Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 278 (2026) 1839–1844

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2025

A multivariate analysis technique to support guidelines for team size in digital projects

Luisa Mich*

Department of Industrial Engineering, University of Trento, Via Sommarive 9, 38123, Trento, Italy

Abstract

The design of increasingly complex digital systems requires teamwork. The size and composition of teams are critical for initial activities focused on understanding the problems to be addressed through digital technology. Regardless of the adopted methodology, from the classic waterfall to the agile approach, the first step is to identify the needs and objectives – requirements in software engineering terms – of the various stakeholders. Requirements elicitation is key to the success of digital projects and requires a multidisciplinary approach. A relevant question then is that of the ideal size of teams working in requirements elicitation sessions. The existing literature provides guidelines, but they are usually not supported by empirical data. A recent survey has partially filled this gap by providing data on the issue of team size. Thanks to the availability of that data, this paper proposes applying the multiple correspondence analysis (MCA), to further exploit them. MCA allows multiple factors to be considered simultaneously, and above all it does not require ex-ante assumptions to be made about the associations between them. MCA, which is not well known in software engineering, makes it possible to identify and graphically visualize relationships between many variables. The ultimate goal is to eventually use the identified profiles to support the guidelines for team staffing. Due to the size of the dataset (92 responses), the results of the MCA are exploratory. Nevertheless, they confirm mainstream guidelines.

Keywords: Group size; Multiple correspondence analysis; Profiles; Requirements Engineering

* Corresponding author. Tel.: +39-0462-282150
E-mail address: luisa.mich@unitn.it

1877-0509

1. Introduction

Teamwork is an important concept in project management [1]. Digital systems methodologies suggest group activities for their design, and so does requirements engineering for the elicitation step (see e.g. [2], [3]). At the same time, companies are under pressure to control their staff costs. In addition, groups dynamics – number of possible lines of communication and polarization of positions among the others [4], [5] – can hinder the work of the team of analysts responsible for requirements elicitation. A digital projects manager looking for advice on how to form teams will often find general guidelines that are not supported by empirical data [6]. The problem is that it is very difficult to gather data on group work in the development of digital systems. In this context, requirements elicitation is a critical multidisciplinary activity, involving many different areas. Consequently, many different factors must be considered when analyzing the available data. To overcome the limitations of descriptive statistics, this paper proposes applying a multivariate technique [7], the multiple correspondence analysis (MCA) [8]. The goal is to identify relationships between these factors without assuming an a priori model of their associations, as is required for the most well-known structural equation modelling (SEM) [9]. In this sense, MCA is an exploratory technique that identifies latent 'profiles' whose characteristics could in turn be used to support decisions on team formation. Although MCA is widely used in the fields of sociology and economics, it is virtually unknown in the computer science and software engineering communities. Section 2 provides a brief overview of the MCA, offering an initial understanding of its key concepts and citing seminal references for further exploration. As an illustration of its potential, Section 3 reports on two examples of its use, exploiting data collected with a survey on how companies carry out requirements elicitation in relation to group or individual activity [10]. The results of the MCA are exploratory due to the size of the dataset (92 responses). However, on one hand they do confirm the guidelines suggested by the literature for the requirements group. On the other hand, the validity of the MCA would require additional efforts to gather data from companies — a significant challenge.

