Тема 1. Робота з даними: наука чи мистецтво?

1. Мистецтво роботи з даними (с.28-38).
2. Статистичні дані. Структура статистичних даних, класифікація статистичних даних (с. 42-61).
3. Розподіл статистичних даних. Перетворення несиметричних статистичних даних у симетричні. Бімодальні розподіли статистичних даних. Викиди даних, їх види. Усунення викидів (с. 71-101).
4. Узагальнюючі показники набору статистичних даних. Типове значення набору статистичних даних (с. 117-151).
5. Мінливість даних, її статистичне оцінювання (с. 169-198).

Сильні сторони статистики:

1. Статистика допомагає вилучати інформацію з даних, розуміти незрозуміле, те, що не лежить на поверхні, і оцінювати якість цієї інформації.
2. Дає можливість зрозуміти ризики і випадковості та забезпечує оцінку правдоподібності отриманих можливих результатів.
3. Статистичні методи – це частина прийняття рішень, що слугує для них обґрунтуванням.
4. Статистика працює як з існуючими даними, так і з потенційними, які ще треба зібрати.
5. Індивідуальний підхід до роботи з даними: від загального до особистого.

Приклад. Види даних в менеджменті.

- Фінансова і статистична звітність.
- Інвестиційні звіти – курси та обсяги цінних паперів, процентні ставки.
- Урядові звіти – стан бюджету.
- Внутрішні поточні звіти – ціни та обсяги продажу.
- Маркетингові звіти – огляди ринків.
- Виробничі звіти – дані про якість продукції.
- Внутрішні дані – продуктивність праці.
- Рекламні звіти – витрати на рекламу і результати рекламної компанії.

Висновок:

- ***Статистика*** *– це одночасно і наука і мистецтво збирання і аналізу даних в усіх сферах людської діяльності.*
- Для статистика стовпчик цифр – це прихована інформація.

Чотири етапи статистичного аналізу:

1. Планування збору даних (планування вибіркового дослідження в маркетингу; планування експерименту в хімії).
2. Первинний аналіз даних (розвідувальний аналіз даних) – перевірка наявності очікуваних зв’язків і відповідність даних запланованим методам аналізу; виявлення в даних неочікуваної структури, що передбачає внесення корективів до плану аналізу.
3. Оцінювання – кількісне представлення невідомої величини.
4. Перевірка гіпотез – відповідність висуненого припущення дійсності. Метод дає можливість зробити вибір при неоднозначності ситуації.

Приклади невідомих величин:

- обсяг продажу в наступному кварталі;
- реакція на населення міста на новий продукт;
- зміна процентних ставок;
- вартість портфеля в наступному році;
- рівень браку;
- зміна продуктивності при зміні стратегії;
- вплив умов праці на продуктивність.

Приклади гіпотез:

- середні витрати мешканців в наступному місяці на купівлю продукту;
- нові ліки безпечні та ефективні;
- новий засіб більш ефективний;
- помилка у звіті менше за деяку величину;
- прогноз ситуації на ринку цінних паперів;
- прогнозна оцінка рівня виробничого браку.

Словник термінів (с.38):

Статистика – statistics

Планування дослідження – designing the study

Попереднє дослідження даних – exploring the data

Оцінювання невідомої величини – estimating an unknown quantity

Перевірка статистичних гіпотез – hypothesis testing

Імовірність – probability

Проект (с.41) :

Знайдіть в газеті, журналі або Інтернет статтю, де представлені результати опитування. Письмово опишіть, який з етапів статистичного аналізу був реалізований при обробці даних.

Набір статистичних даних

це результат експерименту (спостереження за об’єктами), що включає реєстрацію однієї і тієї ж інформації для кожного об’єкта (елементарні одиниці).

Існують чотири способи класифікації даних:

1. За кількістю інформації для кожного об’єкта:
2. За типом виміру (числа або категорії) для кожного об’єкта:
3. За можливістю часової упорядкованості: часові ряди (динаміка фондового індексу, щомісячні обсяги продажу) або дані про один часовий зріз.
4. За цілевою спрямованістю інформації: цільові дані (сбір первинних даних з використанням перинних або вторинних джерел інформації); нецільові (вторинні).

1. За кількістю інформації для кожного об’єкта:

- одновимірний – доходи окремих осіб, кількість дефектів вибірки з 50 виробів, прогноз процентної ставки 25 експертів на ступінь їхньої узгодженості; (*відповіді на питання*: типове значення, ступінь розбіжності об’єктів, наявність незвичних об’єктів);

- двовимірний – витрати на виробництво і кількість виробів на 10 підприємствах, щоденні котировки акцій, факт купівлі продукту і згадки про його рекламу (ефективність реклами) (*відповіді на питання*: чи існує зв'язок між змінними, наскільки вони тісно пов’язані, чи можна оцінити означення однієї, виходячи зі значення іншої, ф з якою надійністю, наявність незвичних об’єктів);

- багатовимірний набір даних – вплив типу стратегії (успішність стратегії) на результати роботи фірм (темпи зростання і тип обладнання, обсяги інвестицій, стиль керівництва), яка комбінація характеристик підвищує вартість дому (*відповіді на питання*: чи існує зв'язок між змінними, наскільки вони тісно пов’язані, чи можна оцінити означення однієї, виходячи зі значення іншої, ф з якою надійністю, наявність незвичних об’єктів).

2. За типом виміру (числа або категорії) для кожного об’єкта:

- кількісні дані (числа): дискретні (кількість укладених контрактів); неперервні (ціна за унцію золота, дохід на одну акцію). Не всі числа мають змістовну інтерпретацію;

- якісні дані: порядкові (посади, рейтинги, експертні оцінки; номінальні (назви фірм, регіони, продукти).

Чотири способи класифікації даних:

3. За можливістю часової упорядкованості: часові ряди (динаміка фондового індексу, щомісячні обсяги продажу) або дані про один часовий зріз.

4. За цілевою спрямованістю інформації: цільові дані (сбір первинних даних з використанням перинних або вторинних джерел інформації); нецільові (вторинні).

Приклад даних:

- Приклад первинних даних: інформація о продуктивності обладнання, дані соціологічного опитування.
- Приклад вторинних даних: економічні або демографічні показники, зібрані статистичною службою, дані зі спеціалізованих журналів дані, зібрані іншими компаніями, що займаються цім професійно (продаж телевізійних рейтингів).

Тренінг:

- 1. Знайти на сайті Державних статистичних служб різних країн світу дані про Індекс споживчих цін (Consumer Price Index) щомісячно (щоквартально) за 10 років і представити у форматі Excel.
- 2. Опишіть і класифікуйте базу даних (дод. 1). Для кожної змінної визначить можливі межі застосування операцій: арифметичні, розподільні, упорядкування, розрахунок структури (с.69).

Словник термінів (с.61) :

- Набір даних – data set
- Елементарні одиниці – elementary units
- Змінна – variable
- Одновимірний – univariate
- Двовимірний – bivariate
- Багатовимірний – multivariate
- Кількісна – quantitative
- Дискретна – discrete
- Безперервна – continuos
- Якісна – qualitative
- Порядкова або ординальне – ordinal
- Номінальна – nominal
- Часові ряди; – time series
- Про один часовий зріз – crosss-sectional
- Первинні дані – primary data
- Вторинні дані – secondary data

Самостійна робота (с.69) :

1. Знайдіть в Інтернет статтю з таблицею даних і надайте відповіді на питання щодо типу даних. Для кожної змінної визначить можливі межі застосування операцій: арифметичні, розподільні, упорядкування, розрахунок структури. На які питання можуть відповісти дані цієї таблиці?
2. Скористуйтесь даними звітності компанії, сформуйте таблицю і надайте відповіді на питання п. 1.
3. Знайдіть в Інтернет дані про інвестиції у компанію. Які дані доступні.

Розподіл дає можливість відповісти на такі запитання:

- Які значення є типовими для даного набору даних?
- Як різняться між собою ці значення?
- Чи присутня в наборі даних концентрація навколо якого-небудь значення?
- Який характер затухання коливань для крайніх розподілів даних, тобто який характер має та чи інша концентрація?
- Чи є значення в наборі даних які потребують окремої уваги – обробки)?
- Чи є типовим даний набір даних, чи має місце розшарування?

Чому це має значення?

- Річ у тім, що більшість кількісних методів аналізу, особливо, пов’язаних зі встановленням наявності зв’язку, потребують відповідності нормальному розподілу.
- В основі вивчення розподілу даних лежать **числові послідовності**, які характеризують деякі властивості об’єкта, який розглядається.
- Самим наочним представленням числових послідовностей є гістограми ***для відображення розподілу частот, а не даних***. Для даних використовують стовпчикові діаграми ***– не плутати***).

Приклад: Рівень ставки за позику під заставу нерухомості 45-ти кредиторів

Кредитор	Процентна ставка	Кредитор	Процентна ставка
Accubanc Mortgage Corp.	7,000	Intercontinental Mrtg	6,500
Alpine Mortgage Services	6,875	Federal Mortgage	6,500
American Investment Mrtg.	6,875	Merrill lynch Credit	7,250
Bay Mortgage	6,750	Millennium Mortgage	6,750
Capital Mortgage Corp.	6,870	Mortgage Broker Services	6,875
Castle Mortgage Corp.	7,250	Mortgage Network Inc.	6,875
Choice Mortgage	6,875	Mortgage Solutions	6,875
Citizen’s Mortgage Inc.	7,000	Nu-West Mortgage	6,875
City Mortgage	6,875	Mortgage	6,500
Community National Mrtg.	7,000	…	…

Гістограма розподілу кредиторів за рівнем процентних ставок під заставу нерухомості

Висновки:

1. Розмах значень перевищує 1 п.п.: від мінімуму 5,875% до максимуму – 7,25%.
2. Типове значення. Найчастіше зустрічаються ставки від 6,8% до 7,1%.
3. Розсіювання. Різниця в процентних ставках складає приблизно 0,5 п.п.: відстань між помірно високими стовпчиками.
4. Загальна конфігурація даних. Більшість організацій скупчені праворуч середини діапазону. Небагато організацій пропонують або зависокі або занизькі ставки. Пограничні значення прийнято відносити до правого стовпчику.
5. Характерні особливості. Жодна компанія не пропонує ставки в межах 6,9%-7,0%. Це викликано необхідністю кратності ставок 1/8: 6,5%; 6,625%; 6,75%; 6,875%; 7,0%.

Приклад: Стартова заробітна плата випускників за галузями економіки (річна)

Галузь	Заробітна плата, дол.	Галузь	Заробітна плата, дол.
Аерокосмічна	62500	Енергетика	63333
Автомобільна	50000	Індустрія розваг	55000
Банківська справа	58611	Фінансові послуги	60175
Комп'ютери	59280	Інвестиційна банківська справа	53500
Консалтинг	61625	Нерухомість	60250
Споживчі товари	59280	Роздрібни торгівля	93300
Електроніка	58016		

Гістограма розподілу галузей за початковим рівнем заробітної плати.

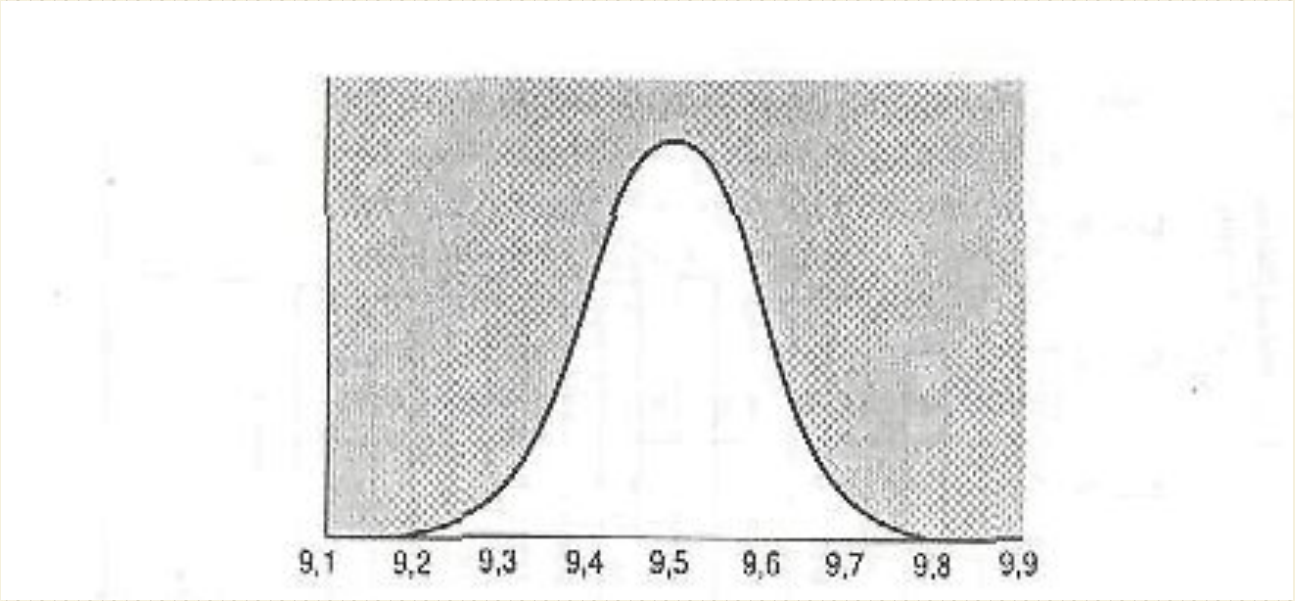
Гістограма розподілу кредиторів за рівнем процентних ставок під заставу нерухомості

Висновки:

- Кожен стовпчик гістограми може представляти більше однієї галузі. Стовпчики показують, які діапазони заробітної плати частіше, а які рідше зустрічаються у цьому наборі даних.
- Кожен стовпчик діаграми характеризує одну галузь промисловості.
- Стовпчикову діаграму краще використовувати у випадку необхідності відображення всіх значень з незначного набору даних, а гістограму для загального уявлення про набір даних.

