

Факторний аналіз

Підготували: Смирнов Євген, 626

Моїсеєнко Анна, 627

сутність

- **Факторний аналіз** - багатовимірний метод, застосовуваний для вивчення взаємозв'язків між значеннями змінних.
- **Факторні ознаки (незалежні)** - ті ознаки, які характеризують причину.
- **Результативні ознаки (залежні)** - ті ознаки, які характеризують наслідок.
- **факторна система** - сукупність факторних і результативних ознак, які знаходяться в одній причинно-наслідкового зв'язку.
- **Модель факторної системи:**

$$Y = f(x_1, x_2, x_3, \dots, x_n)$$

- Цели
 - Определить взаимосвязи между переменными
 - Сокращение числа переменных, необходимых для описания

ВИДИ ЗВ'ЯЗКІВ

- **функціональний зв'язок** - ця така зв'язок, при якій кожному значенню фактора (факторної ознаки) відповідає цілком певний невинпадкове значення узагальненого показника (результативної ознаки).
- **стохастична зв'язок** - такий зв'язок, при якій кожному значенню фактора (факторної ознаки) відповідає безліч значень узагальненого показника (результативної ознаки).

- Методы
 - Детерминированный ФА
 - Метод цепных подстановок
 - Метод абсолютных и относительных разниц
 - Балансовый метод
 - Индексный метод
 - Стохастический ФА
 - Корреляционный метод
 - Регрессионный метод
 - Дисперсионный метод
 - Метод кластерного анализа

Типы факторного анализа

- ФА
 - По порядку
 - Прямой
 - Обратный
 - По периодам
 - Ретроспективный
 - Перспективный
 - По ступеням
 - Одноступенчатый
 - Многоступенчатый
 - По состоянию
 - Статический
 - Динамический

Умови застосування факторного аналізу

- Всі ознаки повинні бути кількісними.
- Число спостережень має бути не менше ніж в два рази більше числа змінних.
- Вибірка повинна бути однорідна.
- Вихідні змінні повинні бути розподілені симетрично.
- Аналіз здійснюється за коррелируючим змінним.

- Сферы применения
 - Психология
 - Политология
 - Экономика
 - Статистика
 - Нейрофизиология
 - Социология

приклад

Загальний фонд відпрацьованого часу залежить від декількох факторів.
Дану залежність можна представити у вигляді формули:

$$\text{ФРВ} = \text{Ч} * \text{Д} * \text{П}$$

| Показатели | План | Фактически | Отклонение от плана (+,-) | | Условное обозначение |
|--|------|------------|---------------------------|-------|----------------------|
| | | | Абсолютное | В % | |
| А | 1 | 2 | 3 | 4 | 5 |
| Численность работников | 1000 | 900 | -100 | -10,0 | Ч |
| Количество человеко/дней работы | 250 | 234 | -16 | -6,4 | ЧД |
| Количество дней работы (стр.2:стр.1) | 250 | 260 | +10 | +4,0 | Д |
| Средняя продолжительность рабочего дня, час (стр.5: стр.2) | 8 | 7,8 | -0,2 | -2,5 | П |
| Общий фонд отработанного времени, тыс. чел./ час | 2000 | 1825 | -175 | -8,8 | ФРВ |

Метод ланцюгових підстановок

- **мета:** визначити вплив окремих факторів на зміну величини результативного показника **ШЛЯХОМ** поступової заміни *базисної величини* кожного факторного показника в обсязі результативного показника на *фактичну* в звітному періоді.

розраунок

| Показатели | 1 расчет | 2 расчет | 3 расчет | 4 расчет | Влияние |
|---------------------------------|-----------|---------------|---------------|-----------|--|
| А | 1 | 2 | 3 | 4 | 5 |
| Численность работников | План 1000 | Факт 900 | Факт 900 | Факт 900 | Фактор влияния численности работников = 2 расчет – 1 расчет = $18000 - 2000 = -200$ тыс. ч/г |
| Количество дней работы рабочего | План 250 | План 250 | Факт 260 | Факт 260 | Фактор влияния количества дней работы работника = 3 расчет – 2 расчет = $1872 - 1800 = +72$ |
| Продолжительность рабочего дня | План 8 | План 8 | План 8 | Факт 7,8 | Фактор влияния продолжительности рабочего дня – 4 расчет – 3 расчет = $1825 - 1872 = -47$ тыс. чел./г. |
| Общий фонд рабочего времени | План 2000 | Условный 1800 | Условный 1872 | Факт 1825 | Совокупное влияние – $200 + 72 - 47 = -175$ тыс. чел./г |

Метод абсолютних різниць

- Величина впливу факторів розраховується множенням абсолютного приросту на базову величину факторів, які знаходяться праворуч від нього, і на фактичну величину факторів, які розташовані зліва від нього в моделі.

