

## АНАЛІЗ ДАНИХ АНКЕТУВАННЯ ЗА ДОПОМОГОЮ КЛАСТЕРНОГО АНАЛІЗУ

Яловенко Олег

Науковий керівник: канд.екон. наук, доцент Фетісов В.С.

*Ніжинський державний університет імені Миколи Гоголя, м. Ніжин, Україна*

*Анотація.* У статті присутнє описання різних методів аналізу. Використання цих методів для аналізу даних анкетування. Особливу увагу приділено методу кластерного аналізу. Описано завдання кластерного аналізу, основні проблеми які він вирішує. Виділено переваги кластерного аналізу над іншими методами аналізу даних. Наведено приклад його застосування, етапи кластеризації даним цим методом та описано результат. Зроблено висновки щодо доцільності використання даного непараметричного методу аналізу даних не тільки в соціологічних дослідженнях, а й розгляд його як складову частину системного аналізу.

**Ключові слова:** *аналіз даних, кластер, кластерний аналіз, анкетування.*

**Analysis of data using cluster**

**O. Yalovenko**

**Scientific supervisor: Candidate of Economics Sciences, Docent Fetisov V.S.**

*Nizhyn Gogol State University, Nizhyn , Ukraine*

*Abstract.* The article describes various methods of analysis. Use these methods to analyze survey data. Particular attention is paid to the method of cluster analysis. The tasks of cluster analysis are described, the main problems it solves. The advantages of cluster analysis over other methods of data analysis are highlighted. An example of its application, stages of clustering by this method are given and the result is described. Conclusions are made on the expediency of using this non-parametric method of data analysis not only in sociological research, but also its consideration as an integral part of systems analysis.

**Key words:** *data analysis, cluster, cluster analysis, questionnaire.*

**Постановка проблеми.** На сьогоднішній день більшість соціологічних опитувань дає лише відсоткове співвідношення голосів: рахується основна кількість людей, які відповіли позитивно, або відсоток незадоволених, але системно це питання не розглядають [1]. Найчастіше опитування не показує тенденції зміни ситуації. У деяких випадках є необхідність рахувати не кількість людей, які «за» або «проти», а відстань, або міру подібності, тобто визначати групи людей, які думають приблизно однаково.

Для виявлення на основі даних опитування деяких реально існуючих взаємозв'язків ознак і породження на цій основі їх типології, можна використовувати процедури кластерного аналізу [2]. Наявність, будь-яких

апріорних гіпотез соціолога при роботі процедур кластерного аналізу не є необхідною умовою.

**Аналіз досліджень і публікацій.** Кластерний аналіз – це група статистичних методів, з допомогою яких дані поділяються на схожі групи об'єктів, тобто на кластери – таким чином щоб кожен із кластерів складався зі схожих об'єктів, а от об'єкти різних класів – мали відмінності. Взагалі кластер це об'єднання кількох, чи декількох однорідних елементів, яке може бути розглянуте як самостійна одиниця зі своїми певними властивостями. Як кластери можна розглядати колег які пов'язані роботою, людей пов'язаних сімейними відносинами, один тип звірів, тощо.

Найважливішим етапом є вибірка ознак за якими об'єкти будуть поділятися, та сама процедура групування.

Галузь застосування кластерного аналізу є досить обширною:

- Маркетинг – сегментація конкурентів та споживачів.
- Менеджмент – розбиття персоналу на рівні за рівнем мотивації групи; кваліфікація постачальників.
- Медицина – класифікація симптомів, пацієнтів чи припаратів.
- Соціологія – розбиття респондентів на однорідні групи.
- Анкетування – обробка результатів, та виділення будь-яких статистичних даних.

Завдання кластерного аналізу полягає у разі необхідності розділення вихідної сукупності об'єктів на групи схожих, близьких між собою об'єктів, тобто на кластери. В даний час відомо близько сотні різних алгоритмів кластеризації. Їх різноманітність пояснюється не тільки різними обчислювальними методами, але і різними концепціями, що лежать в основі кластеризації. Найбільш часто використовуються ієрархічний кластерний аналіз і кластеризація методом k-середніх.