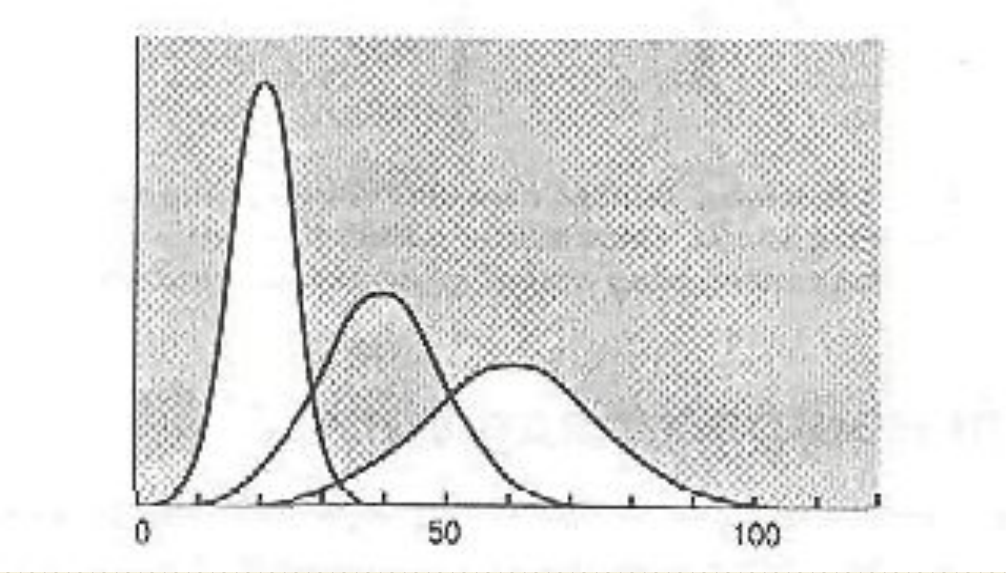
Нормальний розподіл

являє собой теоретичну гладку гістограму у формі колоколу без випадкових відхилень. Така крива представляє ідеальний набір даних, в якому більшість чисел сконцентровано в середній частині діапазону значень, а решта значення із загасанням, симетрично розташовані по обидві сторони від вершини колоколу. ***Такий ступінь гладкості не притаманний реальним даним.***



Нормальний розподіл

Фактично існує багато різних кривих нормального розподілу, форма яких нагадує симетричний колокол. Вони відрізняються розташуванням центру і масштабом (шириною колоколу). Щоб побудувати конкретну криву нормального розподілу, слід базову криву у формі колоколу перемістити по горизонталі в точку, де передбачається розмістити центр, а потім розтягнути (або стиснути). На рис. наведено кілька кривих нормального розподілу.



Чому нормальний розподіл відіграє таку важливу роль у статистиці?

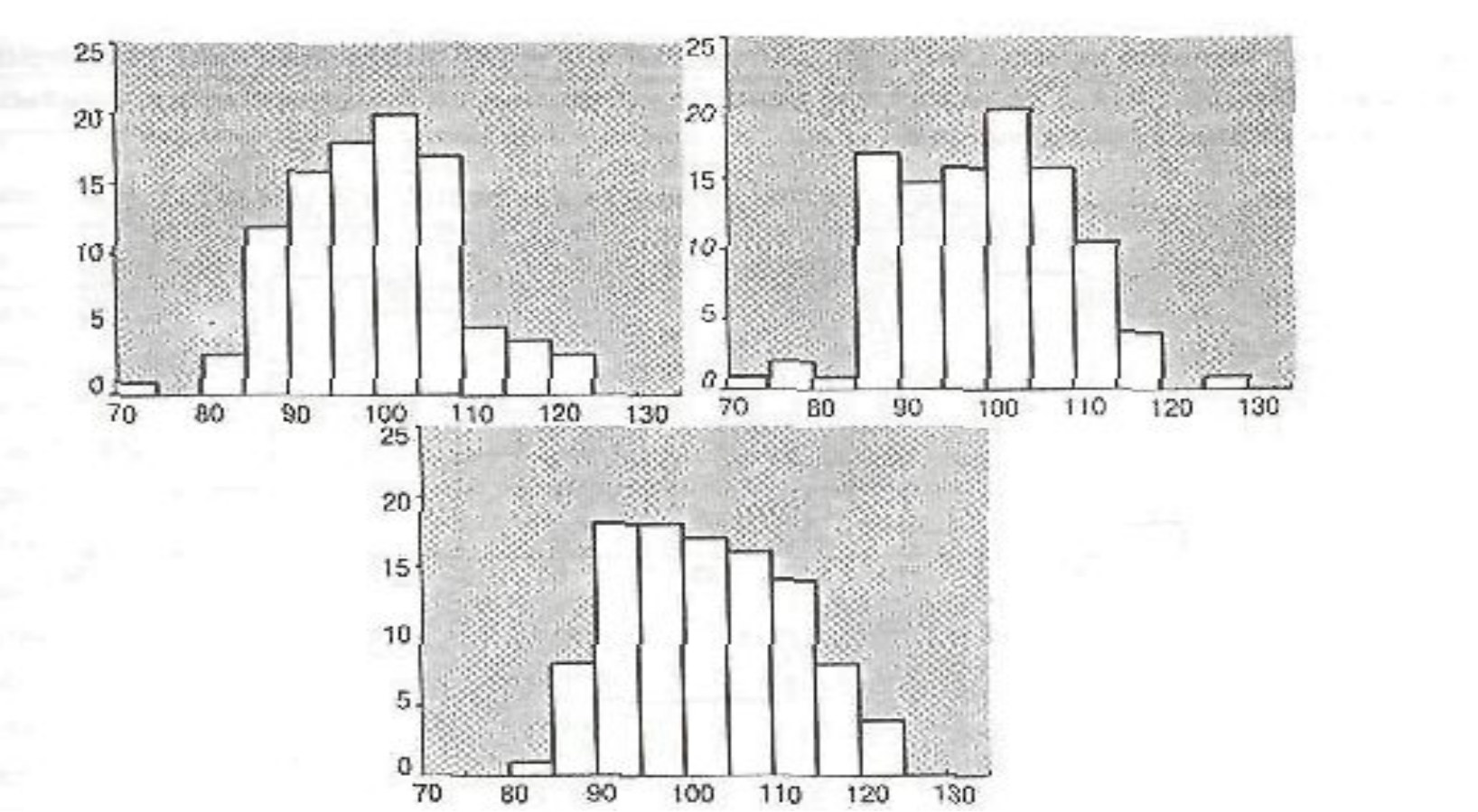
- Зазвичай в статистиці припускають, що розподіл даних приблизно відповідає нормальному.
- Формула кривої нормального розподілу має такий вигляд

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}}$$

де – центр, що визначає горизонтальне положення найвищої точки, – визначає ширину колоколу (мінливість або масштаб).

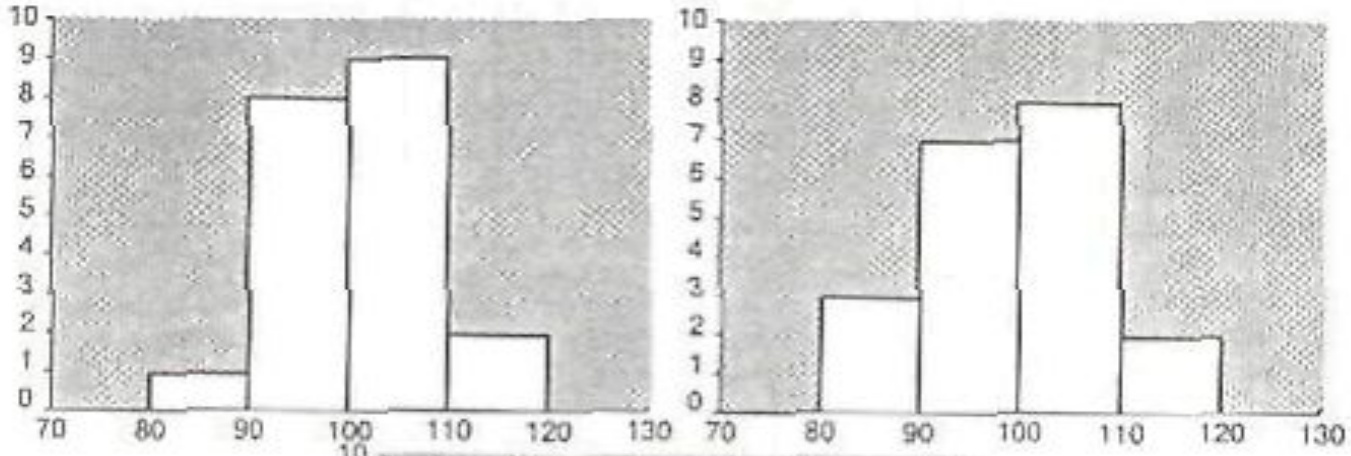
Чи є нормальний розподіл?

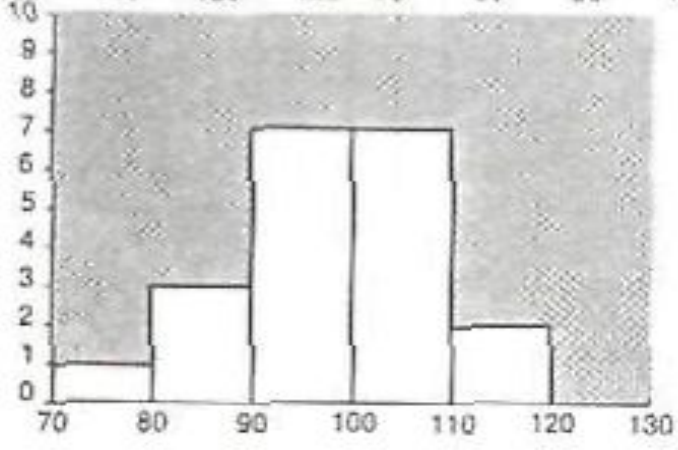
Гістограми для даних, витягнутих з нормально розподіленого набору. Обсяг кожної вибірки дорівнює 100. Порівняння цих трьох гістограм демонструє, який ступінь випадковості можна очікувати.



Чи є нормальний розподіл?

Гістограми для даних, витягнутих з нормально розподіленого набору. Обсяг кожної вибірки дорівнює 20. Порівняння цих трьох гістограм демонструє, який ступінь випадковості можна очікувати.





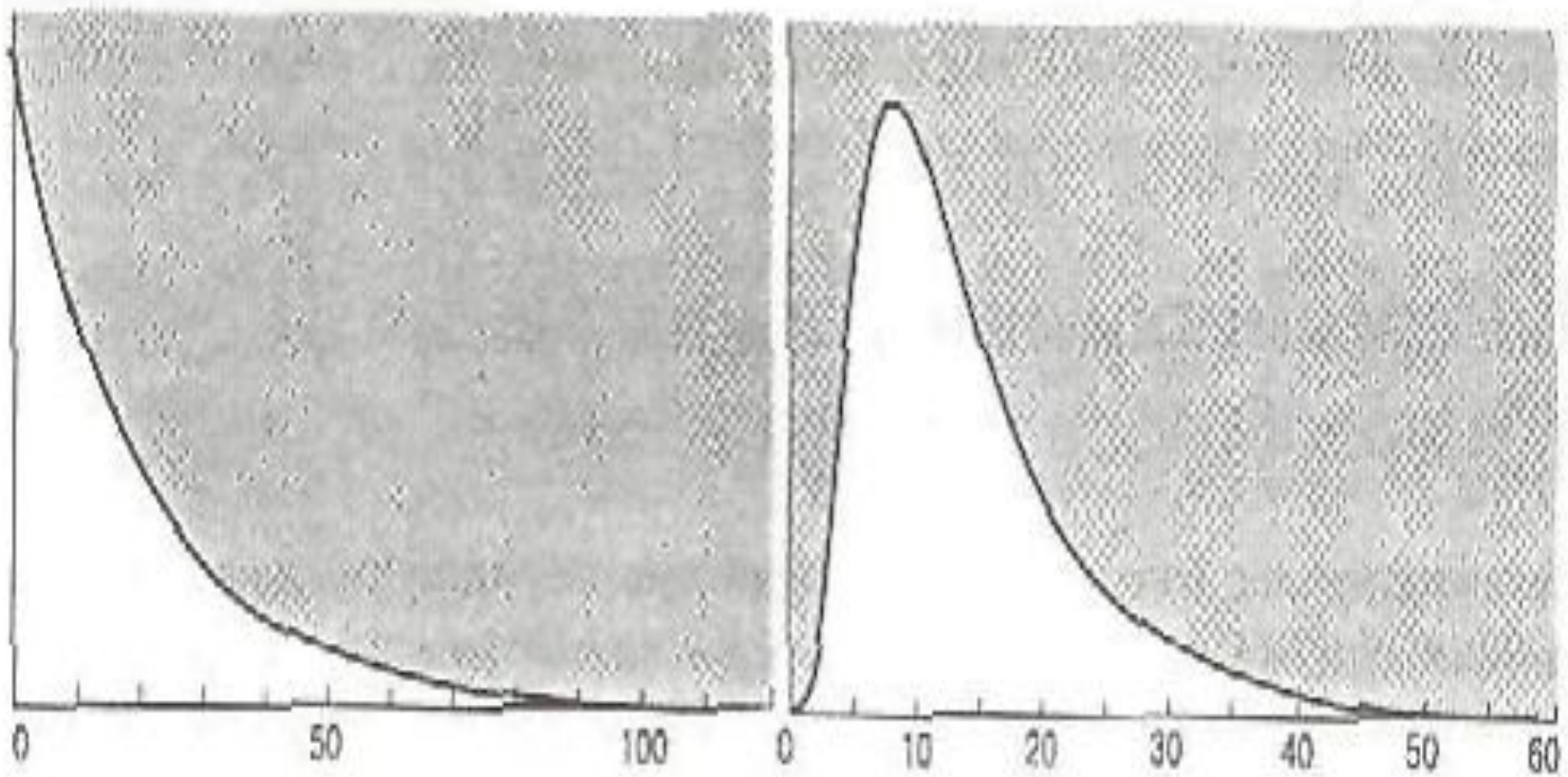
Несиметричний (скошений) розподіл

не є ані симетричним, ані нормальним, оскільки значення даних на одній стороні кривої затухають швидше, ніж па інший. У бізнесі часто можна зустріти асиметрію в наборі даних, які відображають величини, виражені додатними числами (**наприклад, обсяги продажів або розміри активів**).