2. The Multiple Correspondence Analysis

The MCA is a statistical technique for the study of associations between categories of two or more qualitative variables (characters) characterizing aspects (individuals) under investigation [11], [12]. It can be seen as a generalization of Karl Pearson's Principal Component Analysis (PCA) [13]. MCA has been (re)discovered many times and equivalent methods are known under various names, including optimal scaling, optimal or appropriate scoring, dual scaling, homogeneity analysis, scalogram analysis, and quantification method [14]. MCA allows to highlight patterns of relationships between several dependent variables. In a traditional approach, to identify such relationships, the modalities and the frequencies of the characters are usually reorganized into contingency tables that often have a complex structure and tend to distort – if not hide – essential relationships. To bring to light the structure of the relationships (also called associations) among the characters, in MCA the contingency tables are substituted with structures of lower dimensions capable to maintain (or preserve) the relevant part of information available in the rough data. The substitution process aims to remove redundant information (noise) so that the essential associations among characters are highlighted more clearly, all the while attempting to minimize the loss of information. MCA provides a way of visualizing multiple data on a plane. The graphical representation of MCA results in biplots, which make it easier to identify and associations that cannot be inferred from the raw data. In other words, with MCA it is possible to visualize the results in terms of points in a smaller subspace than the original multidimensional space, facilitating their interpretation. Modalities with strong associations are represented as points that are close to each other in the subspace produced by the MCA, while loosely associated or unassociated modalities have points that are further apart: the distance between individual points reflects the dissimilarities between response patterns of individuals. Individuals who choose the same categories are represented by points close together. To interpret the results in an MCA biplot it is important to know that the closeness of the points to a line, plane, or in general to a low-dimensional subspace is defined as the sum of squared distances from the points to the subspace. So, to correctly interpret the distances it is necessary to check the data produced in the process to reduce the original tables: points projected into a plan (bi-dimensional space) could seem close even if they are not.

3. MCA for team size in requirements elicitation

3.1. The data

Data on how companies carry out requirements elicitation (ReqElic) in relation to group or individual activity was collected using an online questionnaire (https://zenodo.org/record/7616991). Results of the survey highlighted that in many companies, requirements are elicited in small groups, and that the ideal size suggested is the same or the closest to that usually adopted in the company. Classical characteristics of the projects – large-size, complexity, and tight-deadline – have also been confirmed as relevant to adopt groupwork for requirements elicitation. A comprehensive descriptive statistical analysis is given in [10]. The explanation given here is abbreviated to what is necessary to understand the potential of the application of the MCA.

The questionnaire included 12 questions The first multiple level-multiple choice question (Q1. My role in ReqElic) aimed to know the role of the respondents in requirements elicitation. Respondents had to choose between 'all our projects', 'most of our projects', 'in some of our projects', or 'none of our projects' for four possible roles: 'business or requirements analyst', 'software engineer', 'project manager', 'representative of the client or of the customer'. In this way, it was possible to consider that people working in requirements elicitation usually plays different roles. The same answers (modalities) were available for the rest of the close-ended questions, from Q2 to Q5. The second multiple level-multiple choice question (Q2. Group vs individual activity in ReqElic) asked if requirements are identified as an 'individual by a single BoRA', 'individual by more than one BoRA' or as 'group activity' (BoRA stands for Business or Requirements Analyst, including all the 4 roles). Questions from Q3 to Q5 investigate the impact of three different factors characterizing projects on the adoption of teamwork, namely, large-size, tight-deadline, and complexity. Then a multiple level-multiple choice was added to investigate the kind of ReqElic technique adopted in group session (Q6), in particular to know if JAD, brainstorming and other creativity techniques were used. The following two questions, Q7 and Q8 were core for the study, as they asked the usual and ideal size of groups, respectively, and answering them involved choosing between some numbers (2, 3, 4 or >5 for usual size and three combinations of 4 BoRAs, namely 4 individuals, 2 groups of 2, and 1 group of 4 for ideal size). Question Q9 asked the respondent to explain the answer to the previous question about the ideal size for the group, and it was an open question. Questions from Q10 to Q12 were semi-open, asking the respondent to give three factors to recommend using groups of BoRAs (Q10), and not using groups of BoRAs (Q11), and three keywords to characterize the domain or sector of their company's software system projects (Q12). Answers to the last question allowed to identify three types of projects: information systems, software systems, and websites.

The descriptive statistics for the results of the questionnaire highlighted that only 15% of the companies do never adopt groups for ReqElic, but also that in about 2/3 of them ReqElic is an individual activity in at least some of the projects (Q2). ReqElic is more often used as a group activity in large (Q3) and complex (Q5) projects than in tight-deadline projects (Q4). The most commonly used group size is 2, with the other sizes (3, 4, 5) used in decreasing frequency (Q7). Question 8 (Q8) on the ideal size suggested three combinations of 4 BoRAs, namely 4 individuals, 2 groups of 2, and 1 group of 4, which were chosen by 21%, 37% and 42% of the respondents respectively.

