

Mathematical Support for Statistical Research Based on Informational Technologies

Roman Kaminskyi, Nataliia Kunanets and Antonii Rzheuskyi

Information Systems and Networks Department,
Lviv Polytechnic National University,
Stepan Bandera street, 32a, 79013, Lviv, Ukraine
Roman.M.Kaminskyi@lpnu.ua, nek.lviv@gmail.com,
antonii.v.rzheuskyi@lpnu.ua

Abstract. The authors developed "case of the information manager", which will include tools for mathematical and statistical analysis, oriented to those who have the skills of working with software only "at the user level" and implemented by means of MS Excel table processor. "Case of the information manager" includes the following methods: descriptive statistics, correlation analysis, classification and cluster analysis methods, paired comparison method, Saaty's method (algorithmic hierarchical procedure) and others.

Keywords: "case of the information manager", correlation analysis, ranking method paired comparison method, Saaty's algorithmic hierarchical procedure, hierarchical agglomerative cluster analysis.

1 Introduction

The use of information technologies in various areas of human activity has given rise to a certain interest in application of mathematical and statistical methods in humanitarian field, particularly, for library staff - information managers. Without good mathematical training, they often use simple methods to improve their work, using, in spite of this, personal computers and elementary mathematical support of MS Excel spreadsheet. In essence, this situation creates an actual scientific problem, the meaning of which in the development of a special manual on the application of mathematical methods for library workers. The aim of this work is the selection of simple and effective mathematical and statistical methods for mathematical study of decision-making information manager. The first among these methods are methods supported by information managers and are the following: correlation method, method of ranking, method of pair comparisons, Saaty's analytic hierarchical procedure and the hierarchical agglomeration cluster analysis. These methods are implemented by a simple step-by-step algorithm.

2 Correlation Analysis

This method has been used to determine the differences between reading rooms in relation to their attendance by readers. As data, the number of visitors per room per month is taken. Mutual and partial correlation coefficients were determined. The names of reading rooms and calculation results are given in Table 1.

Table 1. The results of correlation analysis.

| | | |
|--|--|--|
| Relationship between the reading room № 1 and the room of abstracts № 2 | Relationship between the reading room № 1 and the room of patents № 3 | Relationship between the room of abstracts № 2 and the room of patents № 3 |
| R₁₂ = 0.689 | R₁₃ = 0.823 | R₂₃ = 0.688 |
| The coefficient of partial correlation of rooms № 1 and № 2 relatively to the room № 3 | The coefficient of partial correlation of rooms № 1 and № 2 relatively to the room № 3 | The coefficient of partial correlation of rooms № 1 and № 2 relatively to the room № 3 |
| R_{12.3} = 0.297 | R_{13.2} = 0.664 | R_{23.1} = 0.293 |

Thus, it can be approved that the reading rooms № 1 and № 3 are more similar concerning visits.

3 Ranking Method

The simplest method to evaluate the impact n factors by determining their weight, i.e. setting their ranking is the method of ranking. The essence of this method is represented by the following example. It is necessary to determine in what order in the instruction to put 7 rules for use of reader fund. Participants in the ranking were the 5 specialists. To solve this, a group of specialists was invited. Each of them gives an assessment of the place of the rule on 10-point scale. The results of the method are shown in Table 2.

Table 2. Determination the order of the instruction items.

| | | | | | | | |
|---------|------|------|------|------|------|------|------|
| Rules | P1 | P2 | P3 | P4 | P5 | P6 | P7 |
| Points | 18 | 17 | 33 | 35 | 26 | 23 | 20 |
| Weights | 0.10 | 0.10 | 0.20 | 0.20 | 0.15 | 0.13 | 0.12 |

The rules in the instruction must be in this order: P4, P3, P5, P6, P7, P1, P2. Though $P1 > P2$ and $P4 > P3$, presented values of their weights are equal as a result of rounding. This method is the easiest to evaluate staff. The results of work of employees who are evaluated and ranked from the best to the worst or vice versa are compared. Ranking methods also make it possible to compare employees with each other according to selected criteria.

4 Paired Comparison Method

The paired comparison method is used when the group of objects must be submitted in a certain sequence, i.e., to determine the rank of each object for giving preference. For example, libraries offer 6 services, among which the most demanded must be identified. Three experts were invited. The weights of alternatives, identified by the experts, are shown in Table 3.