Це пов'язано з тим, що такі дані не можуть приймати від’ємні значення (наявність обмеження з одного боку) і значення не обмежені зверху. В результаті на гістограмі багато значень даних сконцентровано навколо нуля, і кількість значень стає все меншим при русі по горизонтальній вісі

Несиметричний (скошений) розподіл

Згладжені ідеальні криві несиметричних розподілів. Реальні розподіли мають деякі відхилення від таких ідеальних кривих



Приклад: активи комерційних банків зі списку Fortune 500, млрд дол. (вибірка 50 банків)

Це яскравий приклад дуже несиметричного розподілу

Банк	Активи	Банк	Активи,
Chase Manhattan Corp.	366	Comerica	36
Citicorp	311	South Trust Corp.	31
National Bank Corp.	265	…	…
J. P. Morgan & Co	262	Compass Bancshares	13
Bank American Corp.	260	Synovus Financial Corp.	9
First Union Corp.	157	First National of	7
Bankers Trust New York Corp.	140	Providian Financial	4

Асиметричний розподіл:

самий високий стовпчик – це банки, які мають активи менше за 50 млрд дол. До 100 млрд дол. активів мають 41 банк.

Проблема з асиметрією:

- більшість найбільш поширених статистичних методів вимагають наявності принаймні приблизно нормального розподілу. Якщо ці методи застосовують до несиметричним даними, то отриманий результат може бути неточним або просто невірним. Навіть тоді, коли результати виходять в основному коректними, буде певна втрата ефективності аналізу, оскільки не забезпечується найкраще використання всієї інформації, що міститься в наборі даних.

Вихід за допомогою перетворення:

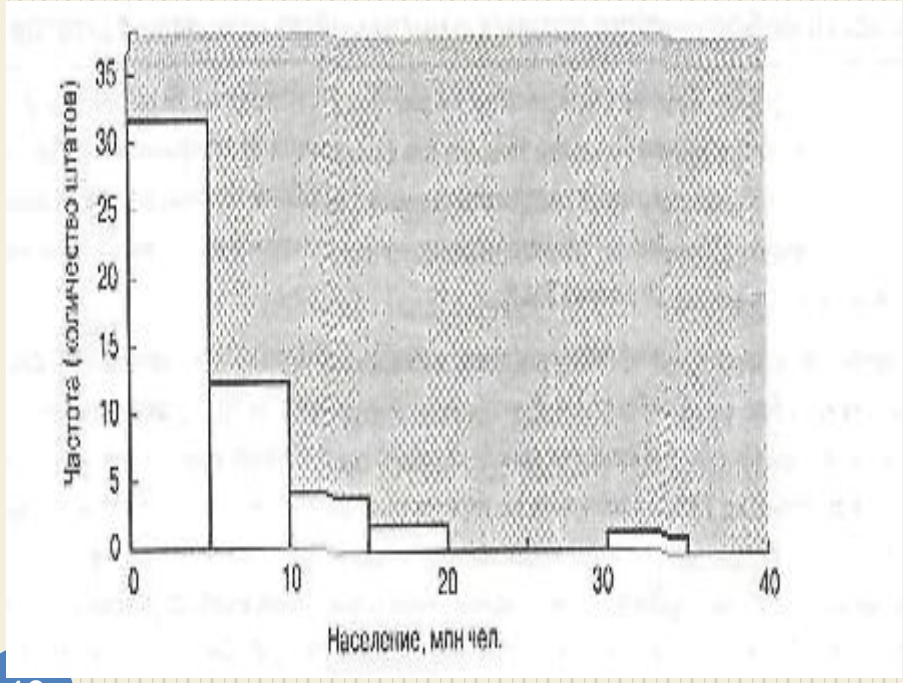
- Один із способів впоратися з проблемою асиметрії полягає у використанні такою **перетворення**, яке переводить несиметричний розподіл в більш симетричний. Перетворення полягає в заміні кожного значення набору даних іншим числом (наприклад, логарифмом цього значення) з метою спростити статистичний аналіз. Найбільш поширеним типом перетворення даних в бізнесі та економіці є **логарифмування**, яке можна використовувати тільки для додатних чисел.

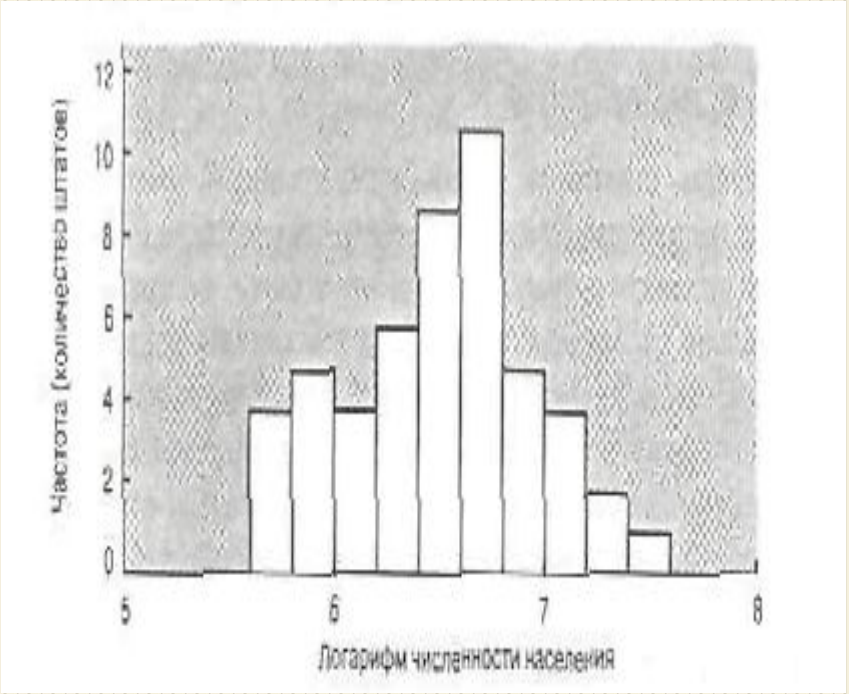
Вихід за допомогою перетворення:

- Логарифмування часто перетворює скошені (асиметричні) дані в симетричні, оскільки відбувається розтягування шкали навколо нуля, що, у свою чергу, призводить до розподілу малих значень, згрупованих разом.
- У той же час логарифмування збирає разом великі значення, які розподілені на правому боці шкали.
- Найчастіше використовують десятковий та натуральний логарифми.

Гістограма чисельності населення штатів (фактичні дані):

Порівнюючи гістограму чисельності населення зліва і справа можна відмітити, що в результаті логарифмування асиметрія зникає але не повністю. Проте можна спостерігати, що крива не ідеально симетрична.





Логарифмічну шкалу можна інтерпретувати скоріше як мультиплікативну або процентну, ніж як адитивну.

Використання логарифмічної шкапи призводить до того, що відстань по горизонталі 0,2 (ширина одного стовпчика) відповідає збільшенню (при русі зліва на право) населення на 58% (оскільки $10^{0,2} = 1,58$, що на 58% більше).

Відстань по горизонтальній вісі у п'ять стовпчиків (з 6 до 7) відповідає 10-ти кратному збільшенню чисельності населення штату (оскільки $10^1 = 10$).

На первісній шкалі, що відбиває фактичну чисельність населення штату, важко проводити порівняй у відсотках.

При русі зліва направо перехід до кожного нового стовпчика означає збільшення населення на 5 мільйонів – на лівій стороні ця різниця в процентах набагато більша ніж на правій.

Висновок:

- *Таким чином, логарифмування стягує разом дуже великі числа, зменшуючи різницю між ними та іншими значеннями в наборі даних (замість різниці в 100 млн разів отримуємо різницю у 8 одиниць) і розтягують маленькі значення, збільшуючи різницю між ними й іншими значеннями*

Порівняльна таблиця фактичних значень й їхніх логарифмів

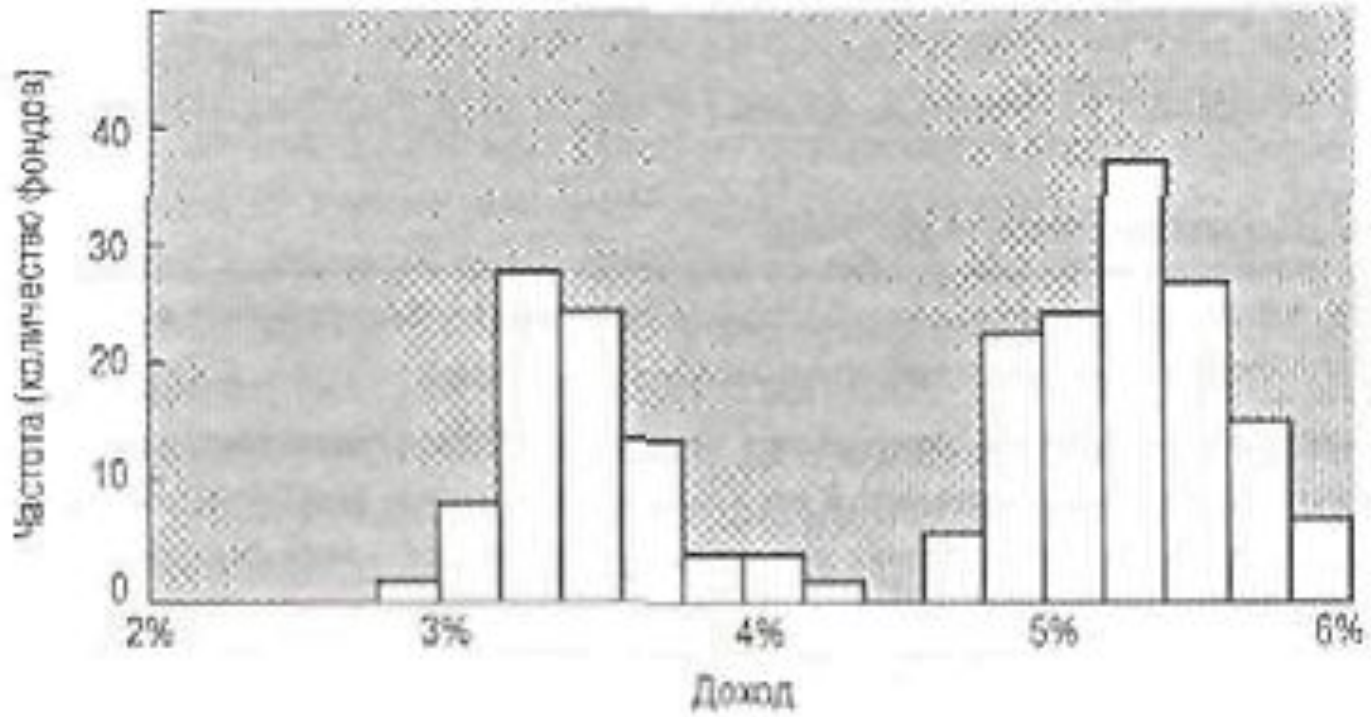
Число	Десятковий логарифм	Натуральний логарифм
0,001	-3	-6,9
0,01	-2	-4,6
0,1	-1	-2,3
1	0	0,0
2	0,301	0,7
5	0,699	1,6
10	1	2,3
100	2	4,6
10000	4	9,2
100000	5	11,5
1000000000	8	18,4

Бімодальні розподіли

- Важливо вміти визначати, коли набір даних складається з двох або більш чітко розрізняються між собою груп, з метою аналізу цих груп окремо. На гістограмі такій ситуації відповідає розрив між двома сусідніми групами стовпчиків. Якщо на гістограмі чітко видні дві окремі групи, то це говорить, про бімодальний розподіл даних. **Бімодальне розподіл** – це розподіл, якому притаманні дві моди або два різних кластера (блоку) даних.
- Наявність бімодального розподілу може свідчити про те, що ситуація є складнішою, а тому потребує більш серйозної уваги. Щонайменше, слід виявити причини наявності двох груп. Можливо, інтерес представляє лише одна група, тому іншу групу можна виключити з розгляду. Можливо і те, що увагу слід приділити двом групам але з уточненням розбіжностей які їм притаманні.

Гістограма розподілу взаємних фондів за доходами на валютному ринку.

Це бімодальний розподіл з двома чітко виділеними групами, що не можна пояснити тільки випадковістю.



Пояснення

Річ у тім, що початковий набір даних містить заголовок «Вільні від податку», який відокремлює у списку звичайні оподатковувані фонди від тих, що вкладають кошти тільки у неоподатковані цінні папери. Оскільки для неоподаткованих фондів не нараховується податок від отриманого відсотка, то ефективний дохід (з поправкою на податок) вище, ніж те значення, яке зазвичай вказують. Таким чином, група з більше низькими доходами в лівій частині гістограми включає фонди, звільнені від сплати податку.

Якщо треба узагальнити поточні ринкові процентні ставки, то доходи фондів, звільнених від податку, необхідно **попередньо обробити**.

Можна не розглядати неоподатковані фонди і проаналізувати тільки доходи звичайних фондів.

З іншого боку, можна попередньо відкоригувати неоподатковані доходи щоб привести їх у відповідність з іншими, а потім провести аналіз.

Середня вартість одного дня перебування у місцевій лікарні, дол.

Штат	Вартість	Штат	Вартість	Штат	Вартість
Alabama	729	Kentucky	674	North Dakota	434
Alaska	1116	Louisiana	836	Ohio	875
Arizona	1051	Maine	674	Oklahoma	740
Arkansas	633	Maryland	806	Oregon	1011
California	1134	Massachusetts	937	Pennsylvania	793
Colorado	904	Michigan	847	Rhode Island	601
Connecticut	1012	Minnesota	618	South Carolina	762
Delaware	920	Mississippi	516	South Dakota	457
Dist. of Colombia	1124	Missouri	732	Tennessee	796
Florida	886	Montana	474	Teas	933
Georgia	721	Nebraska	600	Utah	1036
Hawaii	761	Nevada	952	Vermont	726
Idano	618	New Hampshire	776	Virginia	774
Illinois	849	New Jersey	737	Washington	974
Indiana	822	New Mexico	950	West Virginia	655

Гістограма розподілу штатів за вартістю одного дня перебування у місцевій лікарні, дол.

