**Лекція 2. СТРУКТУРИ ДАНИХ. КЛАСИФІКАЦІЯ,**

**РОЗПОДІЛ І ПЕРЕТВОРЕННЯ ДАНИХ**

* 1. **Сутність даних як ресурсу підприємства**

У більшості сфер діяльності людини, що пов’язані з функціонуванням будь-якого бізнес-процесу, прийняття управлінських рішень базується на використанні й аналізі наявних ресурсів (фінансових, трудових, матеріальних, земельних тощо). Для того, щоб використання ресурсів було оптимальним, необхідно вміти ефективно управляти ними. Управління ресурсами, в загальному сенсі, означає здатність здійснювати над ними певне коло процедур: розподілу, планування, використання, заощадження. Успішне вирішення цих завдань потребує вивчення основних властивостей цих ресурсів – отримання певних ***даних.***

Таким чином, одним із основних видів ресурсів підприємства є **інформація**, тобто сукупність відомостей, які сприймаються з навколишнього середовища, передаються у навколишнє середовище або зберігаються всередині інформаційної системи. Інформація, яка необхідна для управління іншими ресурсами підприємства, витягується із зібраних, проаналізованих ***даних*** про властивості певного виду ресурсу. Можливість моделювання і збереження у пам’яті ЕОМ будь-яких даних (бухгалтерських та інвентарних документів, звітів, картотеки тощо) дозволяє розглядати дані в якості одного з видів ресурсу підприємства.

***Використання даних як ресурсу передбачає:***

* наявність уявлення про саме поняття сутності даних;
* вміння збирати та аналізувати дані про той чи інший об’єкт;
* визначення природи і властивостей даних, для чого необхідно знати як і з якою метою вони використовуються, де зберігаються і звідки беруться; таким чином, для керівництва даними необхідно зібрати найбільш вичерпну інформацію про них;
* наявність середовища для зберігання даних, які можуть бути надійно збережені тільки за умови чітких процедур планування, збирання та накопичення;
* можливість отримання доступу до ресурсу даних усюди, де потрібна інформація, яка є необхідною для управління іншими ресурсами, тобто інтегровані дані;
* наявність адміністратора даних.

***Історично*** розглядання даних як ресурсу стало можливим із появою зовнішніх носіїв інформації та розміщенням на них наборів даних.

Дані мають цілу низку ***специфічних властивостей:***

* вони є невичерпними, що забезпечує можливість їх багаторазового та багатоцільового використання;
* вони не відчужені під час продажу та обміну;
* обсяг даних постійно зростає;
* склад даних постійно змінюється, реагуючи на зміни інформаційних потреб підприємства;
* можливість оперативного пошуку даних;
* наявність засобів зберігання великого обсягу даних;
* наявність доступу до віддалених ресурсів (за допомогою веб- рішень та розвитку хмарних технологій);
* можливість фіксації на одному носії різних за характером даних. Саме дані сьогодні стають одним із найважливіших видів ресурсів сучасного підприємництва, а згодом їх значимість буде тільки зростати. Одне із свідчень феномену полягає в тому, що дані стають ***товаром***, сукупна вартість якого на ринку порівнянна з

вартістю традиційних ресурсів.

Великий обсяг інформації та можливість її подання в електронному вигляді призвели до появи розвиненого ринку інформаційних ресурсів, товарів і послуг. У даний час у багатьох країнах (включаючи й Україну) сформувався для цього національний ринок; можна побачити й явні ознаки відповідного світового ринку інформації. Цей ринок багато в чому подібний до ринку інших ресурсів, оскільки має певну номенклатуру, в якості якої на ньому виступають специфічні набори даних або послуги з їх аналізу, зберігання, трансформації тощо.

* 1. **Способи класифікації наборів даних**

Набір даних містить одне або декілька значень для кожного з окремих об’єктів, які називаються **елементарними одиницями**. У якості таких об’єктів можуть виступати люди, домогосподарства, міста, телевізійні приймачі або що завгодно, що представляє інтерес для вивчення.

Для кожного з об’єктів реєструють одну й ту саму ознаку (або ознаки). Ознака, яку реєструють (наприклад, вартість), називається **змінною**.