Table 3.

| V ₁ | V ₂ | V ₃ | V ₄ | V ₅ | V ₆ |
|----------------|----------------|----------------|----------------|----------------|----------------|
| 0.207 | 0.184 | 0.156 | 0.076 | 0.187 | 0.191 |

So, according to experts, the most demanded service is V₁. Next are the services: V₆, V₅, V₂, V₃, V₄.

5 Saaty's Algorithmic Hierarchical Procedure

This method involves decomposition of the problem into simple components and processing the judgments of decision maker. The essence of this procedure is quantitative expression of qualitative considerations. The structure of the problem is presented with a hierarchy, the top of which is the goal, and the levels are criteria and alternatives. The procedure includes pairwise comparisons, both criteria and alternatives. Quantitative values are estimates in the scale of relations. For example, for the post of librarian, five candidates were selected for 5 criteria: O, I, S, N, J., based on interview results for each of the criteria comprise the matrices of pairwise comparisons and defined the normalized vectors of local priorities, constructed in matrix B, and vector N for it is constructed. To determine global priorities, vector U is constructed by multiplying the matrices $B \times N$. The results are shown in Table 4.

Table 4.

| | B | | | | | N | U |
|---|-------|-------|-------|-------|-------|-------|--------------|
| O | 0,262 | 0,143 | 0,510 | 0,029 | 0,090 | 0,138 | 0,149 |
| I | 0,075 | 0,495 | 0,033 | 0,059 | 0,255 | 0,113 | 0,109 |
| S | 0,507 | 0,047 | 0,064 | 0,502 | 0,138 | 0,160 | 0,179 |
| N | 0,129 | 0,074 | 0,130 | 0,148 | 0,478 | 0,154 | 0,124 |
| J | 0,028 | 0,241 | 0,263 | 0,263 | 0,039 | 0,113 | 0,118 |

In the last column we obtain, as the matrix product, the values of the vector of global priorities for candidates for the position of librarian U. Candidate S has the highest global priority. Thus, with the help of this analytical hierarchical procedure, candidate S is recommended for the post of librarian. This method is widely used in decision making support.

6 Hierarchical Agglomerative Cluster Analysis

In the practice of library and information managers often there are problems of splitting a group of objects into separate classes. If objects are described by identical sets of attributes, then the problem can be solved by means of cluster analysis. Clustering is often the first step in data analysis. The essence of hierarchical agglomerative cluster analysis is that in the beginning all elements are considered as separate clusters, and then the distances between these objects are determined and two of them, the closest ones are grouped together into a separate cluster. Its procedure is iterative and ends with the union of all objects in one cluster, but with a clearly defined structure, which is presented as a dendrogram. As an example, in Fig. 1 the result of the cluster analysis of a group of 20 University libraries based on 12 indicators using the distance of Euclid in the form of a dendrogram is shown. Potential clusters are marked with horizontal dashed lines - similar to each other libraries.

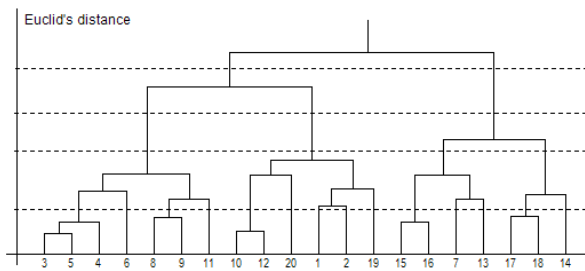


Fig. 1. Dendrogram of splitting into clusters the group of University libraries.

As a result of splitting into clusters this group of libraries, at the lowest level there are 14 clusters, at the next level is 4, then 3 and two. That is, depending on the height of the level, we have a different number of clusters.

7 Conclusions

The listed methods can be supplemented by many others, but even this set can significantly improve the efficiency of the work of the information manager. In this study, the main principle is the presentation of the method, starting with a detailed consideration of solution of specific task with it and discussion of the result. Therefore, at each stage of the used method it is necessary to give detailed interpretation of the obtained intermediate result. Practical methods verification and discussion with the direct users, confirmed the validity of this approach and effectiveness of their use in the practice of information managers.