Це майже нормальний розподіл.

Гістограма розподілу штатів за вартістю одного дня перебування у місцевій лікарні, дол.

Складається враження (невірне), що у наборі даних присутні дві або навіть три групи.

Але це випадковість і не є дійсною бімодальністю.

Викиди – значення, що сильно відхиляються

Існують два види викидів значень:

- помилки;
- коректні значення, що відрізняються від загальних даних.

Вирішення проблеми:

- виключення викидів;
- проведення двох аналізів: з викидами і без них.

Немає вичерпного вирішення цієї проблеми.

Дві проблеми з викидами:

1. Труднощі з інтерпретацією структури у випадку, коли одне значення домінує і привертає до себе підвищену увагу.
2. Як і у випадку асиметрії, більшість сучасних статистичних методів не можна використовувати для аналізу тих даних, розподіл яких сильно відрізняється від нормального. Нормальний розподіл є симетричним і зазвичай не містить викидів.

Приклади викидів

В наборі даних щодо доходів грошового ринку може з'явитися кілька значень доходів фондів, які неоподатковуються. Якщо мета дослідження полягає в аналізі ринкової ситуації для звичайних фондів, то ці викиди краще виключити із загальної картини.

Припустимо, що компанія оцінює новий фармацевтичний продукт. В одному з них лаборант чхнув на зразок перед його аналізом. Якщо ви не вивчаєте нещасні випадки з лабораторними матеріалами, то цей зразок годі й аналізувати.

Приклади викидів

- За повідомленням The Wall Street Journal, чистий дохід за другий квартал найбільших компаній США зріс на 27% за результатами аналізу даних про 677 відкритих акціонерних торгових кампаній. Однак у даних є викиди значень: в результаті відмежування від компанії U.S. West дохід компанії MediaOne склав у другому кварталі 24,5 млрд дол. Якщо це значення виключити з аналізу, то збільшення чистого, доходу фактично знизиться до 1,5%.
- Майже така ж сама ситуація спостерігалася в попередньому кварталі, коли чистий дохід зріс на 20% завдяки продажам компанії Ford Motors. Якщо виключити цей викид, то замість сильного зростання отримаємо просто зростання але 2,5%.

Висновок:

- *Таким чином, наявність викиду дає хибне уявлення про реальне зростання компаній. Може скластися невірна думка про те, що більшість компаній демонструють сильне економічне зростання.*

На прикладі динаміки витрат на телевізійну рекламу провідних компаній

простежимо як наявність викидів впливає на симетричність розподілу

Рекламодатель	Зміна витрат на рекламу, %	Рекламодатель	Зміна витрат на рекламу, %
Procter Gamble	43,2	Warner-Lambert	-22,7
Phillip Morris	27,5	AT&T	73,5
Kellogg	77,9	Grand Metropolitan	14,0
Time Warner	201,0	Johnson & Johnson	16,5
Unilever	16,7	National Education	217,3
Hasbro	54,5	Nestle	31,4
Mattel	47,7	Hershey	42,4
American Home Products	104,4	Regal Communications	2353,7
General Motors	65,7	McDonald's	28,5
Wrigley	66,8	Sara Lee	16,4
Mars	33,3	Himmel Group	684,0
RJR Nabisco	65,9	Bayer Group	12,7
Sears Roebuck	44,7		

Гістограма розподілу процентного зростання витрат на рекламу 25 компаній.

В правій частині присутній викид компанії Regal Communications, що зводить практично всі компанії в один стовпчик.

Гістограма розподілу процентного зростання витрат на рекламу 24 (23) компаній

Після усунення викиду (компанії Regal Communications) спостерігається ще один викид (компанія Himmel Group) приховує деталі більшої частини набору даних.

Решта компаній дають типове збільшення витрат від 0 до 75% (можливо, трохи більше чи менше, за винятком двох компаній з високим, близько 200%, зростанням витрат

Гістограма розподілу процентного зростання витрат на рекламу 21 компанії

Висновки

- Дані цього аналізу свідчать про те, що витрати на рекламу сильно змінюються щодня.
- Крупні рекламодавці не мають постійної стійкої стратегії, яка лише трохи коригується щороку.
- Більшість з 25 провідних рекламодавців для телебачення, мабуть, виявилися в цьому списку завдяки значному збільшенню своїх витрат на рекламу порівняно з попереднім роком.

Самостійно вивчить метод побудови гістограми «Стовбур і листя»! (с. 97-98) і опрацювати у ППП Statistica

Словник термінів (с. 101):

- Послідовність чисел – list of numbers
- Числова вісь – number line
- Гістограма – histogram
- Нормальний розподіл – normal distribution
- Несиметричний скошений розподіл – skewed distribution
- Перетворення – transformation
- Логарифм – logarithm
- Бімодальний розподіл – bimodal distribution
- Викид – outlier
- “Стовбур і листя” – steam-and-leaf

Самостійна робота з використанням бази даних (с. 114):

За даними даних, наведеними в дод. А виконайте завдання.

1. Для заробітної плати службовців:

а) Побудуйте гістограму.

б) Опишіть форму розподілу.

в) Узагальніть інформацію про розподіл, вказавши також розміри найменшої та найбільшої заробітної плати.

2. Для віку службовців:

а) Побудуйте гістограму.

б) Опишіть форму розподілу.

в) Узагальніть інформацію про розподіл.

3. Для стажу роботи службовців:

а) Побудуйте гістограму.

б) Опишіть форму розподілу.

в) Узагальніть інформацію про розподіл.

4. Для заробітної плати службовців різної статі:

а) Побудуйте гістограму тільки для чоловіків.

б) Побудуйте гістограму для жінок, використовуючи той же масштаб, що і в п. "а", з метою порівняння заробітної плати чоловіків і жінок.

в) Порівняйте два розподіли заробітної плати і напишіть резюме, вказавши на відмінності в оплаті праці чоловіків і жінок.

Проекти (с. 115):

- Побудуйте гістограму для кожного з трьох наборів даних, що мають відношення до ваших інтересів в бізнесі (економіці). Підберіть дані, що Вас цікавлять з Internet або зі звітів компаній. Кожен набір даних повинен містити не менше 15 чисел. Для кожного набору даних напишіть сторіночку коментаря (включаючи гістограму), указів наступне:
- а) Яка форма розподілу?
- б) Чи є викиди значень? Що потрібно зробити, якщо вони є?
- в) Узагальніть інформацію про розподіл,
- г) Про що дізналися, вивчив гістограму?

Ситуаційний аналіз: необхідність контролю виробничих втрат (с. 115)

- "Цей Оуен викидає наші гроші на вітер! – Голосно заявив Біллінгс на нараді. – У мене є докази. Ось гістограма вартості використання сировини. Чітко видно дві групи, причому Оуен витрачає на сировину на кілька сотен доларів більше, ніж Парсел ".
- Ви ведете нараду, і вона проходить більш емоційно, ніж хотілося б. Щоб перевести збори в більш спокійне русло, ви чемно намагаєтеся пом'якшити обговорення і досконально обдумати рішення.
- Ви знаєте, як, втім, і більшість інших, що Оуен має репутацію безтурботного людини. Однак ви ніколи не ставили цей порок на перше місце, і вам хотілося б відкласти оцінку Оуена якраз тому, що інші заздрісно підкидають таку пропозицію, й тому, що Оуена поважають за компетентність і працьовитість. Вам також відомо, що Біллінгс і Парсел – хороші приятелі. У цьому, звичайно, немає нічого поганого, але все ж краще познайомитися з усією доступною інформацією перед тим, як робити остаточний висновок.
- Після наради ви просите Біллінгса прислати вам електронною поштою копію даних. Але він надсилає вам тільки перші дві колонки (витрати на матеріали), (табл. 1.7), і вони зам вже знайомі. У вашому комп'ютері вже є звіт, що включає всі три колонки, наведені нижче. Тепер ви готові витратити час на підготовку наради, щоб провести її на наступному тижні.

Ситуаційний аналіз: необхідність контролю виробничих втрат (42 спостереження)

Вартість сировини, дол.	Відповідальний менеджер	Вартість продукції, дол.	Вартість сировини, дол.	Відповідальний менеджер	Вартість продукції, дол.
1459	Оуен	4869	1434	Оуен	4589
1502	Оуен	4806	1127	Парсел	3606
1492	Оуен	4774	1457	Оуен	4662
1120	Парсел	3558	1109	Парсел	3549
1433	Оуен	4746	1236	Парсел	3955
1136	Парсел	3635	1188	Парсел	3802
1123	Парсел	3594	1512	Оуен	4838
1542	Оуен	4934	1131	Парсел	3619
1434	Оуен	4749	1108	Парсел	3546
1379	Оуен	4413	1135	Парсел	3632
1406	Оуен	4499	1416	Оуен	4531
1487	Оуен	4756	1170	Парсел	3744
1138	Парсел	3642	1417	Оуен	4534
1529	Оуен	4893	1381	Оуен	4419
1142	Парсел	3654	1248	Парсел	3994
1127	Парсел	3605	1171	Парсел	3747
1457	Оуен	4662	1471	Оуен	4707
1379	Оуен	4733	1142	Парсел	3654
1407	Оуен	4502	1161	Парсел	3715
1105	Парсел	3536	1135	Парсел	3632
1126	Парсел	3603	1500	Оуен	4800

Ситуаційний аналіз: необхідність контролю виробничих втрат (с. 116)

Питання для обговорення:

1. Чи є розподіл вартості сировини дійсно бімодальний? Або ці дані можна розглядати як одну нормально розподілену групу значень?
2. Чи узгоджуються гістограми, побудовані Для Оуена і Парсела окремо, із твердженням Біллінгса про те, що Оуен витрачає більше?
3. Чи потрібно погодитися з Біллінгсом на наступній нараді? Обґрунтуйте вашу відповідь за допомогою ретельного аналізу наявних даних.

Узагальнюючі показники набору статистичних даних. Типове значення набору статистичних даних

- У складних ситуаціях один з найефективніших способів "побачити всю картину" полягає в узагальненні, тобто використанні одного або декількох відібраних або розрахованих значень для характеристики набору даних.

Докладне вивчення кожного окремого випадку само по собі не є статистичною діяльністю, але виявлення та ідентифікація особливостей, які характерні для розглянутих випадків в цілому є статистичною діяльністю.

- Одна з цілей статистики полягає в тому, щоб звести набір даних до одного числа (або декількох), які виражають найбільш фундаментальні властивості даних.

Узагальнюючі показники набору статистичних даних. Типове значення набору статистичних даних

- **Середнє, медіана і мода** – це різні способи вибору одного числа, яке краще всього описує всі числа в наборі даних. Такий представлений одним числом показник називається типовим значенням або центром (також використовують поняття міра центральної тенденції).
- **Перцентиль (процентиль)** – узагальнює інформацію про ранги, характеризуючи значення, що досягається заданими відсотком загальної кількості даних, після того, як дані упорядковуються (ранжуються) за зростанням.
- **Стандартне відхилення** – характеризує розбіжність між значеннями в наборі даних. Це також називають розкидом або мінливістю.

Як бути, якщо набір даних містить окремі значення, які неадекватно описуються цими показниками?

- Такі викиди можна просто описати окремо. Таким чином, можна охарактеризувати великий набір даних, узагальнив основні властивості більшості його елементів і потім створивши список винятків.
- *Це дає можливість досягти статистичної мети ефективного опису великого набору даних з урахуванням особливої природи окремих елементів.*

Приклад. Аналіз витрат

Фірму цікавить скільки в цілому витрачають на медичні товари мешканці міста. Аналіз вибірки з 300 осіб показав, що в минулому місяці кожен з них витратив приблизно **6,58 дол.** Природно, хтось витратив більше, а хтось менше. Замість того щоб працювати з усіма 300 числами, ми використовуємо середнє, щоб визначити типове значення індивідуальних витрат кожного споживача. *Що особливо важливо, помноживши середнє значення витрат на чисельність населення міста, отримаємо оцінку сумарних витрат на медичні товари мешканців усього міста:*

$$6{,}58 \cdot 503000 = 3309740 \,(\text{дол.})$$

Цей прогноз сумарних продажів на рівні 3,3 млн. дол. є прийнятним і, ймовірно, корисним. Однак це значення не є точним (в тому сенсі, що воно не відображає точну суму витрат). При вивченні довірчих інтервалів далі буде враховано статистичну похибку, яка виникає при поширенні отриманого для вибірки в 300 осіб результату на все населення міста.

В чому неточність цього поширення?

Приклад. Скільки є бракованих деталей?

Кожна партія виробів компанії Globular Ball Bearing Company містить 1000 виробів. Для проведення контролю якості виробів з вироблених за день 253 партій було взято випадковим відбором 10 партій.

Кількість бракованих виробів в кожній партії: **3, 4, 2, 5, 0, 7, 14, 7, 4, 1.**

Середнє для цього набору даних: **5,1 виробу**.

Іншими слонами, рівень браку 0,51%.

Якщо поширити отримане значення середньої на всі випущені за день 253 партії, то можна **очікувати 1290 одиниць браку**.

Зважене середнє

(використовують також термін середньозважене). Схоже на середнє, але дає можливість врахувати різну важливість (значимість), або "вагу", кожному елементу даних.

Зважене середнє гнучко визначає систему важливості окремих елементів даних в тому випадку, коли їх не можна розглядати як рівноцінні.

Якщо у фірми три заводи, при аналізі пенсійних витрат не можна використовувати просте середнє типових розмірів пенсійних витрат на кожному з трьох заводів як типове значення загальних пенсійних витрат, особливо, якщо заводи відрізняються за розміром. *Якщо чисельність службовців на одному в два рази перевищує чисельність службовців на іншому, то його слід врахувати з подвійною вагою. Як правило, ваги – це додатні числі сума яких дорівнює 1.*

Приклад. Розрахунок середнього балу

- **Середній бал (GPA – grade point average)** результатів навчання в університеті обчислюється як зважене середнє. Це пов'язано з тим, що деякі курси оцінюються більшою кількістю очок і, отже, є більш важливими порівняно іншими. Цілком розумно, якщо курсом, який оцінюється в два рази більше, ніж інший, присвоюється удвічі більшу вагу і середній бал це відображає.
- У різних університетах використовують різні системи оцінок. Припустимо, що система оцінок включать оцінки від 2,0 (незалік) до 5,0 (відмінно) і в кінці семестру картка з оцінками має такий вигляд.