Існують чотири основних способи ***класифікації наборів даних***:

* за кількістю змінних (***2.2.1***);
* за типом вимірювання (***2.2.2***);
* залежно від того, чи є важливою впорядкованість набору в часовому зрізі (***2.2.3***);
* залежно від того, чи збиралася інформація спеціально для даного аналізу або дані збиралися раніше кимось іншим для своїх власних потреб (***2.2.4. – 2.3***).
	+ 1. За кількістю змінних розрізняють набори даних:

***одновимірний, двовимірний*** і ***багатовимірний***.

***Одновимірний набір даних*** (одна змінна) містить інформацію тільки про одну ознаку, що зареєстрована для кожного об’єкта. Одновимірний набір даних дозволяє визначити типове значення та характеристику мінливості даних, а також виділити специфічні особливості або проблеми в даних. Типовий приклад одновимірних даних наведений в табл. 2.1.

*Таблиця 2.1*

**Фінансовий результат діяльності підприємств України, січень-вересень 2017 р.**

|  |  |
| --- | --- |
| Галузь економіки | Фінансовий результат, млн. грн. |
| Оптова та роздрібна торгівля, ремонт автотранспортних засобів і мотоциклів | 31890,0 |
| Транспорт, складське господарство, поштова та кур’єрська діяльність | 11178,5 |
| Інформація та телекомунікації | 11003,6 |
| Фінансова та страхова діяльність | 49330,3 |

Можна надати й інші приклади одновимірних наборів даних.

1. Доходи окремих людей, які визначені в рамках перепису населення. Статистичне дослідження цих даних дозволило б виявити структуру доходів, їх середній рівень, варіацію доходів залежно від стажу, рівня освіти, географічного розташування тощо.
2. Тип пакування, який обрано членами фокус-групи в ході маркетингового дослідження. Подальший аналіз даних дозволить обрати найкращий варіант пакування нової продукції.

***Двовимірний набір даних*** (дві змінні) містить дві ознаки, значення яких реєструються для кожного об’єкта. Двовимірні дані, на додаток до інформації про кожну змінну як набір одновимірних даних, дозволяють вивчити зв’язок між двома змінними та передбачити значення однієї змінної на основі значення іншої. Прикладом може слугувати табл. 2.2, яка містить дані про фінансовий результат діяльності підприємств деяких галузей економіки України за три квартали 2017 р., а також зміну цього показника порівняно з відповідним кварталом попереднього року.

*Таблиця 2.2*

**Динаміка фінансового результату промислових підприємств України, січень-вересень 2017 р.**

|  |  |  |
| --- | --- | --- |
| Галузь економіки | Фінансовий результат, млн. грн. | Зміна фінансового результату у % до відповідного періоду 2016 р. |
| Оптова та роздрібна торгівля, ремонт автотранспортних засобів імотоциклів | 31890,0 | + 620,0 |
| Транспорт, складське господарство, поштова та кур’єрська діяльність | 11178,5 | + 16,3 |
| Інформація та телекомунікації | 11003,6 | + 720,4 |
| Фінансова та страхова діяльність | 49330,3 | + 18,1 |

Розглянемо ще декілька прикладів двовимірних наборів даних.

1. Дані за звітний період про витрати на виробництво продукції (***перша змінна***) та обсяг виготовленої продукції (***друга змінна***) великих підприємств однієї галузі, які виробляють однотипну продукцію. Статистичний аналіз таких двовимірних даних зміг би виявити взаємозв’язок між витратами підприємств та обсягом виробництва. Зокрема, в ході аналізу можливо було б виокремити постійні та змінні витрати, а також визначити ефективність функціонування підприємств галузі.
2. Дані опитування серед туристів міста, чи користувались вони послугами екскурсійної агенції (***перша змінна***; відповіді –

«так»/«ні» або 1/0) та чи пам’ятають вони рекламу цієї агенції (***друга змінна***, відповідь до якої можна записати таким самим чином). Такі дані, наприклад, допоможуть визначити ефективність існуючої рекламної кампанії підприємства.