Вплив чисельності працюючих: $\Delta\text{ФРВ}_4 = (\text{ЧФ}-\text{Чпл}) * \text{Дпл} * \text{ППЛ} = (900-1000) * 250 * 8 = -200 \text{ тис. Ч / г.}$

Вплив відпрацьованих людино-днів: $\Delta\text{ФРВ}_д = \text{ЧФ} * (\text{Дф}-\text{Дпл}) * \text{ППЛ} = 900 * (260-250) * 8 = + 72 \text{ тис. Ч / г}$

Вплив тривалості робочого дня:

$\Delta\text{ФРВ}_п = \text{ЧФ} * \text{Дф} * (\text{Пф}-\text{Ппл}) = 900 * 260 * (7,8-8) = - 46,8 \text{ тис. Ч / г} \approx 47$

Метод відносних різниць

- Для розрахунку впливу першого фактора необхідно базисну величину результативного показника помножити на відносний приріст першого фактора.
- Для розрахунку впливу 2-го фактора, потрібно до планової величиною результативного показника додати зміну його за рахунок першого фактора і потім отриману суму помножити на відносний приріст другого.

Етапи стохастичного моделювання

- Етап 1 - Якісний аналіз
- Етап 2 - Попередній аналіз модельованої сукупності
- Етап 3 - Побудова регресійної моделі
- Етап 4 - Оцінка адекватності моделі
- Етап 5 - Інтерпретація і практичне використання моделі

кореляційний аналіз

кореляційний аналіз - це кількісний метод визначення тісноти і напрямку взаємозв'язку між вибірковими змінними величинами.

метою *кореляційного аналізу* є оцінка тісноти зв'язку між ознаками.

тіснота зв'язку *кількісно* виражається величиною коефіцієнтів кореляції.

Практична реалізація кореляційного аналізу включає наступні етапи:

- 1) постановка завдання і вибір ознак;
- 2) збір інформації та її первинна обробка (угруповання, виключення аномальних спостережень, перевірка нормальності одновимірного розподілу);
- 3) попередня характеристика взаємозв'язків (аналітичні угруповання, графіки);
- 4) усунення мультиколінеарності (взаємозалежності факторів) і уточнення набору показників шляхом розрахунку парних коефіцієнтів кореляції;
- 5) дослідження факторної залежності і перевірка її значущості;
- 6) оцінка результатів аналізу і підготовка рекомендацій щодо їх практичного використання.

Кореляційний аналіз:

1. **парна кореляція** - зв'язок між двома ознаками (результативним і факторним або двома факторними).
2. **приватна кореляція** - залежність між результативним і одним факторингу ознаками при фіксованому значенні інших факторних ознак.
3. **множинна кореляція** - залежність результативного і двох або більше факторних ознак, включених в дослідження.

Види кореляційних зв'язків:

- **По формі** кореляційний зв'язок може бути прямолінійною або криволінійною.
- **У напрямку** кореляційний зв'язок може бути позитивною ("прямий") і негативною ("зворотній").
- **за силою** кореляційний зв'язок визначається шкалою Чеддока

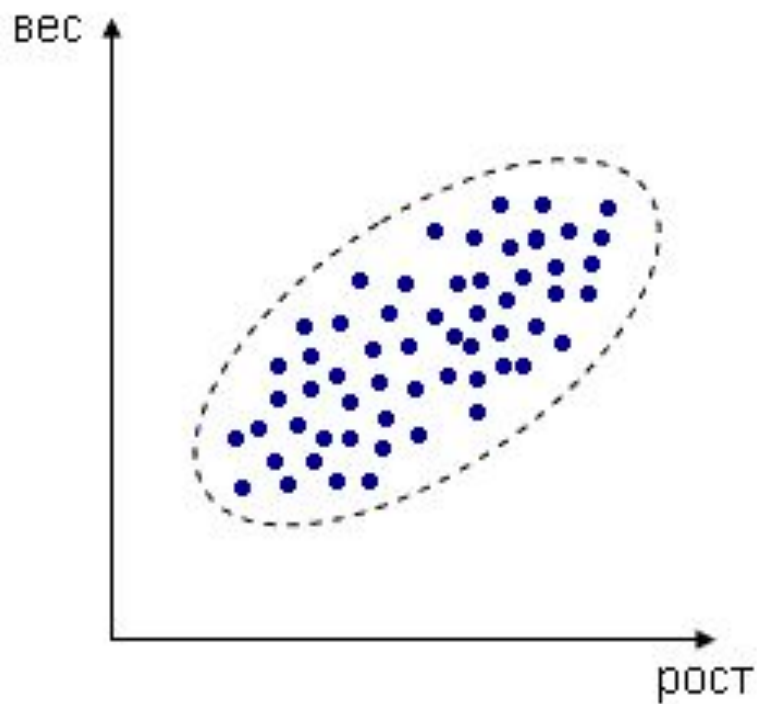
Формула коефіцієнта кореляції при лінійній залежності