Курс (Course)	Очки (Credits)	Оцінка (Grade)	Вага	Зважені оцінки
Статистика (Statistics)	5	4,7	5/16=0,3125	4,7·0,3125
Економіка (Economics)	5	4,3	0,3125	4,3·0,3125
Маркетинг (Marketing)	4	4,5	0,2500	4,5·0,25
Спецкурс (Track)	2	3,8	0,1250	3,8·0,125
Разом	16	x	1,0000	4,41

Приклад. Вартість капіталу фірми

Вартість капіталу фірми обчислюють як зважене середнє. Суть в тому, що фірма збільшує свої грошові кошти за допомогою продажу різних цінних паперів: акцій, облігацій, векселів тощо. Оскільки кожен вид цінного паперу має свою власну доходність (вартість капіталу), корисно об'єднати і узагальнити різні рівні прибутковості в одне значення, яке являє собою сукупну вартість капіталу для цього набору цінних.

Вартість капіталу фірми є простою середньозваженою вартістю капіталу по кожному цінному паперу (доходність або процентна ставка), причому вага визначається у відповідності з повною ринковою вартістю цих цінних паперів.

Розглянемо ситуацію для Leveraged Industries, Inc., гіпотетичної фірми з безліччю боргових зобов'язань, які утворилися внаслідок нещодавньої діяльності, пов’язаною із злиттям і придбанням.

Приклад. Вартість капіталу фірми

Вид цінних паперів	Ринкова вартість, тис. дол.	Доходність, %	Вага	Зважені оцінки
Звичайні акції	100	18,5	0,220	18,5·0,22
Привілейовані акції	15	14,9	0,033	14,9·0,033
Облігації (ставка 9%)	225	11,2	0,495	11,2·0,495
Облігації (ставка 8,5%)	115	11,2	0,253	11,2·0,253
Разом	455	x	1,001	12,94

Середньозважену вартість акціонерного капіталу можна пояснити у такий спосіб: якщо компанія вирішить збільшити додатковий каптал без зміні власної стратегії (типу та ризику проектів) і зберегти той же набір цінних паперів, то необхідно буде сплачувати на рік 12,9%, або 129 дол. на 1000 дол. Ці 129 дол. будуть виплачені по різних типах цінних паперів відповідно до їхньої ваги.

Приклад. Аналіз витрат (продовження)

Розглянемо вибірку 300 мешканців міста з точки зору витрат на медичні товари. Якщо процент людей до 18 років складає 21,7% не відповідає фактичній частці в генеральній сукупності – 25,8%, а витрати для кожної групи людей відрізняються: до 18 років – 4,86 дол., а старше 18 років – 7,06 дол., то при розрахунку середніх витрат будемо враховувати фактичний розподіл у генеральній сукупності. В результаті середні витрати складуть 6,49 дол., а не 6,58, що змінить загальне уявлення про сукупні витрати у генеральній сукупності.

Самостійно розрахуйте загальні витрати населення міста на медичні товари з урахуванням нової інформації.

Приклад. Обвал фондового ринку 19.10.1987 р.

- Обвал фондового ринку 1987 став екстраординарною подією, тоді ринок втратив за один день 70% вартості.
- Розглянемо відсоток втрат вартості акцій 29 компаній зі списку Dow Industrial в проміжок часу між закриттям торгів у п'ятницю 16 жовтня і відкриттям торгів в понеділок 19 жовтня 1987 р. в день краху. З таблиці можна побачити, що навіть при відкритті торгів акції вже втратили значну частину своєї вартості.

Приклад. Обвал фондового ринку 19.10.1987 р.

З таблиці можна побачити, що навіть при відкритті торгів акції вже втратили значну частину своєї вартості.

Фірма	Зміна вартості, %	Фірма	Зміна вартості, %
Union Carbide	-4,1	Primerica	-6,6
USK	-5,1	Navistar	-2,1
Steel	-4,5	General Electric	-17,2
AT&T	-5.4	Westinghouse	-15,7
Boeing	-4,0	Alcoa	-8,9
International Paper	-11,6	Kodak	-15,7
Chevron	-4,0	Texaco	-12,3
Woolworth	-3,0	IBM	-9,6
United Technologies	-4,4	Merck	-12,0
Allied-Signal	-9,3	Phillip Morris	-12,4
General Motors	-0,9	Du Pont	-8,6
Procter & Gamble	-3,5	Sears Roebuck	-11,4
Coca-Cola	-10,5	Goodyear Tire	-10,9
McDonald's	-7,2	Exxon	-8,6
Mining	-8,9		

Приклад. Обвал фондового ринку 19.10.1987 р.

- **Гістограма розподілу процентного падіння вартості 29 промислових компаній зі списку Dow Industrial 19 жовтня 1987 р. в день краху. Середня: -8,22%.**

Застереження: падіння більш ніж на 8% напочатку торгів є загрозливим сигналом.

Приклад. Обвал фондового ринку 19.10.1987 р.

Має місце невелика асиметрія у напряму низьких значень (хвіст зліва злегка довше, ніж праворуч), але незважаючи на це, розподіл приблизно нормальний з випадковими відхиленнями.

Середню процентну зміну -8,2% можна інтерпретувати так: якщо в п'ятницю на момент закриття торгів у вас був портфель інвестицій з однаковою кількістю грошей, вкладених в кожен з цих цінних паперів (відповідно до вартості акцій на момент закриття торгів в п'ятницю), то у понеділок при продажу на початку торгів ваш інвестиційний портфель втратив би 8,2% від своєї вартості.

Якщо ви вклали різний обсяг коштів у різні акції? Тоді втрату вартості портфеля можна було б розрахувати як середньозважене, використовуючи для визначенні ваги розміри вкладених коштів.

У той день середнє падіння індексу Dow Jone Industrial було рекордним – 508 пунктів, або 22,6%. Це стало справжньою трагедією для багатьох людей і організацій.

Самостійно упорядкуйте дані і знайдіть медіану.

Мода в контролі якості: метод Демінга

Будь-яка виробнича діяльність має відхилення від ідеалу. Демінг запропонував систематичний метод вимірювання відхилень виробничого процесу, виявлення причин цих відхилень і їх зменшення, вдосконалення за рахунок цього процесу, а значить, і підвищення якості продукції.

Припустимо, що підприємство реєструє причину браку кожного разу, коли з'являється виріб неприпустимої якості.

Причина проблеми	Число випадків
Пайка з'єднань	37
Пластмасовий корпус	86
Блок живлення	194
Бруд	8
Удар	1

Мода в контролі якості: метод Демінга. Висновки.

- Зрозуміло, що модою в цьому наборі є проблеми з блоком живлення. Мода допомагає зосередити увагу на найважливішій категорії. Немає необхідності розробляти додаткові заходи з підтримки чистоти на робочому місці або з недопущення падіння коробок, оскільки ці причини мало впливають на загальну частоту браку**. В першу черги слід звернути увагу на модальну категорію.**
- **У даній ситуації фірмі слід розібратися з проблемою "блок живлення" і вжити відповідних заходів.** Можливо, цей блок живлення має недостатню потужність для цього виробу і необхідне більш потужне джерело. Можливо, потрібно знайти більш надійного постачальника. У будь-якому випадку, мода допомагає конкретизувати проблему.

Приклад. Зборка системних блоків.

- Розглянемо стан зборки системних комп'ютерних блоків:

Стадія виробництва	Кількість системних блоків
A	57
B	38
C	86
D	45
E	119
F	42
Разом	387

- Так, **медіана** припадає **на стадію виробництва D**, оскільки ця стадія відділяє половину системних блоків, які знаходяться на початкових стадіях, від другої половини системних блоків на кінцевих стадіях збірки. **Проте в даному випадку медіана не збігається з модою.**
- **Модою є стадія E**, на якій знаходиться 119 системних блоків, тобто більше, ніж на будь-який інший стадії. *У такій ситуації керівництво має бути проінформовано про те, що найбільш "вузьке місце" у виробничому процесі.*

Приклад. Зборка системних блоків. Висновки

- *В цьому прикладі стадія Е – це встановлення материнської плати в системний блок. Наявність великої кількості системних блоків на цій стадії може бути свідченням більшою трудоємністю даної операції. Але, з іншого боку, це може бути свідченням наявності проблем у службовців, що працюють на цій стадії (можливо, причино в недостатній кількості людей або великій кількості відсутніх працівників).*

Перцентилі.

це показники набору даних, які характеризують ранги елементів у вигляді відсотків від 0 до 100%, а не у вигляді чисел від 1 до *n*, таким чином, що найменшим значенням відповідає нульовий перцентиль, найбільшому – 100-й перцентиль, медіані – 50-й перцентиль тощо. Перцентилі можна розглядати як показники, що розбивають набори кількісних і порядкових даних на певні частини.

Мета використання перцентилів:

1. Щоб показати значення елемента в даних при заданому перцентильному рангу (наприклад, 10-й перцентиль дорівнює 156293 дол.).

2. Щоб показати перцентильний ранг значення даного елемента в наборі даних (наприклад, ефективність продажів агента по збуту (Джона) становить 296994 дол., що відповідає 55-му перцептилю").

Перцентилі і блочна діаграма.

1. Найменше значення – 0-й перцентиль.
2. Нижній квартиль – 25-й перцентиль.
3. Медіана – 50-й перцентиль.
4. Верхній квартиль – 75-й перцентиль.
5. Найбільше значення – 100-й перцентиль.

Блочна діаграми дає можливість виявити викиди.

За методологією Тьюкі викидом зверху буде таке значення, яке виходить за межі $Q_3 + 1{,}5 \cdot (Q_3 - Q_1)$, знизу – $Q_1 - 1{,}5 \cdot (Q_3 - Q_1)$.

Приклад. Обвал фондового ринку 19.10.1987 р. продовження

Блочна діаграма процентного падіння вартості 29 промислових компаній зі списку Dow Industrial 19 жовтня 1987 р. в день краху.

Самостійно переконайтеся у відсутності викидів!

Функція кумулятивного розподілу даних

- представляється у вигляді графіка, який показує перцентилі шляхом встановлення відповідності між даними і відсотками. Оскільки на вертикальній вісі відкладаються відсотки від 0% до 100%, а по горизонтальній – самі перцентилі (тобто значення даних). Використовуючи цей графік можна легко знаходити перцентилі при заданому значенні відсотка, або значення відсотку, що відповідає певному значенню даних.
- Функція кумулятивного розподілу складається з вертикальних стрибків заввишки $1/n$ для кожного з n значень даних і горизонтальних відрізків, що поєднують точки значень даних.

Приклад. Обвал фондового ринку 19.10.1987 р. продовження

Кумулятивна діаграма процентного падіння вартості 29 промислових компаній зі списку Dow Industrial 19 жовтня 1987 р. в день краху.

Таким чином, 59% компаній втратили 8% і більше вартості цінних паперів. 10% компаній втратили 14% і більше своєї вартості, а 10% – 4% і менше.

Словник термінів (с. 151):

- Узагальнення – summarization
- Усереднення – average
- Середнє – mean
- Зважене середнє – weighted average
- Медіана – median
- Ранг – rank
- Мода – mode
- Перцентиль – percentile
- Екстремуми – extremes
- Квартили – quartiles
- П'ять базових показників – five-number summary
- Блокова діаграма – box plot
- Детальна блокова діаграма – detailed box plot
- Викид – outlier
- Функція кумулятивного розподілу – cumulative distribution function

Самостійна робота з використанням бази даних (с. 164):

1. Для розмірів річної заробітної плати:

а) Визначте середню.

б) Визначте медіану.

в) Побудуйте гістограму і визначте приблизне значення моди.

г) Порівняйте ці три показника. Що ви можете сказати про типовий розмір заробітної плати в цьому адміністративному підрозділі?

2. Для розмірів річної заробітної плати:

а) Накресліть функцію кумулятивного розподілу.

б) Знайдіть медіану, квартили й екстремуми.

в) Побудуйте блокову діаграму і прокоментуйте її.

г) Визначте 10-й и 90-й перцентилі.

д) Чому дорівнює перцентильний ранг для службовця під номером 6?

3. Розглядаючи стать службовців:

а) Узагальнити дані, обчисливши відсоток чоловіків і жінок.

б) Знайдіть моду. Про що вона свідчить?

4. Стосовно віку: дайте відповідь на питання 1.

5. У відношенні віку: дайте відповідь на питання 2.

6. У відношенні стажу роботи: дайте відповідь на питання 1.

7. У відношенні стажу роботи: дайте відповідь на питання 2.

8. Стосовно рівня підготовки: дайте відповідь на питання 3.

Проекти (с. 164):

1. Використовуючи Internet чи економічні журнали, підберіть набір даних з 25 чисел, що характеризують цікаву для вас фірму або галузь промисловості. Узагальнити ці дані, використовуючи всі вивчені вами методи, які можна застосувати в даному випадку. Використовуйте, як числові, так і графічні методи. Представте результати у вигляді короткої (дві сторінки) аналітичної записки, вказавши в першому абзаці свої рекомендації. (Не використовуйте великі графіки)
2. Знайдіть статистичні характеристики для двох обраних вами одновимірних кількісних наборів даних, які пов'язані з роботою, фірмою або галуззю промисловості. Для кожного набору даних:

А) Визначте середнє, медіану і моду.

Б) Як кожен з цих показників характеризує набір даних і економічну ситуацію?

В) Побудуйте гістограму і вкажіть значення цих трьох характеристик на горизонтальній осі. Прокоментуйте форму розподілу та взаємозв'язок між гістограмою і цими характеристиками.

Г) Побудуйте блокову діаграму і прокоментуйте переваги і недоліки гістограми в порівнянні з блочною діаграмою.

