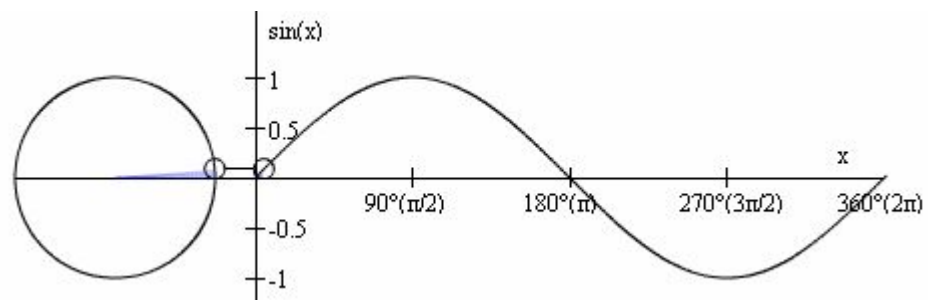




# Багатовимірна статистика

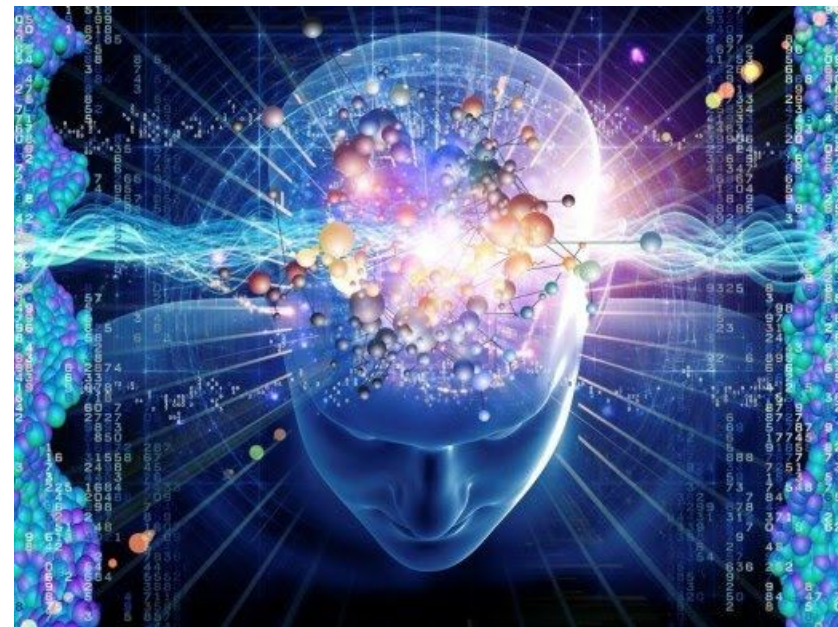
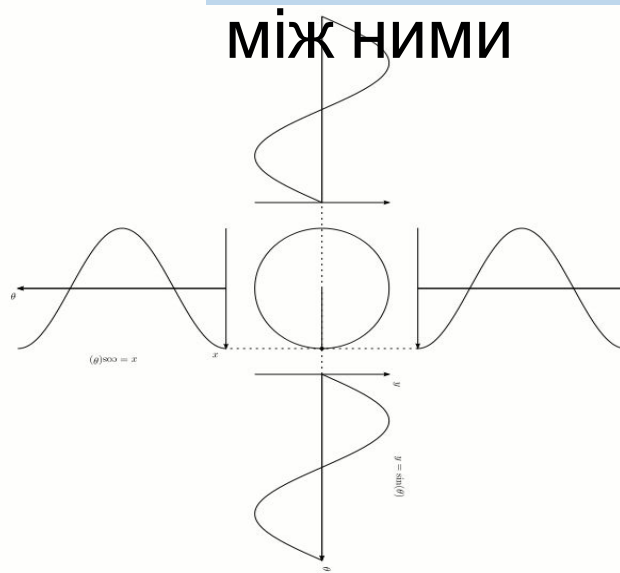




# Багатовимірна статистика - це підрозділ статистики, який охоплює одночасне спостереження та аналіз більш ніж однієї змінної

Застосування багатовимірної статистики називають **багатовимірним аналізом**

Основна задача багатовимірної статистики — це зрозуміти досліджувані явища та процеси, оцінити взаємозв'язок **МІЖ НИМИ**

