13,7171	
2-3	-1,76833	13,7171	
2-4	-13,4333	18,4961	
2-5	*-42,5733	18,4961	
2-6	5,68	11,698	
2-7	*49,5967	18,4961	
2-8	*-14,3083	13,7171	
3-4	-11,665	21,4905	
3-5	*-40,805	21,4905	
3-6	7,44833	16,0181	
3-7	*51,365	21,4905	
3-8	-12,54	17,5469	
4-5	*-29,14	24,8151	
4-6	19,1133	20,2615	
4-7	*63,03	24,8151	
4-8	-0,875	21,4905	
5-6	*48,2533	20,2615	
5-7	*92,17	24,8151	
5-8	*28,265	21,4905	
6-7	*43,9167	20,2615	
6-8	*-19,9883	16,0181	
7-8	*-63,905	21,4905	

В стовпці «Однорідні групи» вертикальними стовпцями зірок виділені можливі однорідні групи рівнів обробки величини ознаки. Маємо декілька наборів однорідних груп; до першого набору увійшли групи 6-та, 2-га, 3-тя та 4-та; до другого набору: 3-тя, 4-та, 8-ма; до третього: 4-та, 8-ма, 1-ша. Отже, для даних

груп об'єктів можна застосовувати однакові коефіцієнти в їх моделях, розробляти та використовувати однакові методи управління економіки. Існують групи, що в сукупності не мають однорідних, це - 7-ма та 5-та, вони не репрезентативні для виявлення загальних закономірностей досліджуваних об'єктів, але потребують детального вивчення, оскільки мають особливе в сукупності об'єктів.

Обґрунтування тенденції зміни величини ознаки в динаміці маємо зі змісту Рис. 7., що демонструє напрямок зміни – спадна одностороння, що свідчить про правильну стратегію системи управління на підприємствах по відношенню до даної ознаки їх розвитку.