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

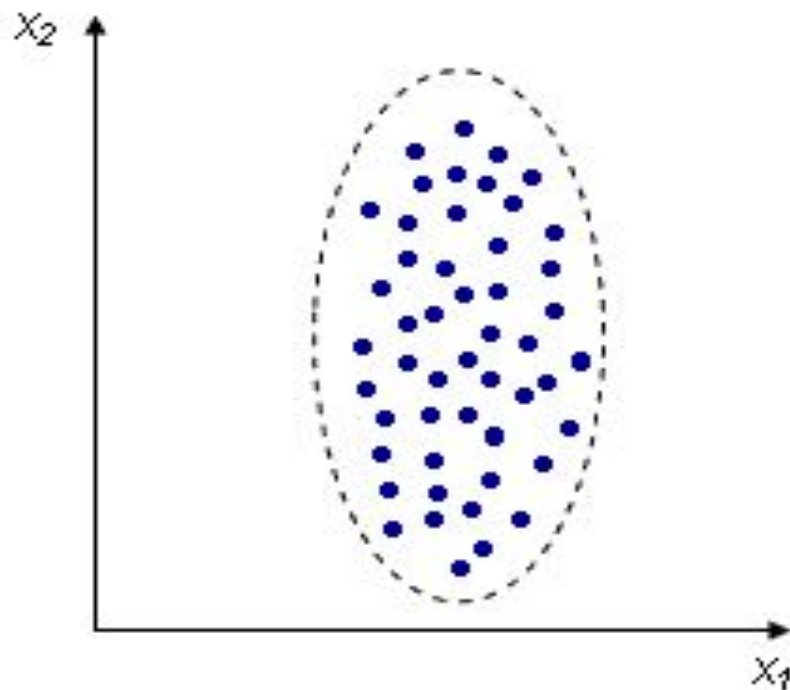
Величина коефіцієнта лінійної кореляції Пірсона не може перевищувати +1 і бути менше ніж -1. Ці два числа **+1 і -1** - є межами для коефіцієнта кореляції.

Коли при розрахунку виходить величина більша +1 або менша -1 - отже сталася помилка в обчисленнях.