***Багатовимірний набір даних*** (багато змінних) містить три або більше ознак, значення яких реєструються для кожного об’єкта. Багатовимірні дані дають можливість вивчити зв’язок попарно між змінними та передбачити значення однієї змінної на основі значення інших. У якості прикладу наведемо табл. 2.3, де деталізовано інформацію про функціонування галузей економіки України.

*Таблиця 2.3*

**Статистична характеристика діяльності промислових підприємств України, січень-вересень 2017 р.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Галузь економіки | Фінансо- вий результат, млн. грн. | Зміна фінансового результату у % до відповідного періоду 2016 р. | Рента- бельність діяльності,% | Кількість зайнятих працівни- ків,тис. осіб |
| Оптова та роздрібна торгівля, ремонт автотранспорт-них засобів і мотоциклів | 31890,0 | + 620,0 | 25,9 | 2182,3 |
| Транспорт, складське господарство, поштова та кур’єрськадіяльність | 11178,5 | + 16,3 | 6,6 | 912,4 |
| Інформація та телекомунікації | 11003,6 | + 720,4 | 19,3 | 282,8 |
| Фінансова та страхова діяльність | 49330,3 | + 18,1 | 3,2 | 209,5 |

Існують й інші приклади наборів багатомірних даних.

1. Заробітна плата робітника (***виділена змінна***) та характеристики, які на неї впливають (***інші змінні*** – рівень освіти, стаж роботи, категорія робіт і рівень продуктивності праці). Статистичне дослідження зможе виявити основні фактори, що впливають на рівень заробітної плати, наприклад, у середині підприємства.
2. Ціна на квартири однієї будівельної компанії (***виділена змінна***) та ***ряд змінних***, які на неї впливають (поверх, на якому

знаходиться квартира; кількість кімнат; загальна площа; наявність душової кабіни; кількість санвузлів тощо). Аналіз даних показав би принцип, згідно з яким компанія проводить оцінювання побудованих осель, а також виявити, яка комбінація характеристик найбільш суттєво підвищує ціну квартири.

* + 1. Друга характеристика набору даних, за типом вимірювання, розрізняє: ***категорії*** та ***числа***.

Якщо змінна містить інформацію про те, до якої з декількох нечислових категорій належить об’єкт, то вона називається ***якісною змінною*** або ***категорією***.

Існують два типи якісних змінних (категорій): ***порядкові***

(***ординальні***) та ***номінальні***.

Якщо якісні дані можна звичайним чином і змістовно впорядкувати, виділити перший, другий, третій, найкращий, найгірший показник, то мова йде про порядкові дані, для яких під час статистичного аналізу розраховують ***медіану***. У якості прикладу можна назвати посади для кожного члена управлінського персоналу (президент, віце-президент, чальник відділу, заступник начальника тощо), кваліфікації робітників (розряди), привабливість міст країни для туризму за результатами опитування та ін.

Якщо порядок у даних відсутній, то мова йде про ***номінальну якісну змінну***. Наприклад, номінальними якісними змінними є стать людини, віросповідання, цех (його назва або номер). При статистичному аналізі номінальних якісних змінних зазвичай визначають ***моду***.

Незважаючи на те, що часто значення якісної змінної можна записати за допомогою чисел, така змінна все ж залишається якісною, оскільки ці числа не мають будь-якої інтерпретації.

Значення змінних, які реєструються як числа, що мають змістовний сенс, називають ***кількісними даними***. Вони можуть бути ***дискретними*** та ***безперервними***.

***Дискретна*** – це така змінна, яка може набувати значень тільки з деякого списку певних чисел. Прикладами такої змінної є кількість комп’ютерів на підприємстві; число викликів майстра підприємства для післяпродажного обслуговування; число клієнтів, які звернулися на підприємство за певний проміжок часу; число випадків відмов обладнання протягом доби тощо.

***Безперервною*** вважають будь-яку змінну, яка не є дискретною. Вона може приймати значення з деякого проміжку. Прикладами

безперервної змінної є заробітна плата робітників підприємства, стаж їх роботи, прибуток підприємств промисловості тощо.