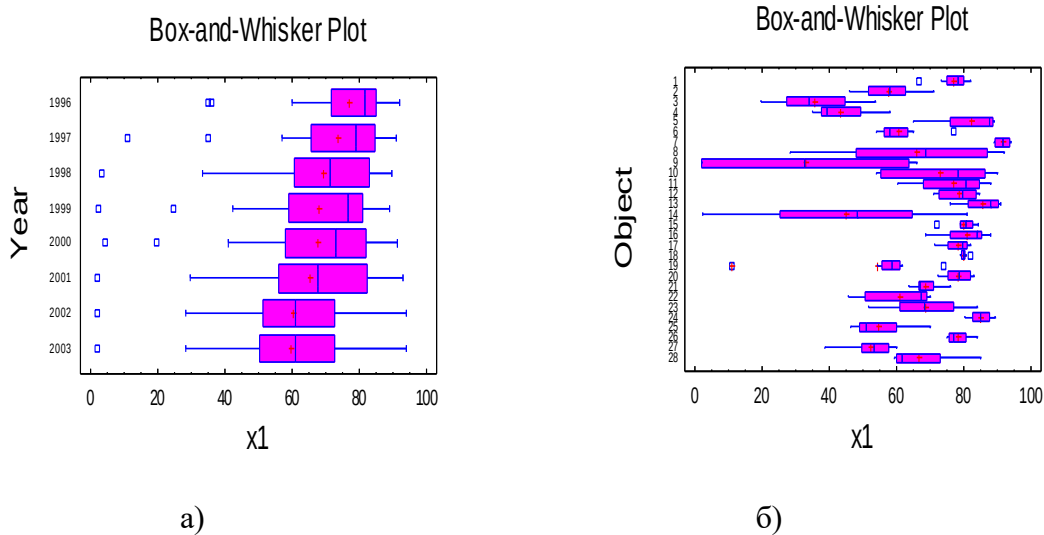
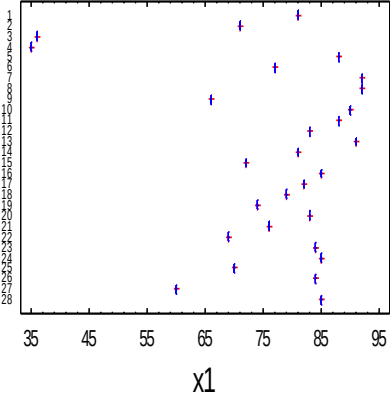
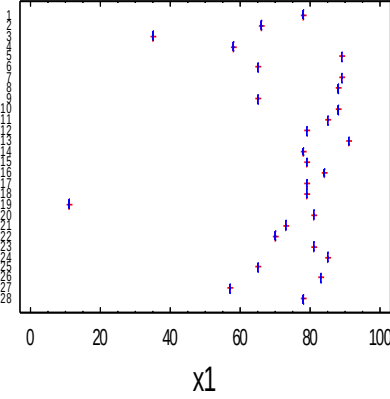
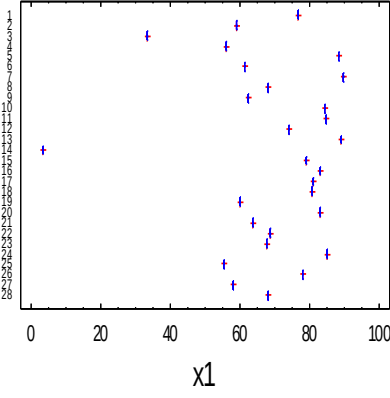