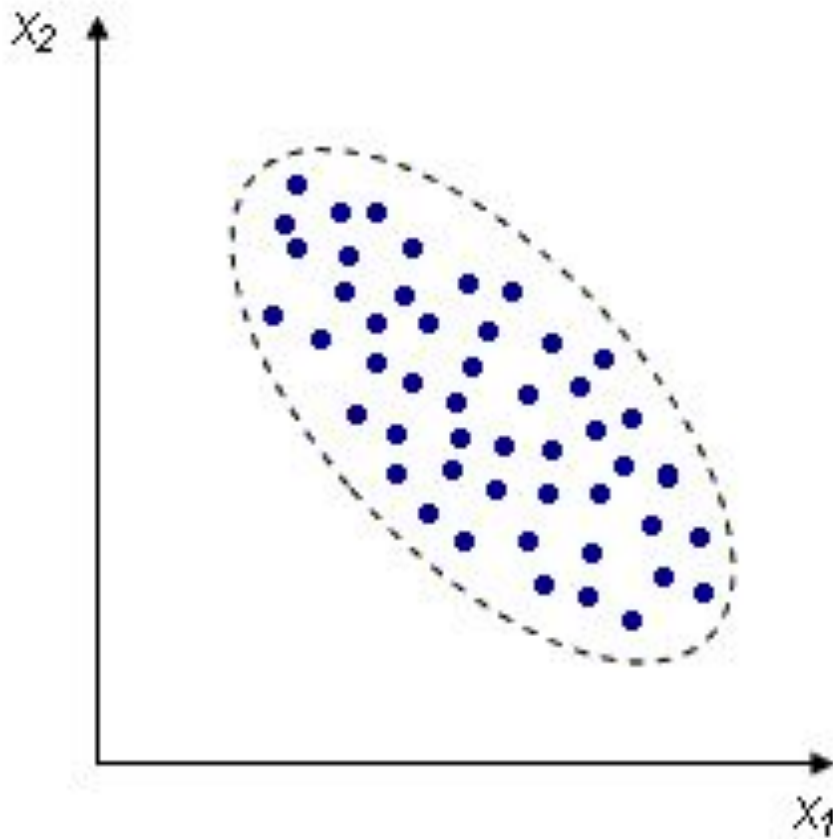
ПОЗИТИВНА КОРЕЛЯЦІЯ



відсутність кореляції



негативна кореляція



шкала Чеддока

| Теснота связи | Значение коэффициента корреляции при наличии: | |
|----------------|---|----------------|
| | Прямой связи | Обратной связи |
| Слабая | 0,1–0,3 | (-0,3)–(-0,1) |
| Умеренная | 0,3–0,5 | (-0,5)–(-0,3) |
| Заметная | 0,5–0,7 | (-0,7)–(-0,5) |
| Высокая | 0,7–0,9 | (-0,9)–(-0,7) |
| Весьма высокая | 0,9–1 | (-1)–(-0,9) |

регресійний аналіз

- **регресійний аналіз** - це кількісний метод визначення виду математичної функції в причинно-наслідкового залежності між змінними величинами.
- Метою регресійного аналізу є встановлення форми залежності.

види регресий

лінійна регресія: $y = a + b x$

регресії, нелінійні по пояснює змінним:

- 1) поліноми різних ступенів $y = a + b_1x + b_2x^2 + \dots$;
- 2) рівнобічна гіпербола $y = a + b / z$.

регресії, нелінійні по оцінюваним параметрам:

- 1) статечна $y = ax^b$;
- 2) показова $y = ab^x$;
- 3) експоненціальна $y = ea + b x$.

оцінки параметрів a і b
знаходяться за формулами:

$$a = \bar{y} - b \cdot \bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\text{cov}(x, y)}{\text{var}(x)},$$

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 - \text{дисперсия признака } x$$

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x} \cdot \bar{y}$$

a і **b**

формально a - значення y при $x = 0$.

Якщо ознака-фактор x НЕ може мати нульового значення, то вищевказана трактування вільного члена a не має сенсу, тобто параметр a може не мати економічного змісту.

параметр b називається коефіцієнтом регресії.

його величина показує середня зміна результату зі зміною фактора на одну одиницю.

коефіцієнт детермінації

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2},$$

$$\sigma_{\text{ост}}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2, \quad \sigma_y^2 = \frac{1}{n} \sum (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2$$

приклад:

| Номер региона | Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x | Среднедневная заработная плата, руб., y |
|---------------|--|---|
| 1 | 78 | 133 |
| 2 | 82 | 148 |
| 3 | 87 | 134 |
| 4 | 79 | 154 |
| 5 | 89 | 162 |
| 6 | 106 | 195 |
| 7 | 67 | 139 |
| 8 | 88 | 158 |
| 9 | 73 | 152 |
| 10 | 87 | 162 |
| 11 | 76 | 159 |
| 12 | 115 | 173 |

Рішення:

| | x | y | xy | x^2 | y^2 |
|-------------------------|-------------|--------------|----------------|---------------|----------------|
| 1 | 78 | 133 | 10374 | 6084 | 17689 |
| 2 | 82 | 148 | 12136 | 6724 | 21904 |
| 3 | 87 | 134 | 11658 | 7569 | 17956 |
| 4 | 79 | 154 | 12166 | 6241 | 23716 |
| 5 | 89 | 162 | 14418 | 7921 | 26244 |
| 6 | 106 | 195 | 20670 | 11236 | 38025 |
| 7 | 67 | 139 | 9313 | 4489 | 19321 |
| 8 | 88 | 158 | 13904 | 7744 | 24964 |
| 9 | 73 | 152 | 11096 | 5329 | 23104 |
| 10 | 87 | 162 | 14094 | 7569 | 26244 |
| 11 | 76 | 159 | 12084 | 5776 | 25281 |
| 12 | 115 | 173 | 19895 | 13225 | 29929 |
| ИТОГО | 1027 | 1869 | 161808 | 89907 | 294377 |
| <i>среднее значение</i> | <i>85,6</i> | <i>155,8</i> | <i>13484,0</i> | <i>7492,3</i> | <i>24531,4</i> |