* + 1. Залежно від того, чи є набір даних часовим рядом чи ні, розрізняють: ***часові ряди*** (***ряди динаміки***) і ***дані одного часового зрізу***.

Якщо послідовність запису даних має змістовний сенс (хронологію), то відповідний набір даних є часовим рядом (наприклад, рис. 2.1).

160

140

120

100

80

60

40

20

0

135,3

117,3

115,7 120,8

124,1

123,3 123

114,1

114,3 118,7 114,2

131,8 125,4

135,7

116,5

105

100,9 105,7 111,1

109,5

100,3 101,7

Індекси цін виробників промислової продукції, % до попереднього року

***Рис. 2.1. Динаміка індексу цін виробників промислової продукції України, 1996-2017 рр.***

Розглянемо ще декілька прикладів часових рядів.

1. Ціна на цукор за останні 50 років з урахуванням інфляції. Якщо припустити, що зміни, які були в минулому, будуть повторюватися й у майбутньому, то можна використовувати ці тренди для довгострокового прогнозування.
2. Місячні обсяги продажу цифрової техніки за останні десять років. Статистичний аналіз цих даних допоможе виявити зміну в їх структурі під впливом розвитку нових технологій, а також наявну сезонну особливість із піком протягом грудня-січня.

Якщо послідовність запису даних не важлива, то відповідний набір містить дані про ***однин часовий зріз***. Слова «про один часовий

зріз» у даному випадку означають лише те, що немає ніякого впорядкування у часі, а є лише інформація по деяких об’єктах у певний момент часу (свого роду «моментальний знімок»). Наприклад, курс долара в деяких обмінних пунктах міста на сьогоднішню дату, виміряна тривалість сну перед іспитом обраної групи студентів тощо. Аналіз часових рядів є більш складним, ніж аналіз даних по одному часовому зрізу, оскільки вимагає ретельного обліку порядку

спостережень.

* + 1. Залежно від того, чи збиралася інформація спеціально для цього аналізу або дані збиралися раніше кимось іншим для своїх потреб, бувають дані: ***первинні*** та ***вторинні***. Ми розглянемо їх більш детально у наступному питанні.
	1. **Джерела даних**

Достатньо важливим питанням дослідження даних виступають джерела їх походження. Вибір цих джерел здійснюється підприємством залежно від їх надійності, доступності, вартості та важливості для економічної діяльності. Якщо підприємство здійснює цей збір самостійно (або доручає іншим особам за відповідним договором), то в результаті отримують ***первинні дані***. Дані, що зібрані іншими підприємствами або людьми для власних цілей, називають ***вторинними***.

Первинними даними для підприємства можуть слугувати:

* ***безпосереднє спостереження***, яке здійснюється шляхом реєстрації даних та їх ознак на основі підрахунку, зважування, обмірювання, зняття показників приладів, реєстрації цін і обсягу проданої продукції, інвентаризації залишків товарно-матеріальних цінностей на складах підприємства тощо;
* ***документальний облік*** – спосіб спостереження, при якому використовують різні документи первинного обліку підприємств. Наприклад, документальне дослідження постачальника: якості його обладнання, сировини, рівня браку, динаміки цін на продукцію. Таке дослідження зазвичай проводить відділ постачання або технічного забезпечення. Безпосереднє спостереження та документальний облік дають найбільш достовірні дані;
* ***опитування***, в ході якого джерелом даних є відомості, які дають опитувані особи. Наприклад, дані опитування відділу

маркетингу серед споживачів продукції підприємства з приводу розширення асортименту.

Основними ***перевагами*** первинних даних є їх унікальний, ексклюзивний характер, недоступність конкурентам та іншим стороннім особам, можливість отримання якісної інформації, що характеризує ставлення покупців до товарів ті їх якості, цін, самого підприємства та його бренду, можливість виявити мотиви поведінки покупців тощо.