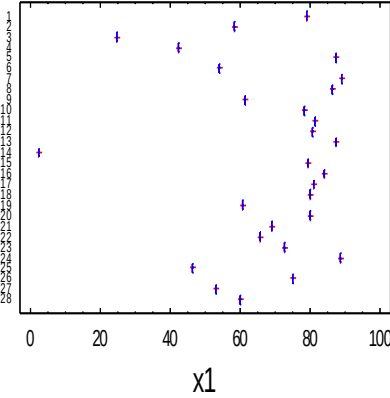
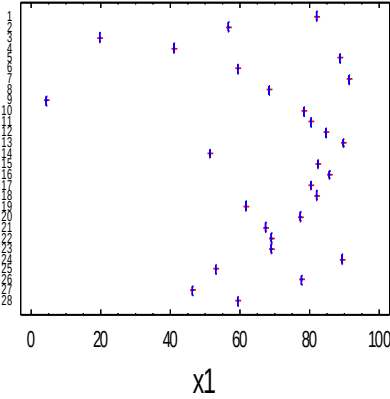
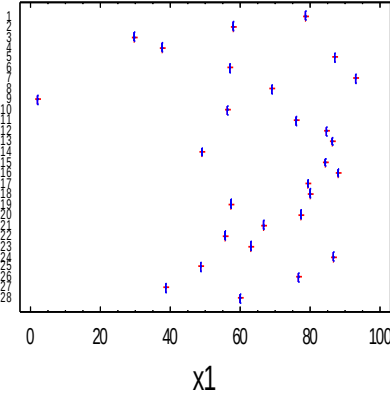
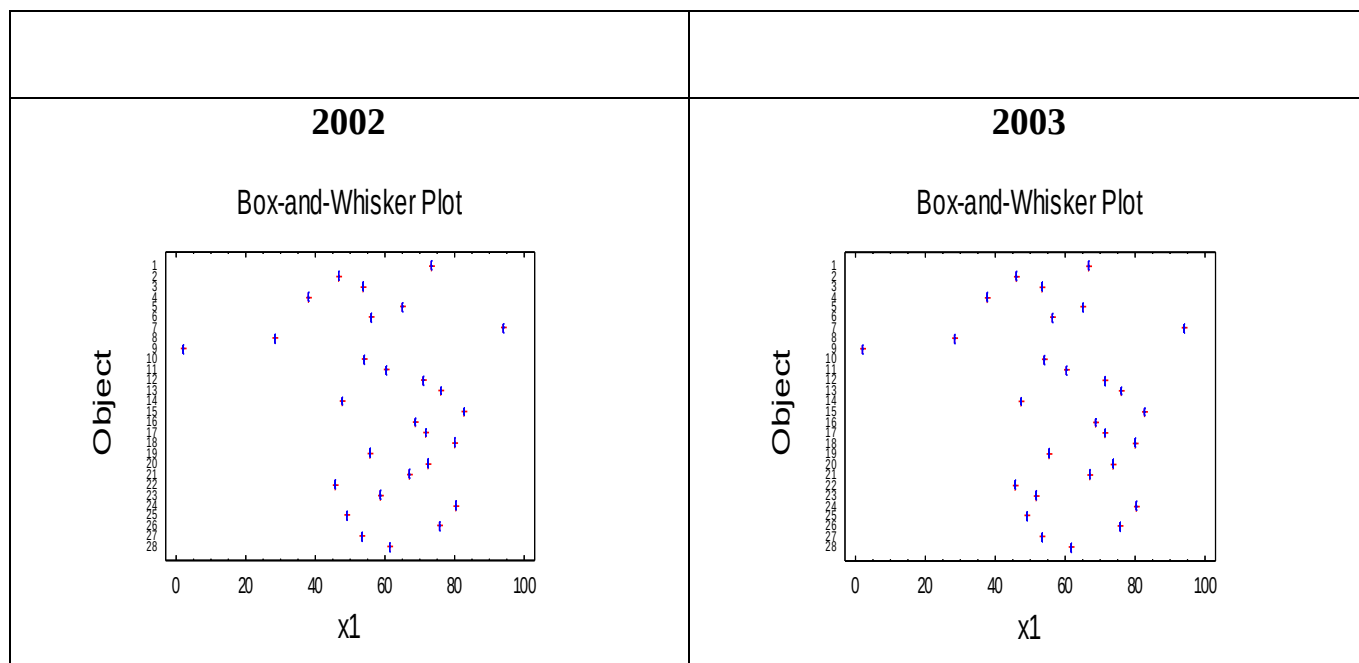