Кластерний аналіз в порівнянні з іншими типами аналізу має декілька важливих переваг:

- Нормальність даних не є обов'язковою вимогою;

- Не потребує великих обсягів даних, може працювати навіть з незначними вибірками;

- Ознаки даних можуть бути виражені будь-якою довільною шкалою.

Етапи кластерного аналізу:

- 1) Спочатку вибираються об'єкти які порівнюються (будь-який предмет, який може бути класифікований за будь-якими параметрами)

- 2) Потім виділяються ознаки за якими об'єкти будуть порівнюватись

- 3) Обчислюються риси подібності чи відмінності об'єктів за допомогою обраних вище ознак.

- 4) Об'єкти групуються в кластери за допомогою обраної вище процедури об'єднання.

- 5) Перевіряється можливість використання даного отриманого кластерного рішення.

**Мета статті:** Проведення дослідження, розкриття поняття, та виділення переваг кластерного аналізу.

**Виклад основного матеріалу дослідження.** Припустимо, проведено анкетування співробітників і потрібно визначити, яким чином можна найбільш ефективно керувати персоналом. Тобто необхідно розділити співробітників на групи і для кожної з них виділити найбільш ефективні важелі управління. При цьому відмінності між групами повинні бути очевидними, а всередині групи респонденти повинні бути максимально схожі.

Виходячи з цього, нам потрібно провести кластерний аналіз результатів соціологічного опитування працівників підприємства. Спочатку застосуємо розвідувальний аналіз даних. Анкета, яка використовувалася в опитуванні, має 63 показники якості трудового життя працівників, розділених на 7 блоків:

- 1) трудовий колектив (психологічний клімат, стреси на роботі, відносини з адміністрацією і ін.);

- 2) оплата праці (заробітна плата, поточне преміювання, відчуття економічного добробуту та ін.);

- 3) робоче місце (територіальна близькість до дому, офіс та меблі, рівень організації робочого місця та ін.);
- 4) керівництво організацією (задоволеність стилем керівництва, відносини з безпосереднім начальником, повага підлеглих і ін.);
- 5) службова кар'єра (планування службової кар'єри, бажання змінити роботу, заохочення навчання персоналу та ін.);
- 6) соціальна гарантія (оплата лікарняних листів, надання відпустки за графіком, медичне страхування та ін.);
- 7) соціальні блага (отримання форменого одягу, компенсація транспортних витрат, оплата витрат по дитячим установам і ін.).

Кожен показник оцінювався за десятибальною шкалою. При цьому 10 балів характеризує найвищу оцінку, а один бал - найнижчу. Загальна кількість опитуваних – 60 осіб.

Відповідаючи на поставлені питання, співробітники зіставляють з дійсністю свої уявлення про себе, свої цінності, здібності і можливості розвитку. Оціночне ставлення задоволеності працею відображає соціально-психологічну реакцію працівників на те, наскільки їх очікування від роботи відповідає реальному досвіду.

В результаті дослідження була побудована гістограма середніх значень (рис. 1).



Рис. 1. Гістограма задоволеності співробітників

Дана гістограма наочно відображає ставлення співробітників до якості своєї трудової діяльності. Основні показники, якими не задоволені працівники:

- соціальні блага (середній бал 2,115);
- оплата праці (середній бал 3,022);
- службова кар'єра (середній бал 3,991).

Для обробки результатів опитування співробітників за допомогою кластерного аналізу пропонується використовувати метод К-середніх [3], який полягає в тому, що визначаються еталонні кластери, а далі кожен об'єкт приєднується до найближчого еталону. В якості критерію використовується мінімальна відстань всередині кластера щодо середнього.