Ситуаційний аналіз (с. 165): Управлінські прогнози виробництва та маркетингу, або "Випадок підозрілого споживача"

Прийшовши на роботу, містер Б. Р. Харріс, як і очікував, виявив у себе на столі рекомендації містера Х. Е. Макроурі. У них містилися основні дані для квартальної презентації Харріса щодо обсягів виробництва на наступні три місяці, яку він мав провести сьогодні для вищого керівництва. Ці прогнози повинні були лягти в основу планування і показати теоретичні обсяги закупівель, запасів і робочих ресурсів в найближчому майбутньому. *Проте споживачі поводяться всупереч очікуванню, тому подібні прогнози завжди складні і, як правило, включають елемент припущень (суб'єктивної думки).*

Харріс і Макроурі вирішили змінити традицію і підготувати більш об'єктивне обґрунтування для цих прогнозів. Макроурі останнім часом аналізував дані опитування споживачів (нова експериментальна процедура, заснована на відповідях 30 репрезентативних споживачів, табл. 3) і підготував звіт, в якому, зокрема, стверджувалося: "У наступному кварталі ми очікуємо обсяг продажів на суму 477108 дол.

Прогнози обсягів продажів по регіонах наведені в табл. 1. Ми рекомендуємо збільшити виробництво до рівня, який узгоджується з очікуваним зростанням обсягів продажів ...".

Ситуаційний аналіз: Управлінські прогнози виробництва та маркетингу, або "Випадок підозрілого споживача"

Таблиця 1

Показники	II кв. поточного року (прогноз)	I кв. поточного року	II кв. минулого року
Обсяги продажу			
Північно-Схід	441058	331309	306718
Північно-Захід	291948	22185	200201
Південь	149518	118151	101721
Середній захід	370577	277952	254315
Південно-захід	224007	165332	157843
Разом	1477108	1114929	1020798
Продукція (оптова вартість)			
Стільці	514 458	425926	389115
Столи	228314	201125	197250
Книжні полки	272624	209105	180475
Шафи	461702	276500	295400
Разом	1477108	1112655	1071240
Виробництво (штук)			
Стільці	11433	9465	8647
Столи	1827	1609	1578
Книжні полки	4194	3217	2915
Шафи	1319	790	844

Ситуаційний аналіз: Управлінські прогнози виробництва та маркетингу, або "Випадок підозрілого споживача"

- Харрісу було нелегко. Прогноз містив велике збільшення обсягів як відносно поточного кварталу (на 32,5%), так і до аналогічного кварталу минулого року (на 44,7%). За останні роки темпи зростання фірми не були такими високими. Разом з тим, рекомендації містили пропозиції про збільшення обсягу виробництва у зв’язку з очікуваним збільшенням продажів.

Ситуаційний аналіз: *Чому виникають сумніви?*

- Тому що, якщо прогноз невірний і обсяг продажів не збільшиться, фірма отримає великий і дорогий запас готової продукції (яка, до того ж, проведена з підвищеними, порівняно зі звичайними, витратами через оплату понаднормових робіт, зарплата додаткових робочих та оренди додаткового обладнання) на додаток до своїх звичайних поточних витрат (включаючи відсоток, який фірма могла б отримати з суми грошей, яку вона змушена була витратити на виробництво додаткової продукції).
- Харріс висловив свої сумніви і Макроурі теж завагався. Так, все здавалося просто: *ліпити з результатів опитування середнє прогнозоване значення споживчих витрат і помножити його на загальну чисельність споживачів в даному регіоні.*

Ситуаційний аналіз: У чому може бути помилка?

- Харріс і Макроурі вирішили уважніше вивчити дані. Нижче наведена таблиця 2, яка включає загальну інформацію (оптова ціна кожного найменування продукції та кількість реальних покупців по регіонах) і результати вибіркового дослідження.
- Кожен з 30 відібраних споживачів вказав, скільки одиниць кожного з найменувань товару він планує замовити в наступному кварталі. Колонка "Вартість" містить обсяг готівки, які отримає фірма (*наприклад, покупець 1 планує придбати 3 стільці по 45 дол. і 4 книжкові полиці по 65 дол., на загальну суму 395 дол.*).

Ситуаційний аналіз: Управлінські прогнози виробництва та маркетингу, або "Випадок підозрілого споживача"

Таблиця 2

Продукція	Ціна, дол.
Стільці	45
Столи	125
Книжні полки	65
Шафи	350
Реальні покупці	Кількість
Північно-Схід	303
Північно-Захід	201
Південь	103
Середній захід	255
Південно-захід	154
Разом	1016

Ситуаційний аналіз:

Таблиця 3

Покупець	Стільці, шт.	Столи, шт.	Книжні полиці, шт.	Шафи, шт.	Вартість, дол.
1	3	0	4	0	395
2	9	1	6	1	1270
3	23	2	1	2	2050
4	7	0	3	0	510
5	4	0	0	0	180
6	14	1	5	0	1080
7	6	0	5	0	505
8	14	1	0	0	755
9	1	5	17	3	2825
10	2	0	4	1	700
11	16	1	1	1	1260
12	4	0	4	0	440
13	6	0	4	1	680
14	2	1	8	2	1435
15	42	15	21	16	11430
16	3	0	0	2	835
17	7	3	0	0	690
18	1	4	2	o	675
19	43	0	4	0	2195
20	6	*2*	4	2	1480
21	3	1	1	0	325
22	45	6	1	0	2840
23	0	2	7	1	1055
24	13	6	3	0	1053
25	19	0	2	2	1685
26	0	0	0	0	0
27	8	0	7	0	815
28	14	3	3	1	1550
29	6	o	1	2	1035
30	17	0	6	0	1155
Разом по вибірці	338	54	124	39	43670
Середнє	11,267	1,8	4,133	1,3	1455667
Середня вартість	507	225	268667	455	1455667
Загальний прогноз для всіх покупців (помножено на 1016 покупців)					
Вартість	514468	228314	272624	461702	1477108
Кількість одиниць	11433	1827	4194	1319	

Ситуаційний аналіз: Питання для обговорення (с. 168)

1. Чи підходить в даному випадку звичайний метод, що застосовується Харрісом і Макроурі, метод, заснований на середньому, або цей метод заздалегідь невірний? Обґрунтуйте вашу відповідь.
2. Вивчить дані, використовуючи статистичні характеристики та графіки. Який можна зробити висновок?
3. Що б ви порекомендували зробити Харрісу і Макроурі для підготовки до сьогоднішньої презентації?

Мінливість даних, її статистичне оцінювання

1. **Продуктивність праці працівників**. Цілком очевидно, що ефективність роботи відділу визначається загальною продуктивністю праці всіх його співробітників. Однак будь-які зусиллі, спрямовані на підвищення продуктивності праці, мають враховувати індивідуальні особливості працівників. Визначення мінливості продуктивності праці дає можливість виявити розкид таких індивідуальних відмінностей і отримати корисну інформації: для планування заходів підвищення загальної продуктивності праці.

2. **Фондова біржа.** Фондова біржа в середньому забезпечує більш високу прибутковість вкладених коштів, ніж, наприклад, фонди грошового ринку. Однак робота на фондовій біржі пов'язана з великим ризиком, а інвестування в акції може призвести до реальних втрат. Таким чином, середня, або очікувана доходність не відображає повною мірою всю картину. Міра мінливості прибутковості окремих інвестицій буде відображати рівень ризику, пов’язаного з кожним конкретним вкладенням коштів.

3. Стратегічне планування. Припустимо, що ви порівнюєте маркетингові витрати своєї фірми з аналогічними витратами фірм, що працюють у вашій галузі промисловості, і виявляєте, що витрати вашої фірми менше витрат, типових для даної галузі. Для того, щоб оцінити витрати на майбутнє, дуже корисним може виявитися облік розкиду відповідних даних по галузі. Знайшовши різницю між значенням витрат фірми і середнім значенням по галузі і порівнявши отриману величину з мірою мінливості витрат у галузі, можна зробити висновок про те, чи знаходиться маркетингова діяльність вашої фірми порівняно з іншими аналогічними фірмами лише па дещо нижчому рівні або ж ваша фірма є деяким винятком із загальної картини.

Три способи опису ступеня мінливості набору даних

1. Стандартне відхилення (середнє квадратичне відхилення)

При розрахунку стандартного відхилення суму відхилень ділять на n-1 замість n це пов'язано з поправкою, обумовленою тим фактом, що при роботі з вибіркою справжнє значення середнього генеральної сукупності невідомо. Ця поправка обумовлена втратою при обчисленні відхилень однієї порції інформації (однієї ступені свободи). Втраченої є інформація про істинні значення даних (оскільки тепер, при роботі з відхиленнями, дані розподіляються не навколо середнього, а навколо нуля).

Важливо: *Чим менший обсяг вибірки ми маємо, тим більше проявляються розбіжності. Так, у випадку 10 елементів стандартне відхилення вибірки перевищує стандартне відхилення генеральної сукупності на 5,4%. При 35 елементах відмінність становить 2,1%. Зі збільшенням обсягу вибірки ця розбіжність зменшується, добігаючи до 1,0% для 50 елементів і 0,5% для 100 елементів.*

Три способи опису ступеня мінливості набору даних: приклад

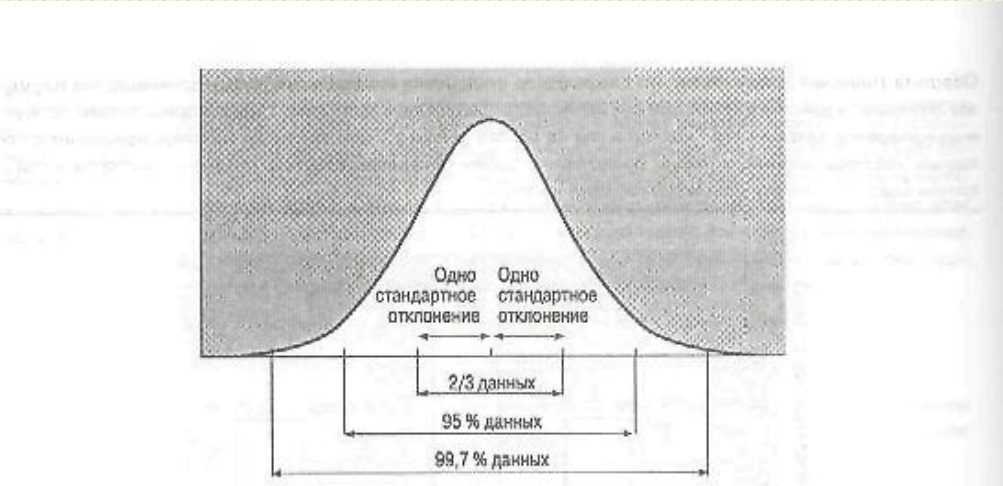
Витрати на рекламу. Припустимо, що фірма витрачає на рекламу 19 мільйонів доларів на рік і керівництво фірми бажає знати, чи відповідає це сума реальним потребам. Незважаючи на те, що існує досить багато способів оцінки цієї стратегічно важливої величини, завжди корисно порівняти себе з конкурентами.

Нехай інші працюючі у вашій сфері фірми, що мають приблизно такий саме розмір, в середньому витрачають на рік на рекламні цілі 22,3 мільйона дол. Можна скористатися стандартним відхиленням для того, щоб виходячи з різниці (22,3 – 19=3,3 млн дол.) оцінити, наскільки витрати на рекламу вашої фірми менше, ніж в інших аналогічних фірмах.

Розглянемо витрати на рекламу (в млн дол.) групи з 17 фірм, схожих на вашу: Легко переконатися, що середнє становить 22,3 млн дол. (результат округлення 22,29411 млн дол.) і стандартного відхилення 9,18 млн дол. (результат округлення значення 9,177177). Оскільки різниця між витратами на рекламу на фірмі і середніми витратами на рекламу в групі фірм (3,3 млн дол.) навіть менше одного стандартного відхилення (9,10 млн дол.), то можна зробити висновок, що бюджет рекламної діяльності вашої фірми досить типовий. Незважаючи на те що він менше середнього значення, він ближче до цього середнього, ніж бюджет типовою фірми з даної групи,

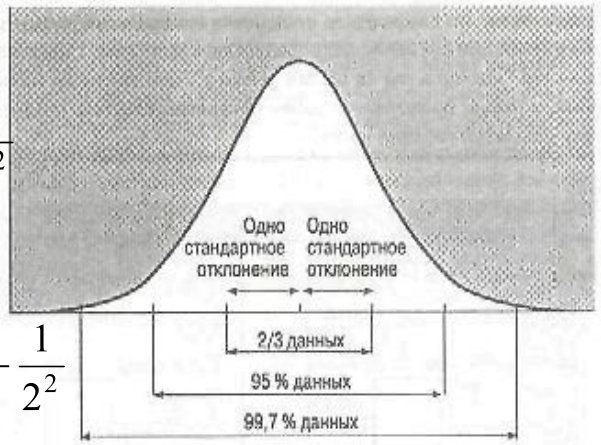
Три способи опису ступеня мінливості набору даних: приклад

Ваша фірма з бюджетом реклами в 19 млн дол. дійсно виявляється досить типовою. Незважаючи на те, що різниця в 3,3 млн дол. між бюджетом вашої фірми і середнім значенням в грошовому вираженні здається доволі значною, воно незначне порівняно з розходженнями, що існують між бюджетами фірм, що входять до групи, З точки зору обсягу бюджету реклами становище вашої фірми не набагато нижче середнього.



Три способи опису ступеня мінливості набору даних

У випадку, якщо набір даних не підкоряється закону нормального розподілу можна скористатися правилом Чєбишева відповідно до якого як мінімум значень $1-\frac{1}{a^2}$ потрапляє в проміжок, що лежить в межах ***a*** стандартних відхилень від середнього значення. Наприклад, при ***a=2*** щонайменше 75% даних (це значення розраховується як $1-\frac{1}{2^2}$ має знаходитися на відстані подвійного стандартного відхилення від середнього, навіть якщо розподіл не є нормальним (порівняйте з величиною для нормального розподілу, що становить приблизно 95%). Якщо ***a=3***, щонайменше 88,9% даних буде знаходитися в межах потрійного стандартного відхилення від середнього значення.