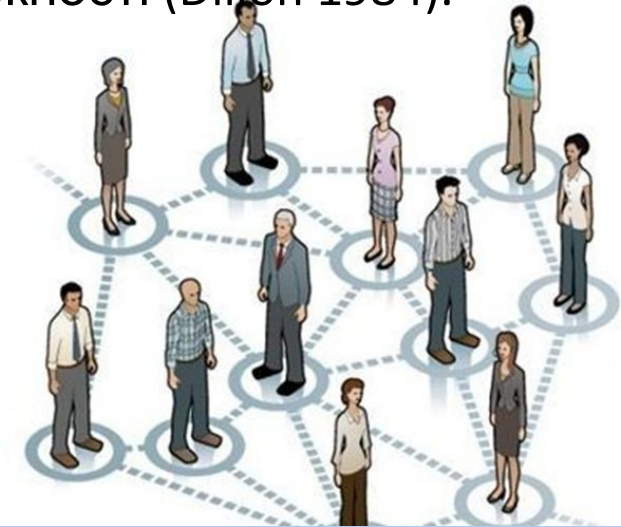


# Методи багатовимірної статистики



Традиційна класифікація багатовимірних статистичних методів, запропонована Кендаллом, базується на концепції залежності між змінними (Kendall 1957). Якщо інтерес зосереджується на асоціації між двома наборами змінних, де один набір є реалізацією залежної змінної (або змінних), а інший набір - реалізацією ряду незалежних змінних.

Якщо зацікавленість зосереджується на взаємній асоціації між усіма змінними без різниці між типами змінних, використовуються багатовимірні методи взаємозалежності (Dillon 1984).



*Практична* реалізація багатовимірної статистики для конкретної проблеми може включати декілька типів одновимірного і багатовимірного аналізів з метою зрозуміти відношення між змінними та їх доцільність для досліджувачого фактора



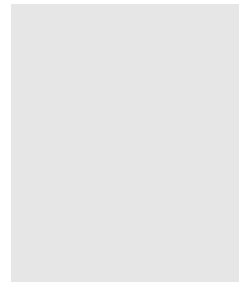
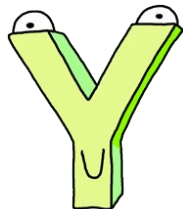
# Класифікація багатовимірних статистичних методів:

методи для результуючих змінних:

- **одна результуюча змінна:** крос-табуляція, дисперсійний аналіз та коваріаційний аналіз, множинна регресія, дискримінантний аналіз

- **декілька результуючих змінних:** множинний дисперсійний аналіз та коваріаційний аналіз, аналіз канонічної кореляції, множинний дискримінантний аналіз

YYYYYY





# Типи аналізів

- Багатофакторний дисперсійний аналіз розширює дисперсійний аналіз, щоб покрити випадки, коли є більш ніж одна залежних змінних, для одночасного їх аналізу.
- Багатовимірний регресійний аналіз намагається визначити формулу, яка зможе описати як елементи векторних змінних міняються при зміні інших елементів. При лінійних співвідношеннях регресійний аналіз має за основу форми загальної лінійної моделі.
- Аналіз головних компонент створює новий набір ортогональних змінних, які містять ту ж інформацію, що й вихідний набір. Таким чином повертаючи осі змінних, створюється новий набір ортогональних осей, впорядкованих так, що вибіркова дисперсія даних вздовж кожної координати була максимальною, за умови ортогональності всіх попередніх координат



# Типи аналізів

- Факторний аналіз схожий на аналіз головних компонент, але дозволяє користувачеві вилучати вказану кількість штучних змінних, яка менша за кількість змінних всього набору, залишаючи непоясненні змінні як помилку. Вилучені змінні відомі як приховані змінні або фактори, припускається що кожна з них пояснює коваріацію в групі досліджуваних змінних.
- Канонічний аналіз кореляції знаходить лінійні зв'язки між двома наборами змінних, це узагальнена версія парної кореляції.
- Аналіз надмірності схожий на канонічний аналіз кореляції, але на відміну від нього дозволяє користувачеві отримати вказану кількість штучних змінних від одного набору незалежних змінних, які пояснюють як можливо багато змінних в іншому наборі незалежних змінних. Це — багатовимірний аналог регресу.