У той же час, збір первинних даних має низку суттєвих ***недоліків***, як, наприклад, великі витрати фінансових і трудових ресурсів, відносно тривалий характер отримання та обробки інформації, можлива неточність результатів, що виникає у випадку порушення правил збору первинної інформації. Тому до збору первинних даних прибігають тоді, коли необхідно отримати відомості для вирішення конкретного завдання, впевнитися у повноті, актуальності та достовірності вторинних даних, або отримати недоступну конкурентам інформацію з метою отримання конкурентних переваг.

Розглянемо детальніше ***джерела вторинних даних***. Зазвичай вони включають як внутрішні, так і зовнішні публікації та документи, які дозволяють отримати різноманітну інформацію: про стан розвитку галузі, про динаміку цін, про рух робочої сили, напрямки науково-дослідної діяльності та нових технології, нові джерела енергії та загальний стан її споживання тощо.

Усі ***зовнішні джерела*** вторинних даних можна розділити на чотири основні групи (табл. 2.4).

В якості ***внутрішніх джерел*** вторинних даних підприємства розглядають власні документи та звіти попередніх досліджень, наприклад, дані про запаси продукції, інформація про клієнтів та обсяги їх закупівель, сезонний та географічний розподіл продажів, дані про прибутки та збитки підприємства в цілому та його окремих підрозділів тощо. Ці дані є достовірними, важливими та ексклюзивними. Однак, попередні дані можуть не враховувати всі потреби маркетингових відділів підприємства; вони можуть бути дубльованими, а доступ до них може бути обмежений внаслідок внутрішніх перепон. Основним призначенням внутрішніх джерел даних є оцінка прийнятих маркетингових рішень, рівня обслуговування, якості продукції, ретроспективний аналіз ринку тощо.

*Таблиця 2.4*

**Класифікація джерел даних**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Тип джерела | Приклад | Переваги | Недоліки | Основне призначення |
| Офіційні | - специфічні | - високий | - неспівставність | - більш чітке |
| видання та | видання; | ступінь | даних; | визначення |
| документи | - економічна та | достовірності | - неповнота | проблеми |
|  | технічна література; | та | інформації; | дослідження; |
|  | - річні звіти органі- | доступності; | - дані можуть бути | - аналіз ринку; |
|  | зацій і підприємств; | - невеликі | використані | - пошук даних |
|  | - документи | витрати. | конкурентами; | для більш |
|  | професійних |  | - швидке старіння | детального |
|  | асоціацій; |  | даних. | розуміння |
|  | - засоби масової |  |  | первинної |
|  | інформації. |  |  | інформації. |
| Неофіційні | - спілкування з | - можливість | - неструктурова- | - збір даних про |
| джерела | постачальниками, | отримання | ність даних; | клієнтів, |
|  | клієнтами, | ексклюзивних | - трудомісткість | партнерів, |
|  | посередниками, | даних. | організації | конкурентів або |
|  | торговим |  | контактів; | інших осіб, які |
|  | персоналом тощо. |  | - нерегулярність | мають суттєвий |
|  |  |  | контактів; | вплив на |
|  |  |  | - сумнівна | діяльність |
|  |  |  | достовірність даних; | підприємства. |
|  |  |  | - багато зайвих |  |
|  |  |  | даних. |  |
| Специфічні | - придбання товарів | - відносна | - неповнота даних; | - збір даних про |
| джерела | конкурентів або | доступність | - зайві дані; | переваги |
| інформації | використання їхніх | даних; | - висока | конкурентних |
|  | послуг; | - конкретна | трудомісткість; | товарів, послуг, |
|  | - відвідування | спрямованість | - високий рівень | технологій. |
|  | виробництва на | отриманої | фінансових витрат. |  |
|  | підприємствах тієї ж | інформації. |  |  |
|  | галузі; |  |  |  |
|  | - інші специфічні |  |  |  |
|  | акції. |  |  |  |
| Синдика- | - дані про товарні | - високий | - дані можуть бути | - відстеження |
| тивні | ринки; | рівень якості | куплені | зміни цінностей |
| джерела | - бази даних; | даних; | конкурентами; | споживачів; |
| даних – дані | - відомості про | - регулярність | - неспівставність | - сегментація |
| досліджень | споживачів. | оновлення | даних різних | ринку; |
| спеціалізо- |  | даних; | досліджень; | - дослідження |
| ваних фірм |  | - вартість | - неможливість | ринкових |
|  |  | даних нижча | впливати на склад | тенденцій; |
|  |  | за | даних дослідження. | - відстеження |
|  |  | індивідуальні |  | наявності |
|  |  | маркетингові |  | товарів. |
|  |  | дослідження. |  |  |