Рішення:

$$r = \frac{\overline{yx} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{13484 - 155,8 \cdot 85,6}{7492,3 - 85,6^2} \approx 0,92$$

$$a = \bar{y} - b\bar{x} = 155,8 - 0,92 \cdot 85,6 \approx 77$$

кластерний аналіз

Кластерний аналіз являє собою клас методів, використовуваних для класифікації об'єктів або подій у відносно однорідні групи, які називають **кластерами (clusters)**.

кластерний аналіз

Об'єкти в кожному кластері повинні бути схожі між собою і відрізнятися від об'єктів в інших кластерах.

Кластерний аналіз також називають класифікаційними аналізом (classification analysis) або чисельної таксономії (систематикою) (numerical taxonomy).

Ідеальна ситуація кластеризації



Рис. 20.1. Идеальная ситуация кластеризации

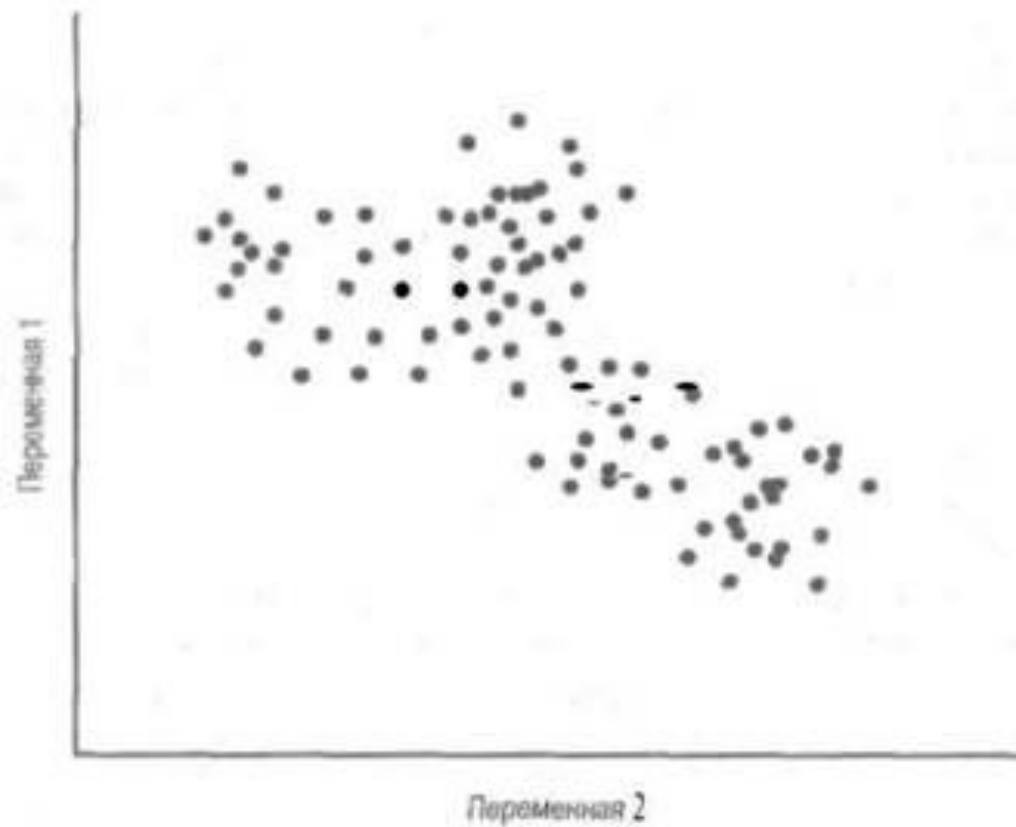


Рис. 20.2. Реальная ситуация кластеризации

дисперсійний аналіз

- *дисперсійний аналіз* - це статистичний метод, що дозволяє підтвердити або спростувати гіпотезу про те, що дві вибірки даних відносяться до однієї генеральної сукупності.

Наприклад, стосовно аналізу діяльності підприємства можна сказати, що дисперсійний аналіз дозволяє визначити, до однієї і тієї ж сукупності даних чи ні відносяться групи різних спостережень.

спасибі за увагу