# Типи аналізів

- Аналіз відповідностей або взаємне усереднення, знаходить штучні змінні, що узагальнюють початковий набір. Основна модель приймає відмінності  $\chi^2$ -квадрат серед записів (випадки). Існує канонічний (або обмежений) аналіз відповідностей: робить висновки про спільні змінні двох наборів (схоже як канонічний аналіз кореляції).
- Багатовимірне шкалювання (масштабування) включає в себе різні алгоритми для визначення набору штучних змінних, які найкращим чином зображують попарні відстані між записами. Первісний метод — це аналіз головних координат.
- Дискримінантний аналіз або канонічний аналіз варіаційної величини намагається встановити, чи може набір змінних використовуватись для відокремлення двох або більше груп випадків.



# Типи аналізів

- Лінійний дискримінантний аналіз обчислює лінійного предиктора двох наборів нормально розподілених даних з метою класифікації нових спостережень.
- Кластеризація систем призначають об'єкти в групи (так звані кластери), таким чином що об'єкти з одного кластера більш схожі один на одного. Ніж об'єкти з різних кластерів.
- Рекурсивне розбиття створює дерево рішень, яке намагається правильно класифікувати елементи множини на основі дихотомічна залежна змінна.
- Штучні нейронні мережі розширюють регресію та метод кластеризації до нелінійних багатовимірних моделей.





# Багатовимірний аналіз

є сукупністю формалізованих статистичних методів, які базуються на представленні вихідної інформації у багатовимірному геометричному просторі та дозволяють визначати неявні (латентні), але об'єктивно існуючі закономірності у структурі та тенденціях розвитку досліджуваних соціально-економічних явищ та процесів.

## Основними задачами БСА є наступні:

- статистичне дослідження залежностей між ознаками;
- класифікація елементів (об'єктів або ознак);
- зниження виміру простору ознак і відбір найбільш інформативних показників.





# Програмне забезпечення

Якщо потрібний потужний, загальноvizнаний пакет з простим і зрозумілим навіть для початківців інтерфейсом, то краще скористатися SPSS.



Для початківців і професіоналів, яким потрібна підказка і розвинена документація російською мовою, можна рекомендувати STATISTICA. Це потужне ПЗ з професійними можливостями. Переваги використання пакету STATISTICA:

- краще представлення 3D графіків та ілюстрацій;
- дане ПЗ є сертифікованим та надійним засобом обробки статистичних даних;
- дозволяє використовувати сучасні та надійні методи покладені в розробку ПЗ.



Для невибагливих користувачів, які обмежуються у своїх дослідженнях стандартними статистичними методами можна рекомендувати Excel



Для ознайомлення із статистичною обробкою даних у навчальному процесі рекомендуємо ознайомитись з такими ПЗ: універсальними - Minitab, MatLab, STATISTICA, STATA, S-PLUS, STATGRAPHICS, Stadia, SYSTAT, SPSS, Excel; професійними - BMDP, SAS; спеціалізованими- DATASCOPE, BioStat, MESOSAUR.



# Рекомендації щодо застосування методів аналізу

Кореляційний і регресійний аналіз

Як зміниться обсяг продажу, якщо витрати на рекламу скоротяться на 10%?

Які характеристики товару цікаві даній групі споживачів? Які характеристики товару можна об'єднати в один фактор?

Яка буде ціна на послугу в наступному році?

Дискримінантний аналіз

Як розділити споживачів на групи всередині кластера?

Які характеристики працівників служби маркетингу найбільш істотні для їх поділу на лідерів і невдах?

Чи можна позичальника, враховуючи вік, дохід, освіту, вважати досить надійним для видачі кредиту?



# Рекомендації щодо застосування методів аналізу

## Факторний аналіз

Чи можна скоротити безліч характеристик, які клієнти компанії вважають важливим, до невеликої кількості?  
Як можна описати різні компанії з точки зору цих факторів?

## Кластерний аналіз

Чи можна клієнтів розділити на групи за їхніми потребами?  
Чи має компанія різні типи клієнтів?  
Як можна класифікувати клієнтів за видами вкладів, які їх цікавлять?

## Багатовимірне шкалювання

Наскільки продукт або компанія відповідає "ідеалу" клієнта?  
Який імідж має компанія?  
Як змінилося ставлення клієнта до продукту протягом ряду років?