Усі джерела даних можна також розділити на два великих класи:

***документальні*** та ***електронні*** (рис. 2.2).

***Рис. 2.2. Класифікація джерел даних залежно від їх носія***

Серед цього переліку відмітимо наукові журнали: згідно даних статистичних досліджень вони складають близько 70% усіх наукових документів, а близько 80% фахівців різних рівнів вважають науковий журнал основним джерелом науково-технічної інформації.

Однак особливої уваги заслуговують електронні дані, що знаходяться в мережі Інтернет або у специфічних базах. На кінець 2017 р. в мережі знаходилося біля 1,3 млрд. сайтів, що майже в 2 рази більше, ніж 5 років тому. Протягом дня тільки пошукова система Google робить біля 6 млн. запитів, у рамках того ж дня офіційно продається більше 600 тис. комп’ютерів і 3,5 млн. смартфонів. Все це свідчить про зростаючу важливість електронних даних.

Класифікація Інтернет-ресурсів даних наведена в табл. 2.5.

*Таблиця 2.5*

***Класифікація Інтернет-ресурсів залежно від способу наведення даних***

|  |  |  |
| --- | --- | --- |
| **Група** | **Характеристика** | **Види** |
| Ресурси прямого доступу | Основним типом інформаційних ресурсів у Мережі є сторінка гіпертексту та (або) інформація у довільній формі: графічній, звуковій, відео та ін. | Офіційні сторінки організацій; персональні сторінки; різноманітні портали. |
| Ресурси вторинних даних | Ресурси, що містять огляди Інтернету, тематичні каталоги. Доступ до потрібної інформації здійснюється через посилання на інші сайти. | Каталоги посилань; торрент-трекери тощо. |
| Ресурси інтерактивних даних | Ресурси, що дають можливість запитувати та розміщувати власну інформацію. | Форуми; блоги; чати; телеконференції та ін. |

Залежно від типу даних електронні ресурси даних можна також поділити на ***тематичні*** (що містять найбільш цікаву інформацію, присвячену певній галузі), ***наукові публікації*** (електронний варіант статей, збірників тощо), ***рекламні*** (більшість web-сайтів комерційних компаній, представлених у Мережі, які мають тією чи іншою мірою рекламний характер; даний вид інформації може також допомогти в пошуках потрібних даних), ***довідкові*** (різноманітні довідкові матеріали, нормативні бази) та ***новини*** (ця інформація легко доступна в Мережі, являє собою вид необроблених даних, які є цінними не стільки самі собою, скільки в контексті інших даних або в динаміці, і часто потребують подальшої обробки та аналізу).

Стосовно банків та баз даних звернемо увагу, що тільки в ***Україні*** для підприємців працює більше 140 відкритих джерел цього типу. До них відносяться:

* єдиний державний реєстр юридичних осіб та фізичних осіб- підприємців (irc.gov.ua/ua/Poshuk-v-YeDR);
* пошук виданих всіляких ліцензій (irc.gov.ua/ua/Poshuk-v-YeLR);
* інформація з фондового ринку України (smida.gov.ua/db);
* перевірка дипломів (osvita.net/checkdoc);
* реєстр зареєстрованих товарних знаків і послуг (uipv.org/ua/bases2);
* державна служба фінансового моніторингу (sdfm.gov.ua/articles) тощо.

До ***міжнародних баз даних***, які можуть цікавити національні підприємства, можна віднести базу ***LinkedIn***, яка обіймає велику кількість підприємств та потенційних працівників; ***Enterprise Europe Network***, що допомагає знайти партнерів по всій Європі і стати на шлях інноваційного розвитку; ***International Network, Local Hubs***, що спрямована на розвиток громад і відкриття нового бізнесу.