- Особливого сенсу цей розподіл набуває в картах контролю, які широко використовують в аналізі контролю якості продукції.
- *В цьому випадку заслуговують на увагу лише ті результати спостережень, які відстоять від середнього на відстані більш ніж три сігми.*

Приклад. Зміна прибутку на біржі

Розглянемо непостійність фондової біржі за період часу, що передував обвалу 1989 р. на прикладі Індекса Доу Джонса (DJI) на момент закриття біржі. Індекс Доу Джонса обчислюється як середнє значення ринкових цін акцій 30 великих промислових компаній. Зазвичай інвестори вивчають такі дані у вигляді графіка залежності індексу цін від часу

IDJ	i	IDJ	i	IDJ	i	IDJ	i
2572,07 (31.07.1978)	x	2685,82	0,012	2545,12	-0,006	2570,17	0,001
2557,08	-0,006	2706,79	0,008	2549,27	0,002	2601,50	-0,105
2546,72	-0,004	2709,50	0,001	2576,05	0,011	2581,57	0,122
2566,65	0,008	2697,07	-0,005	2608,74	0,013	2596,28	0,006
2594,23	0,011	2722,42	0,009	2613,04	0,002	2653,20	0,022
2592,00	-0,001	2701,85	-0,008	2566,58	-0,018	2640,99	-0,005
2635,84	0,017	2675,06	-0,010	2530,19	-0,014	2640,18	0,000
2680,48	0,017	2639,35	-0,013	2527,90	-0,001	2548,63	-0,035
2669,32	-0,004	2662,95	0,009	2524,64	-0,001	2551,08	0,001
2691,49	0,008	2610,97	-0,020	2492,82	-0,013	2516,64	-0,014
2685,43	1,855	2602,04	-0,003	2568,05	0,030	2482,21 (9.10.1987)	-0,014
2700,57	-0,649	2599,49	-0,001	2585,67	0,007		
2654,66	-0,017	2561,38	-0,015	2566,42	-0,007		

Приклад . Індекс Доу Джонса цін акцій 30 великих промислових компаній за період з 31 липня 1987 р. по 9 жовтня 1987 р.

Приклад: продовження

Розподіл денного прибутку акцій 30 великих промислових компаній за період з 1 серпня 1987 р. по 9 жовтня 1987 р., у%. *Середній прибуток приблизно дорівнює нулю, що означає: короткочасні зростання і зниження були рівноцінними. Стандартне відхилення, яке складає 1,194 п.п. відображає величину звичайних добових флуктуацій. Впродовж цього часу вкладений на фондовому ринку долар міг змінитися на 1 цент.*

Приклад: пояснення

- Середній денний прибуток за цей період часу становив -0,066%, тобто він приблизно дорівнює нулю (середнє зниження склало сім сотих відсотка). Таким чином, на ринку в цей час тримався середній курс. Стандартне відхилення становить 1,194 п.п., що означає 1 дол., вкладений у фондовий ринок, в середньому змінювався за добу на 0,01194 дол., в тому сенсі, що вкладення $ 1 могло призвести за добу до прибутку або втрати приблизно в 0,01194 дол. Крайні значення з обох боків від центру, демонструють максимальний розмір зростання і падіння за один день. Так, 22 вересня на ринку спостерігався підйом з 2492,82 до 2568,05, що склало зростання на 75,23 пункти, з денним прибутком 3% (прибуток в розмірі 0,03 дол. на один долар, вкладений на день раніше). А 6 жовтня на ринку відбулося зниження з 2640,18 до 2548,63, тобто на 91,55 пункти. Денний прибуток при цьому склав -3,5% (втрати в розмірі 0,035 дол. на один долар, вкладений на день раніше).

Показник	Середнє	Стандартне відхилення	Мінімум	Максимум
Денний прибуток, %	-0,06555	1,194	-3,47	3,01

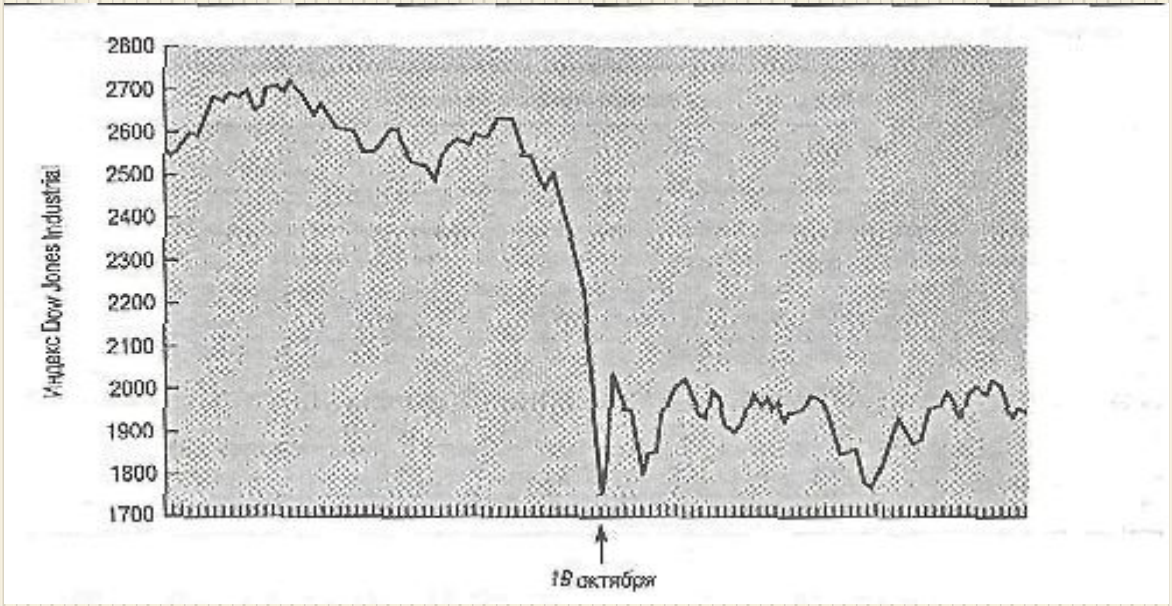
Приклад: висновки

- Для того, щоб знаходитися в межах значення **одного стандартного відхилення** (1,194) від середнього значення (-0,07), розмір денного прибутку повинен знаходитися в межах від (-0,07–1,194 = -1,264%) до (-0,07+1,194 = 1,124%). З **49 наведених значень денного прибутку цій вимозі відповідають 32**. Таким чином, 32/49, або *65,3% значень денного прибутку віддалені від середнього значення на відстань, що не перевищує одного стандартного відхилення*. Цей відсоток досить близький до значення 2/3 (або 66,7%) – приблизно тієї частини від загальної кількості значень, яку ми могли б очікувати у разі ідеального нормального розподілу. Отже, можна вважати, що "правило двох третин" працює.
- Для того щоб залишатися в межах відстані **в дві величини стандартного відхилення** від середнього значення, денний прибуток має знаходитися в межах від (-0,07–2·1,194 = -2,458%) до (-0,07+2·1,194 = 2,318%). З **49 значень денного прибутку цій вимозі відповідають 47** (всі, за винятком двох крайніх значень, на які ми вже звернули увагу раніше). Таким чином, 47/49, або 95,9%, величини денного прибутку розташовані по відношенню до середнього значення на відстані, що не перевищує подвійного стандартного відхилення. Отримане значення досить близько до значення 95%. Яке ми могли б очікувати у разі ідеального нормального розподілу.

Приклад: Обвал на фондовій біржі у 1987 р.: 19 стандартних відхилень.

В понеділок 19 жовтня 1987 р. індекс Доу Джонса втратив 508 пунктів, з 2246,74 (у попередню п'ятницю) до 1738,74. Це відповідає денному доходу (-0,2261); таким чином, фондовий ринок втратив 22,61% своєї вартості.

Таке неочікуване падіння вартості, показане на рис. було найбільшим з часу "Великого кризи" 1929 року.



Індекс Доу Джонса цін акцій 30 великих промислових компаній за період з 31 липня 1987 р. по 31 грудня 1987 р.

Приклад: Обвал на фондовій біржі у 1987 р.: 19 стандартних відхилень.

Для того щоб представити собі, наскільки екстремальною з точки зору статистики виявилася ситуація при цьому обвалі, порівняємо її з тією, яку слід було б очікувати відповідно до попередньої поведінки ринку. В якості базового періоду скористаємося попереднім прикладом, в якому розглянуто проміжок часу з 31 липня по 9 жовтня, до п'ятниці за тиждень до обвалу. Для базового періоду мм визначили, що середнє значення денного прибутку становить -0,07%, а стандартне відхилення дорівнює 1,194 п.п. Поставимо питання у такий спосіб: скільки величин стандартного відхилення необхідно відкласти вниз від середнього значення, щоб отримати втрати, понесені 19 жовтня? Відповідь на це питання така:

$$\frac{(-22{,}61-(-0{,}07))}{1{,}194}=-18{,}87$$ стандартних відхилень

Приклад: Обвал на фондовій біржі у 1987 р.: 19 стандартних відхилень.

Якби денний дохід на біржі **дійсно мав нормальний розподіл** (і розподіл не було б схильним до швидких змін), такого **екстремального результату не могло б виникнути ніколи**. У такому випадку досить часто (приблизно в одній третині випадків) можна було б очікувати денний прибуток, що відрізняється від середнього значення більш ніж на одне стандартне відхилення. Різниця у два або більше стандартних відхилень спостерігалася б час від часу (приблизно у 5% випадків). Відмінність, що становить три стандартних відхилення і більше, могло б спостерігатися тільки дуже рідко приблизно в 0,3% випадків, або, для більшої наочності, можна сказати, що це відбувалося б не більше одного разу на рік. Навіть відхилення, яке становить п'ять стандартних відхилень, було б уже досить не характерним для ідеального нормального розподілу. **Різниця у 19 стандартних відхилень здається зовсім неймовірною.**

Висновок.

Це показує, що денний прибуток на фондовій біржі не підпорядковується ідеальному нормальному розподілу. Це не означає, що теорія в чомусь невірна. Це тільки вказує на те, що теорія в донному випадку непридатна. Незважаючи на те, що нормальний розподіл описує денний прибуток для тривалішого проміжку часу роботи фондової біржі, обвал 1987 р. нагадує про необхідність перевірки правильності всіх припущень для захисту власних інтересів в особливих випадках.

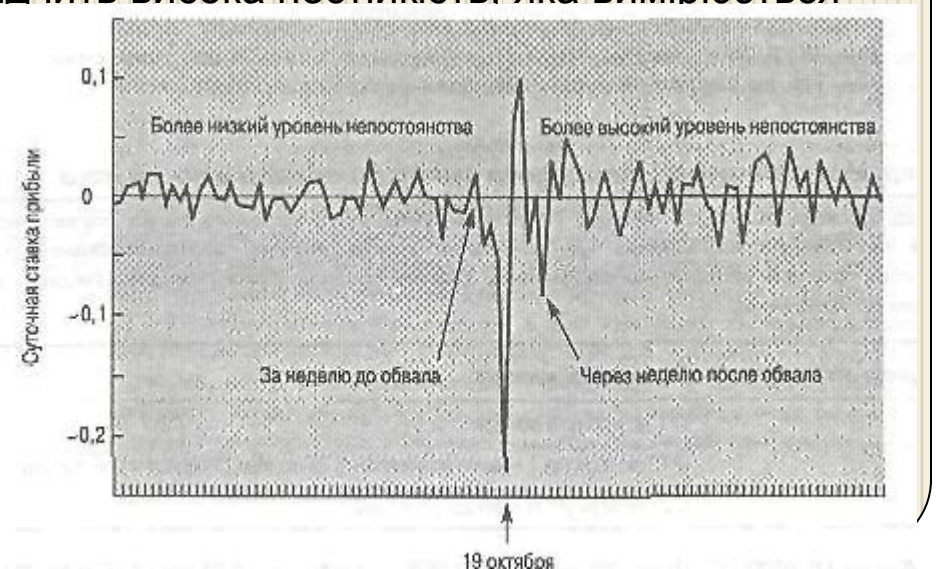
Приклад: продовження. Нестійкість фондового ринку до обвалу і після

У період після обвалу 19 жовтня 1997, стан ринку характеризувався високою нестійкістю. Ступінь нестійкості, оцінена за допомогою стандартного відхилення денного прибутку за різні періоди часу, склала

Стандартне відхилення, п.п.	Проміжок часу
1,19	З 1 серпня по 9 жовтня
8,36	З 12 жовтня (за 1 тиждень до обвалу) по 26 жовтня (1 тиждень після обвалу)
2,09	З 27 жовтня по 31 грудня

У період часу, який безпосередньо примикає до обвалу, стандартне відхилення було приблизно в сім разів вище, ніж до цього періоду. Після обвалу, стандартне відхилення зменшилася, проте залишилося приблизно в два рази вище, ніж до нього (2,09% порівняно з 1,19%). Ринок після обвалу, безумовно, повернувся до ділової активності, однак він залишився "неспокійним", про що свідчить висока нестійкість, яка вимірюється стандартним відхиленням.

Якщо абстрагуватися від сильних коливань ринку напередодні і відразу після 19 жовтня, можна побачити, що розмах по вертикалі коливань графіка праворуч від цієї дати приблизно в два рази вище, ніж зліва від неї.



Приклад: продовження. Нестійкість фондового ринку до обвалу і після

- Останнім часом нестійкість фондового ринку значно знизилася. Нижче наведена таблиця стандартних відхилень денного прибутку для кожного року з 1990 по 1998, що розраховане для фондового індексу S&P500. Зверніть увагу, що типова зміна цін у 1995 становило приблизно половину відсотка (від вартості всього портфеля) на день, проте потім нестійкість ринку стала зростати.

Рік	Стандартне відхилення, п.п.
1990	1,00
1991	0,89
1992	0,60
1993	0,53
1994	0,61
1995	0,48
1996	0,73
1997	1,12
1998	1,28

Самостійно: проаналізуйте варіацію стандартного відхилення. Зробіть висновки.

Приклад: Диверсифікація на фондовому ринку

- Розглянемо величину ризику для трьох випадків:
- (1) володіння тільки акціями Boeing,
- (2) володіння тільки акціями Johnson & Johnson і
- (3) володіння портфелем з названих акцій в рівних частках.