Провівши дослідження за всіма показниками якості трудового життя, найбільший інтерес викликали оплата праці, робоче місце, соціальні блага і службова кар'єра. Аналіз даних, що відображає вищевикладені показники, дозволив зробити наступні висновки:

1. З рис. 2 видно, що маємо дві групи людей, середні значення для параметрів заробітної плати яких, розрізняються. Крива 2, швидше за все, відображає керівництво.

Незважаючи на те, що працівники діляться на дві категорії, вони відносно однаково відповідають на питання. Іншими словами, якщо щось не влаштовує загальну групу працівників, те ж саме не влаштовує і вище керівництво, і навпаки. Узгодження графіків дозволяє зробити висновки про те, що добробут однієї групи позначається на добробуті іншої.

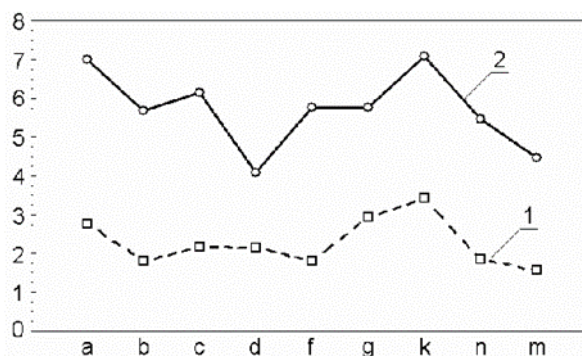


Рис. 2. Графік координат центрів кластерів для пунктів анкети, що відносяться до оплати праці

- a - вживана система оплати праці;
- b - заробітна плата;
- c - співвідношення кваліфікації та оплати праці;
- d - доплати за суміщення професій, ступінь, звання;
- f - винагорода за кінцевий результат;
- g - поточне преміювання;
- k – обґрунтованість доплат;
- n - відчуття справедливості в оплаті праці;
- m - відчуття економічного добробуту

Аналізуючи рис. 3, можна помітити, що одна з груп співробітників не задоволена територіальною близькістю до дому. Даною групою є основна частина працівників, які в основному приходять на підприємство з різних сторін міста. Отже, можна запропонувати головному керівництву направити частину прибутку на будівництво житла для співробітників підприємства.

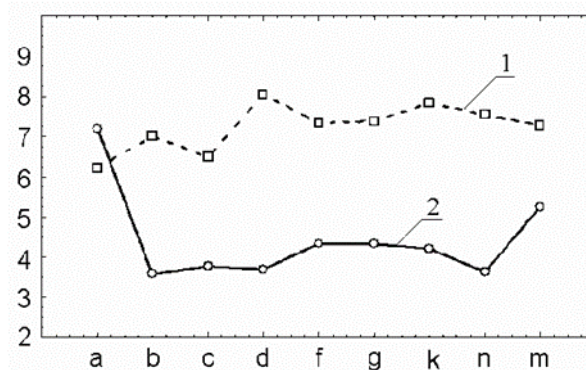


Рис. 3. Графік координат центрів кластерів для пунктів анкети, що відносяться до робочого місця

- a - територіальна близькість до будинку;
- b - офіс і меблі;
- c - оргтехніка;
- d - службовий транспорт, автостоянка;
- f - фізіологічні умови праці;
- g - рівень організації робочого місця;
- k - доведення планового завдання до цеху, робочого місця;
- n - рівень нормування праці;
- m - відчуття особистої безпеки.

2. На рис. 4 проглядаються суттєві відмінності щодо двох груп людей до службової кар'єри. Ті співробітники, яких влаштовують такі показники, як просування по заслугах і кваліфікації, об'єктивність атестації кадрів, заохочення навчання персоналу, планування службової кар'єри, у яких високий

збіг особистих цілей і цілей організації – не мають бажання змінити роботу і відчують задоволеність результатами праці. І навпаки, співробітників, які бажають змінити роботу і незадоволених результатами праці, не влаштовують викладені показники. Керівництво повинно звернути особливу увагу на ситуацію, що склалася.