* 1. **Типи розподілу даних у сукупностях**

Перш ніж приступати до аналізу наявних матеріалів (вихідних даних), слід визначити їх розподіл, оскільки різні розподіли даних описуються за допомогою різних прийомів статистики. Визначення ***виду розподілу*** даних допоможе в майбутньому вибрати найбільш прийнятний статистичний критерій для перевірки гіпотез. Крім того, якщо розподіл далекий від нормального, то не можна автоматично робити узагальнюючі висновки про сукупність у цілому.

У більшості досліджень у першу чергу проводиться перевірка на

«***нормальність розподілу***».

Нормальний розподіл (***розподіл Гауса***) являє собою пологу криву у формі дзвону без випадкових відхилень (рис. 2.3).

Така крива є ідеальним набором даних, у якому більшість чисел сконцентровані в середній частині діапазону значень, а решта значень із згасанням симетрично розташовані по обидва боки від вершини дзвона.

Перевірка нормальності розподілу може проводитися ***трьома способами***:

* за допомогою описової статистики (на комп’ютері: середня ( *х* ), медіана (*Ме*), дисперсія ( 2 ), стандартне відхилення ( ), мінімальне і максимальне значення, коефіцієнти асиметрії (*А*) та ексцесу) (*Е*));
* графічно;
* з використанням статистичних критеріїв.



***Рис. 2.3. Крива нормального розподілу (розподіл Гауса)***

Розподіл може бути ***одновершинним***, ***двовершинним***

(***бімодальним***) і більше.

Серед одновершинних розподілів є ***симетричні*** та ***асиметричні***

(скошені), ***гостро- і плосковершинні***.

У симетричному розподілі рівновіддалені від центру значення змінної мають однакові частоти, в асиметричному – вершина розподілу зміщена. Якщо вершина зміщена вліво, маємо правосторонню асиметрію, і навпаки (рис. 2.4).

У симетричному розподілі характеристики центру – середня, мода, медіана – мають однакові значення, в асиметричному між ними існують певні відмінності. У випадку правосторонньої асиметрії

*х*  *Ме*  *Мо* , а у випадку лівосторонньої, навпаки, *х*  *Ме*  *Мо* . Чим

більшою є асиметрія, тим більшим є відхилення *х*  *Мо* .

Для оцінки симетричності розподілу розраховують коефіцієнт асиметрії:

*х*  *Мо*

*А*  

(2.1)

Цей коефіцієнт характеризує напрямок і міру скошеності (зміщення) в середині розподілу: при правосторонній асиметрії *А* > 0, при лівосторонній – *А* < 0. Існує й більш детальна градація:

* якщо | *А* | < 0,25, то асиметрія вважається незначною;
* якщо 0,25 < | *А* | < 0,50, то асиметрія вважається помірною;
* якщо | *А | >* 0,50, то асиметрія вважається суттєвою.



***Рис. 5.4. Правосторонній та лівосторонній асиметричний розподіл***

Одновершинні розподіли можуть бути ***гостро-*** і ***плосковершинними.*** Для їх оцінки розраховують коефіцієнт ексцесу, який ґрунтується на стандартизованому моменті 4-го порядку:

*Е*   : 4

4

(2.2)

де  4 – момент четвертого порядку, який розраховується

наступним чином:

4 

(*x*  *x*)4  *f*

 *f*

(2.3)

У симетричному, близькому до нормального, розподілі *Е*  3 ;

при гостровершинному *Е*  3 ; при плосковершинному *Е*  3 .

* 1. **Основні методи зберігання даних**

Дані є похідними від інформації, вони являють собою запис у відповідному коді спостереження, факт, об'єкт, що є придатним для інтерпретації, передачі, обробки та отримання інформації. Отже, поняття інформації передбачає існування матеріального носія та форми, способу, яким вона пов'язана зі своїм носієм. Таким чином, **дані** – це інформація, яка представлена у вигляді, що дозволяє автоматизувати її збір, зберігання та подальшу обробку людиною або інформаційним засобом [17, c. 3].