3.2. The results of MCA

As a first step, we applied MCA to different combinations of the questions included in the above-mentioned survey (using standard libraries in R). As an initial step, we applied MCA to different combinations of questions from the aforementioned survey. In this paper, we report on the two combinations that most closely align with the guidelines suggested in the literature: namely, that small team size is better, and that project characteristics are the most relevant factors. The guideline regarding small teams could be formulated as follows: Are the roles involved in ReqElic (Q1), considering if ReqElic is a group or individual activity (Q2), satisfied with their company's adopted team size (usual size, Q7), or would they prefer a different size (ideal size, Q8)? The results of the application of MCA are given in Fig. 1. An initial interpretation of the biplot can start from the horizontal axis (in the MCA, axes are not defined in advance but depend on the projection of the points that maximize the information in the data). It can be interpreted as representing the usual size of groups in ReqElic: in fact, moving from left to right, the size of usual groups increases.

Vertical axis divides respondents playing more often (in all or most projects) the roles involved in ReqElic, from those playing those roles less frequently (in some or none).

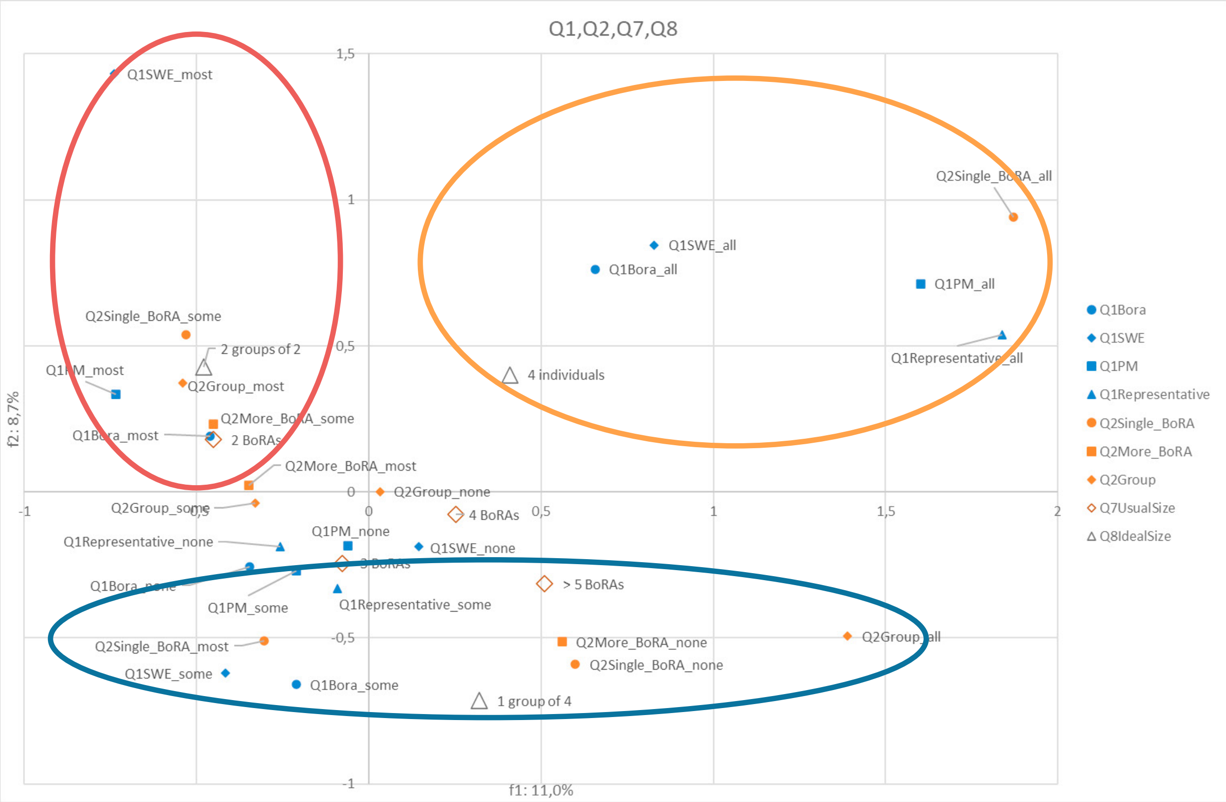

Fig. 1. MCA biplot: Role (Q1) ReqElic as group vs. individual activity (Q2), usual size of groups (Q7), ideal size of groups (Q8)