Стандартне відхилення денної ставки прибутку для кожного з цих випадків (за 1994 і три перші квартали 1995)

Портфель	Стандартне відхилення, п.п.
Johnson & Johnson	1,39
Boeing	1,46
Обидві компанії	0,99

Зверніть увагу на зниження ризику у випадку володіння акціями більш ніж однієї компанії (ризик знижується приблизно з 1,46 п.п. на день до величини порядку 1 п.п. на день). Якщо портфель містить акції більшої кількості компаній, ризик можна знизити ще більше. Ризикованість акцій S&P500 (що включають акції 500 різних компаній) була в цей період ще менше – порядку 0,6 п.п.

Три способи опису ступеня мінливості набору даних

2. Розмах варіації.

Розмах легко обчислюється, проте дає лише поверхневе уявлення про мінливість даних і має обмежене застосування. Ця величина описує межі зміни даних в наборі і являє собою відстань між мінімальним і максимальним значеннями.

Приклад. Корисність розмаху при первинному аналізі інформації: випадок з практики.

Цей гіпотетичний набір даних (**тривалість перебування в лікарні, ліжко-днів**) оснований на досвіді одного з дослідників центру економічних досліджень і проблемах, які у нього виникли, коли він впродовж двох тижнів намагався застосувати комп'ютер для аналізу записів медичної статистики при вивченні ефективності різних систем надання послуг охороні здоров'я.

17	33	5
16	5	6
1	1	16
1	7	12
7	4	386
74	13	2
2	6	7
163	33	28
51		

Робота лікарень зараз більш схожа на комерційну діяльність, ніж це було раніше. Багато організацій, які надають послуги в області охорони здоров'я, просто наймають лікарів як службовців, в той час як в традиційних лікарнях лікарі мають більшу незалежність. Ще одна причина комерціалізації охорони здоров'я полягає в тому, що відповідно до програми охорони здоров'я Medicare в даний час є тенденція до фіксованих виплат на основі діагнозу, а не гнучкі виплати залежно від тривалості лікування. Це сприяє виникненню сильної тенденції до **скорочення тривалості лікування у разі конкретної хвороби пацієнта.**

Приклад. Корисність розмаху при первинному аналізі інформації: випадок з практики.

Висновок:

*При ретельній перевірці було виявлено помилку друку. Реальне значення **286** було помилково записано як **386**. Таким чином, у виправленому наборі даних розмах становив **285**.*

В якості одного з показників інтенсивності лікування виступає кількість днів перебування пацієнта в лікарні. Розмах ряду становить 386-1 = 385 днів, що являє собою занадто велике значення, оскільки в році тільки 365 (або 366) днів, а цей набір даних, належить до одного року. Даний приклад ілюструє користь застосування поняття розмаху для редагування набору даних з метою виявлення помилок перед початком аналізу даних. Для цього також корисно уважно дослідити найменші і найбільші значення.

17	33	5
16	5	6
1	1	16
1	7	12
7	4	386
74	13	2
2	6	7
163	33	28
51		

Три способи опису ступеня мінливості набору даних

3. Коефіцієнт варіації. Коефіцієнт варіації зазвичай обирається як відносна міра мінливості. Цей показник використовується досить часто. Він показує, наскільки сильно зазвичай відрізняється результат конкретного спостереження від середнього значення, в процентному відношенні до середнього

Важливим є також **оцінювання впливу на мінливість даних зміни шкали вимірювання** (наприклад, перехід від японської ієни до доларів США або перехід від кількості одиниць випущеної продукції до грошової вартості цієї продукції).

Приклад. Невизначеність прибутковості портфеля інвестицій

Ви вклали **10000 дол. у 200 акцій корпорації**, які продаються по **50 дол**. за штуку. Ваш знайомий придбав **100 акцій цієї ж корпорації 5000 дол**. Ви очікуєте, що вартість акцій зросте в майбутньому році до **60 дол**. за акцію, що відповідає ставці прибутку **20%**, Ви також вважаєте маркетингову стратегію корпорації досить ризикованою, оскільки вона характеризується стандартним відхиленням курсу акцій 9 дол.

Обсяг ваших інвестицій зросте наступного року до січня 12000 дол. (60·200), зі стандартним відхиленням $ 1800 (9·200). **Інвестиції вашого знайомого**, як очікується, наступного року зростуть до 6000 дол., зі стандартним відхиленням 900 дол.

Складається враження, що ваш ризик в два рази більше, ніж ризик вашого знайомого. І це дійсно, так, оскільки ваші інвестиції в абсолютному вираженні в два рази більше. Однак ви робите вкладення в одні й ті ж цінні папери, а саме в акції однієї і тієї ж корпорації. Таким чином, у всіх відносинах, за винятком обсягу інвестицій, ваша схильність до ризику буде однаковою. У **відносному вираженні ризики мають бути однаковими**. У цьому можна переконатися, обчисливши коефіцієнт варіації, який буде дорівнювати в обох випадках 15%.

Приклад. Продуктивність праці у відділі торгівлі по телефону

- Розглянемо відділ торгівлі па телефону, в якому працюють 19 співробітників, що займаються продажем квитків на концерт симфонічної музики. У середньому кожен співробітник продає **23 квитки за годину.** Стандартне відхилення становить **6 квитків на годину**. Це означає, що будь-який з співробітників може продавати на годину в середньому на 6 квитків більше або менше середнього значення. Відмінності в роботі співробітників складають 6/23=0,261, або 26,1%. Це означає, що варіація продуктивності праці співробітників складає приблизно 26,1% від середнього рівня продажів. ***Використання коефіцієнта варіації є особливо корисним при проведенні порівнянь в умовах різних обсягів.***
- Розглянемо ще один відділ торгівлі по телефону, що займається продажем квитків в театри, і в якому середній рівень продажів складає 35 квитків на годину, а стандартне відхилення дорівнює 7. Оскільки продуктивність праці при продажу театральних квитків виявляється в цілому вище продуктивності при продаж квитків на концерти симфонічної музики, природно, що варіація буде вищою. Проте коефіцієнт варіації для відділу, що працює з театральними квитками, становить 20,0%. *Порівнюючи цю величину з коефіцієнтом 26,1% , що характеризує варіацію продажів білетів на симфонічні концерти, менеджери можуть зробити висновок про те, що група, яка працює з театральними квитками фактично більш однорідна.*

Приклад. Загальна вартість виробленого товару

- Розглянемо виробництво продукту для якого фіксовані витрати складають 1 млн дол., а змінні витрати 0,50 дол. на одиницю. На основі ретельного аналізу ринкового попиту менеджери передбачили у наступному місяці випуск 1200 тис. од. Виходячи з попереднього досвіду невизначеність для прогнозованого обсягу виробництва можна оцінити на рівні 250 тис. од. Таким чином, очікується випуск в середньому 1200тис.±250 тис.
- **Якщо для обсягу виробництва існує такий прогноз, то яким буде прогноз для витрат?** Зверніть увагу на те, що обсяг виробництва переводиться у витрати шляхом множення кількості одиниць товару на 0,50 дол. з додаванням 1 млн дол. Таким чином, у нашому випадку загальна вартість становить: 0,50·1200+1000=1600 тис. дол., стандартне відхилення вартості становить: 0,50·250=125 тис. дол. *Отже, кошторис витрат складений. Очікуються витрати 1,6 млн дол. зі стандартним відхиленням (невизначеністю) 125 тис. дол.*
- Коефіцієнт варіації для кількості одиниць виробленої продукції складе 250/1200·100%=20,8%. Коефіцієнт варіації для витрат дорівнює 125/1600·100%=7,8%. *Зверніть увагу, що відносна варіація у вартісному вираженні виявляється значно меншою, оскільки великі постійні витрати призводять до збільшення бази порівняння і відповідно до помітного зниження варіації.*

Словник термінів (с. 198):

- Мінливість – variability
- Різноманітність – diversity
- Невизначеність – uncertainly
- Розсіювання – dispersion
- Розкид – spread
- Стандартне відхилення – standard deviation
- Відхилення – deviation,
- Дисперсія – variance
- Стандартне відхилення вибірки – sample standard deviation
- Стандартне відхилення генеральної сукупності – population standard deviation
- Розмах – range
- Коефіцієнт варіації – coefficient of variation

Самостійна робота з використанням бази даних (с. 215):

Зверніться до бази даних про найманих працівників у додатку А.

1. Для розміру заробітної плати за рік:

а) Знайдіть розмах.

б) Знайдіть стандартне відхилення.

в) Знайдіть коефіцієнт варіації.

г) Порівняйте три показника. Як вони характеризують типову заробітну плату в розглянутому відділі?

2. Для розміру заробітної плати за рік:

а) Побудуйте гістограму і покажіть на ній середнє значення і стандартне відхилення.

б) Скільки працівників мають зарплату, відмінну від середньої не більше ніж на одну величину стандартного відхилення?

Як ця кількість узгоджується з тим числом, яке можна було б очікувати у разі нормального розподілу?

в) Скільки працівників мають зарплату, відмінну від середньої не більше ніж на два стандартних відхилення?

Як це кількість узгоджується з тим числом, яке можна було б очікувати в разі нормального розподілу?

г) Скільки працівників мають зарплату, відмінну від середньої не більше ніж на три стандартних відхилення?

Як це кількість узгоджується з тим числом, яке можна було б очікувати в разі нормального розподілу?

3. Для віку співробітників дайте відповіді на запитання вправи 1.

4. Для віку співробітників дайте відповіді на запитання вправи 2.

5. Для кваліфікації (досвіду роботи) співробітників дайте відповіді на запитання вправи 1.

6. Для кваліфікації (досвіду роботи) співробітників дайте відповіді на питання вправи 2.

Проекти (с. 216):

1. У відповідності до власних інтересів візьміть набір значень для підприємств двох галузей промисловості (не менше 15 підприємств у кожній групі).

А) Для кожної групи:

1) охарактеризуйте мінливість властивості, скориставшись описаними методами, які можуть бути застосовані до ваших даних;

2) для кожного з наборів даних зобразите отримані характеристики мінливості на гістограмі та/або блокової діаграмі.

3) опишіть, що ви дізналися про галузь промисловості на основі проведеного аналізу мінливості.

Б) Проведіть для обох груп наступні порівняння:

1) порівняйте стандартні відхилення;

2) порівняйте коефіцієнти варіації;

3) величини розмаху.

4) коротко опишіть, що ви дізналися про результат порівняльного аналізу розглянутих галузей промисловості, а саме: яка з характеристик мінливості виявилася найбільш корисною?

2. Візьміть набір даних, що включає не менше 25 значень, що характеризують підприємство або галузь промисловості, яка вас цікавить. Опишіть дані, скориставшись усіма вивченими до цього моменту методами, які застосовні до ваших даних. Використовуйте як чисельні, так і графічні методи; звертайте увагу як на типове значення, так і на мінливість. Представте отримані результати у вигляді двосторінкового звіту для керівництва, сформулювавши рекомендації у першому абзаці.

Ситуаційний аналіз (с. 216-217): Чи слід продовжувати роботу з цим постачальником?

- Ви і один з ваших співробітників, Б.У. Келлерман, отримали завдання – оцінити нового постачальника деталей до обладнання, яке випускається вашою фірмою для догляду за будинком і садом. Одна з деталей повинна мати розмір 8,5 см. Однак допускається також будь-який розмір в межах від 8,4 до 8,6 см. Келлерман нещодавно доповів про дослідження розмірів 99 поставлених деталей. Зроблений Келлерманом перший начерк звіту містить такі рекомендації.
- Якість деталей, що поставляються фірмою НурaTech, не відповідає "нашим вимогам. Незважаючи на те, що ціни цієї фірми досить низькі і привабливі, а поставки відбуваються відповідно до графіку, якість виробів недостатньо висока. Ми рекомендуємо серйозно розглянути питання про використання альтернативних джерел поставок.
- Тепер ваша черга. Після аналізу отриманих Келлерманом цифр і проекту звіту перед вами стоїть завдання підтвердити його рекомендації (або спростувати) на основі власного незалежного дослідження.
- Висновки Келлермана представляються осмисленими. Основний аргумент полягає в тому. що, незважаючи на середнє значення, яке становить 8,494 см і дуже близьке до стандарту – 8,5 см, стандартне відхилення досить значне і дорівнює 0,103. В результаті цього дефектні деталі складають приблизно третину всіх поставляються виробів. Дійсно, Келлерман явно пишаэться тим, що пам'ятає знання, отримані давним-давно при вивченні статистики, – щось про те, що потрапляння в межі одного стандартного відхилення від середнього спостерігається приблизно в третині випадків. У даному конкретному випадку при такій ціні можна допустити 10, або навіть 20% дефектних деталей, однак 33% виходить за рамки розумного.

Ситуаційний аналіз (с. 216-217): Чи слід продовжувати роботу з цим постачальником?

- Ситуація видається цілком очевидною, однак для того, щоб переконатися в правильності отриманих Келлерманом висновків, ви вирішуєте все-таки самостійно швидко переглянути дані. Природно, ви очікуєте, що висновки підтвердяться. Ось цей набір даних:

8,503	8,503	8,500	8,496	8,500	8,503	8,497	8,504	8,503	8,508
8,502	8,501	8,489	8,499	8,492	8,497	8,508	8,502	8,505	8,489
8,505	8,499	8,890	8,505	8,504	8,499	8,499	8,505	8,493	8,494
8,510	8,310	8,804	8,503	8,787	8,502	8,509	8,499	8,493	8,493
8,346	8,499	8,505	8,509	8,499	8,503	8,494	8,511	8,501	8,497
8,501	8,502	8,780	8,494	8,500	8,498	8,500	8,502	8,501	8,491
8,511	8,494	8,374	8,492	8,497	8,150	8,496	8,501	8,489	8,506
8,493	8,498	8,535	8,900	8,433	8,601	8,497	8,501	8,438	8,503
8,508	8,501	8,499	8,504	8,505	8,461	8,497	8,495	8,504	8,501
8,493	8,504	8,897	8,505	8,490	8,492	8,503	8,507	8,497	

Питання для обговорення:

1. Чи правильні результати обчислень Келлермана?
2. Уважно подивіться на дані, використовуючи належны статистичні методи.
3. Чи вірні висновки, які зробив Келлерман? Якщо так, чому ви так вважаєте? Якщо ні, то чому, і що слід зробити для вироблення правильних рекомендацій?

Головне – правильно розподілити