Відомі два способи організації масивів даних: ***файлова організація*** та організація у формі ***баз даних***. З точки зору прикладної

програми **файл** – це іменована область зовнішньої пам’яті, на яку можна записувати дані або зчитувати з неї. Система управління файлами бере на себе розподіл зовнішньої пам’яті, відображення назв файлів у відповідні адреса зовнішньої пам’яті та забезпечення доступу до даних.

Перш за все файли застосовуються для зберігання текстових даних: фінансових документів, звітів тощо. Такі файли зазвичай утворюються та модифікуються за допомогою різноманітних текстових та графічних редакторів (***MS Word, MS Excel, Adobe Photoshop***).

Однак із розвитком цифрових технологій традиційна файлова система зберігання перестає задовольняти наростаючий потік даних та вимоги щодо їх обробки, що призводить до появи іншої форми організації даних – баз даних (***БД***).

У широкому сенсі **база даних** – це набір постійно збережених даних, що використовуються прикладними програмними системами підприємства.

***Перша відмінність*** БД від файлової організації даних полягає в тому, що БД зберігає дані в структурованому і спеціально організованому вигляді, що виключає спільне зберігання різнорідної інформації або її дублювання. Наприклад, у текстових файлах на порядок розміщення даних не накладається скільки-небудь серйозних обмежень, і дані можуть бути розташовані довільно. Деякі дані можуть неодноразово повторюватися. У електронних таблицях дані по рядках і стовпчиках розташовуються вже впорядковано, але все ще досить довільно. Особа сама вирішує в момент створення таблиці, як краще та наочніше розмістити дані. І лише в БД структура даних суворо фіксована та визначається стандартом використовуваної моделі даних.

***Друга відмінність*** полягає в тому, що БД зберігає не тільки дані, але й описання їх структури (мета-дані). ***Мета-дані*** зберігаються окремо від самих даних у так званому словнику (системному каталозі) даних. Таким чином, будь-яка БД може працювати з різними наборами даних, оскільки структура їх зберігання доступна при читанні даних. У файлової системи спосіб зберігання даних – справа кожної програми, що здійснює зберігання та обробку даних. Структура даних вбудована в програму доступу і не може бути прочитана іншими програмами.

Системне програмне забезпечення керує всіма компонентами обладнання підприємства і забезпечує доступ до нього всіх інших додатків, що працюють на комп'ютері. Прикладами системного програмного забезпечення БД можуть виступати ***Windows, Linux, UNIX, MVS, Mac OS, Open Solaris*** тощо.

Саме створення баз даних стало першим кроком **інформатизації підприємств** – комплексу робіт із створення єдиного інформаційного середовища підприємства, яке включає сукупність технічних, програмних, телекомунікаційних засобів, що дозволяють застосовувати в процесах управління підприємством нові інформаційні технології, здійснювати збір, зберігання та обробку даних.

**Питання для самоконтролю**

1. Що дало підставу вважати дані ресурсом підприємства?
2. Назвіть специфічні властивості даних.
3. Назвіть чотири основні класифікації наборів даних.
4. Що таке одновимірні, двовимірні та багатовимірні дані?
5. Що можна робити із багатовимірним набором даних?
6. Охарактеризуйте класифікацію наборів даних залежно від критерію часу.
7. Що є складнішим для аналізу: ряди динаміки чи ряди одного часового зрізу?
8. Дайте визначення первинних даних.
9. Як можна класифікувати первинні дані?
10. Дайте визначення вторинних даних.
11. Яка відмінність між первинними та вторинними даними?
12. Які існують способи класифікації вторинних даних?
13. Наведіть класифікацію Інтренет-ресурсів.
14. Наведіть приклади національних і міжнародних баз даних.
15. Який розподіл даних можна назвати нормальним?
16. Назвіть способи перевірки розподілу на нормальність.
17. Наведіть класифікації розподілу даних.
18. Що характеризує коефіцієнт асиметрії і які значення він приймає?
19. Якими можуть бути одновершинні розподіли?
20. Що характеризує коефіцієнт ексцесу і які значення він має?
21. Що таке база даних?
22. Для чого використовують файлову систему збереження даних?
23. Що таке інформатизація підприємств?