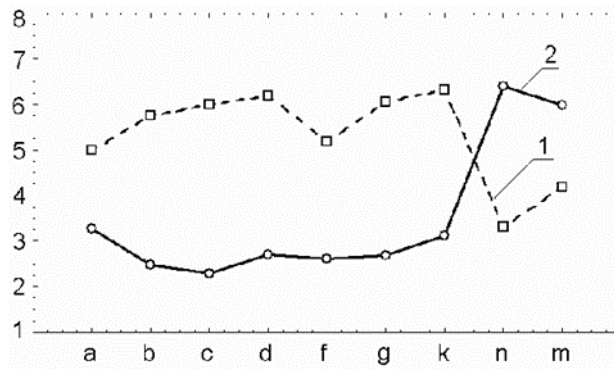


Рис. 4. Графік координат центрів кластерів для пунктів анкети, що відносяться до службової кар'єри

- a - виявлення лідерів і робота з ними;
- b - планування службової кар'єри;
- c - керівництво сприяє кар'єрі;
- d - поєднання особистих цілей і цілей організації;
- f - заохочення навчання персоналу;
- g - просування по заслугах і кваліфікації;
- k - об'єктивність атестації кадрів;
- n - бажання змінити роботу;
- m - відчуття незадоволеності результатами праці

3. На рис. 5 явно проглядаються дві групи людей, які відповідають приблизно однаково, але є істотні відмінності в другому показнику - формений одяг. Швидше за все, одній з них (керівництву) не надається формений одяг, тому що в цьому немає необхідності.

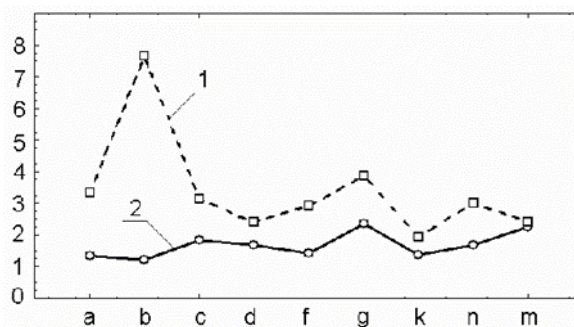


Рис. 5. Графік координат центрів кластерів для пунктів анкети, що відносяться до соціальних благ

- a - виплата матеріальної допомоги до відпустки;
- b - отримання форменого одягу і взуття;
- c - оплата спортивно-оздоровчих послуг;
- d - подарунки до ювілейних дат;
- f - пільгові кредити на житло і покупку автомобіля;
- g - компенсація транспортних витрат;
- k - компенсація витрат на харчування;
- n - оплата витрат по дитячим установам;
- m - відчуття соціального благополуччя

### **Висновки та перспективи подальших пошуків у напрямі дослідження.**

Можна помітити, що при обробці результатів анкетування кластерний аналіз виявився потужним методом, що дозволяє зробити висновки, до яких неможливо прийти, побудувавши гістограму середніх або порахувавши відсоткове співвідношення задоволених різними показниками якості трудового життя.

Кластерний аналіз дозволив виявити проблемні ситуації і указав на шляхи їх вирішення. Отже, цей метод непараметричної статистики можна використовувати не тільки в соціологічних дослідженнях, а й розглядати як складову частину системного аналізу.

#### **Список літератури**

1. Отрут В.А. Соціологічне дослідження: методологія, програма, методи. - Самара: Самарський університет, 1995. - 332 с.
2. Мендель І.Д. Кластерний аналіз. - М.: Финанси і статистика, 1981. - 133 с.
3. Вуколов Е.А. Основи статистичного аналізу. Практикум по статистичним методам і дослідженню операцій з використанням пакетів STATISTICA і EXCEL. - М.: ФОРУМ ИНФРА-М, 2004. - 464 с.