Рис. 7. Блокові діаграми значень величини ознаки в динаміці всієї сукупності підприємств (а) та на кожному підприємстві за весь період дослідження (б).

В Табл. 5. приведені блокові діаграми значень величини ознаки в статистиці та динаміці у всіх можливих зрізах.

Таблиця 5

1996 Box-and-Whisker Plot 	1997 Box-and-Whisker Plot 
1998 Box-and-Whisker Plot 	1999 Box-and-Whisker Plot 
2000 Box-and-Whisker Plot 	2001 Box-and-Whisker Plot 



В даній таблиці спостерігаємо розмитість значень величини, яка не дозволяє виявити яку-небудь закономірність в змінах значень ознаки об'єктів сукупності, що підтверджує об'єктивність групового підходу до визначення статистичних величин ознак об'єктів.

Описаний статистичний аналіз був виконаний для кожної з 45 ознак розвитку 28 промислових підприємств Харківського регіону. Основні економіко-статистичні висновки в першій групі підприємств зведені в Табл. 6. за визначеними загальними критеріями: закону розподілу, існування викидів, тенденції змін, що утворюють простір формування закономірностей статистичних величин (Рис. 8.).

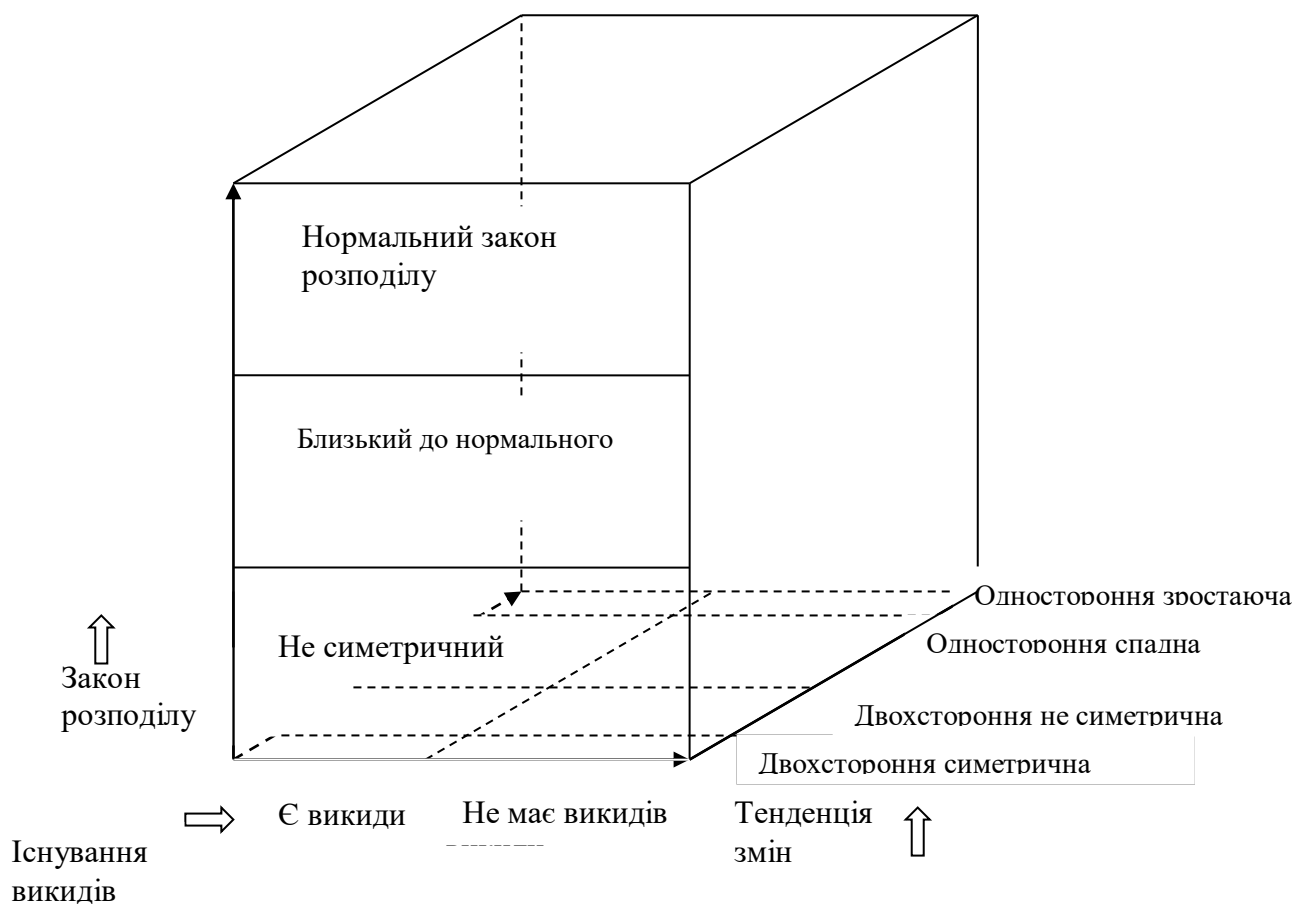


Рис. 8. Схема критеріїв описової статистики величин ознак

Таблиця 6

Закон розподілу	Тенденція змін					
	Одностороння спадна, симетрична	Одностороння спадна, несиметрична	Одностороння зростаюча, симетрична	Одностороння зростаюча, несиметрична	Двохстороння симетрична	Двохстороння несиметрична
Нормальний закон	x_1, x_{17}	x_{10}^*	x_{19}	x_9^*, x_{18}^*		x_{16}, x_{26}
Близький до нормального закону (майже симетричний)		x_6^*, x_7^*		x_5^*, x_8^*		x_{22}
Не симетричний закон				x_4, x_{20}^*, x_{21}^*	x_{13}^*, x_{23}	x_{14}, x_{25}^*, x_{24}
Закон розподілу абсолютних ознак			$x_{27} \div x_{48}$			

Ознаки, що відмічені символом «*» в Табл. 6. мають величини, тенденція змін

значень яких в даному періоді протилежна закономірній тенденції функціонування підприємства. Органам управління в регіоні слід ретельно проаналізувати причину усталених негативних тенденцій і стратегічні плани розробляти з урахуванням зміни напрямків розвитку, що склалися.

Таким чином, розглянута технологія проведення статистичного аналізу величин ознак є вимірюванням статистичної величини ознаки, вторинним чи похідним вимірюванням для проведення любого економічного аналізу, а також, першим необхідним етапом в побудові любой економіко-математичної моделі об'єкту в економіці. Для виявлених однорідних груп підприємств можна встановлювати одні нормативи, еталони, оскільки для них, як було визначено, спостерігаються однакові реперні значення величини на шкалі значень ознаки.