As regards the ideal size, the graph reflects the associations identified with the descriptive statistics. In particular, the point representing the usual size of 2 BoRAs is close to the 2 groups of 2 as ideal size (64% of those choosing 2 BoRAs for the usual size indicated 2 groups of 2 as ideal size). The MCA map suggests three different profiles of respondents, indicated with an ellipse to facilitate their individuation. In the first quadrant, the ellipse corresponds to respondents (a) playing a same role in ‘all ‘the projects (Q1), and (b) working in companies in which requirements are identified as an individual activity by a ‘single BoRA’ working alone in all the projects (Q2): these respondents suggest 4 individuals as ideal size (Q8). The ellipse on the left side indicates a profile including respondents that (a) have one of the roles involved in ReqElic in ‘most’ of the projects (Q1), and (b) working in companies in which requirements are identified as a group activity in ‘most’ of the projects, but also as individual activity by a ‘single BoRA working alone’ or by ‘more than one BoRA’ each working separately in ‘some’ of the projects (Q2): they suggested as ideal group size 2 groups of 2 (Q8). The third bottom ellipse includes respondents that (a) have one of the roles involved in ReqElic in ‘some’ of the projects (Q1), and (b) working in group in ‘all’ the projects (and in ‘none’ of the other modalities for Q2), whose usual group size is large (Q7), and, coherently, suggest ‘1 group of 4’ as ideal size for ReqElic groups (Q8). Altogether, Fig. 1 suggests that the respondents tend to confirm as an ideal size for ReqElic groups (Q8) the usual size of groups (Q7). This trend is also related with ReqElic as group vs. individual activity (Q2) and with the frequency of more specific roles (Q1: all, most, some). In other words, compared to those involved in 'more' or 'some' projects, those involved in 'all' ReqElic projects ('all' projects) prefer to work alone or in smaller groups. That means people with a specific role, i.e. more specialized in any of them, do not believe in the advantages related to larger groups and prefer to minimize risks connected with teamwork (as highlighted by answers to Q9). On the other hand, people used to work in group and playing different roles in the company digital projects, suggest larger groups, confirming the importance of teamwork.

The second MCA investigated how the ideal group size (Q8) and the usual size (Q7) depends on the characteristics of the projects – Q3 Group activity in ReqElic as a function of the size of the project, Q4 Group activity in ReqElic as a function of the tight deadline of the project and Q5 Group activity in ReqElic as a function of the complexity of the project – and on the type of the project (Information system, Software system or Web-based system, Q12). MCA allowed to identify two profiles. The ellipse on the right in Fig. 2, corresponds to respondents who adopt groups in ReqElic for 'all' the 'large-size', 'tight-deadline', and 'complex' projects. Usual sizes of the groups are '4 BoRas' and 'more than 5 BoRAs' (Q7). Ideal size for this profile is '1 group of 4', but also '4 individuals'. Interesting, on one hand respondents say that ReqElic is a group activity in all the 'all' the 'large-size', 'tight-deadline', and 'complex' projects, but on the other hand they suggest as an ideal size 4 individual working separately. This confirms the duality inherent in teamwork, which can optimize some factors but at the expense of others. The ellipse on the left indicates that there are companies that do not identify requirements as a group activity for any of the large, tight-deadline, and complex projects. They usually work in groups of two for small, simpler projects. Answers to Q3, Q4, Q5 show a high association degree: the more ReqElic is a group activity for one of the characteristics of the projects, the more it is for any of the other. Altogether, MCA in Fig. 2 highlight an association between group activity in ReqElic for the different kind of projects (large-size, tight-deadlines, complex) and the usual size of the group, supporting guidelines which suggest adapting the team size to the characteristics of the project. This result corresponds to what we might expect, but with MCA we have empirical confirmation of which dimensions are adopted by companies. Finally, as regards the type of system – Information system, Software system, and Website – coded as yes or no for Q12, they are very close to the center of the MCA, so that they are not associated with any of the profiles; in other words, the type of system is less relevant than the characteristics of the project for team size guidelines.

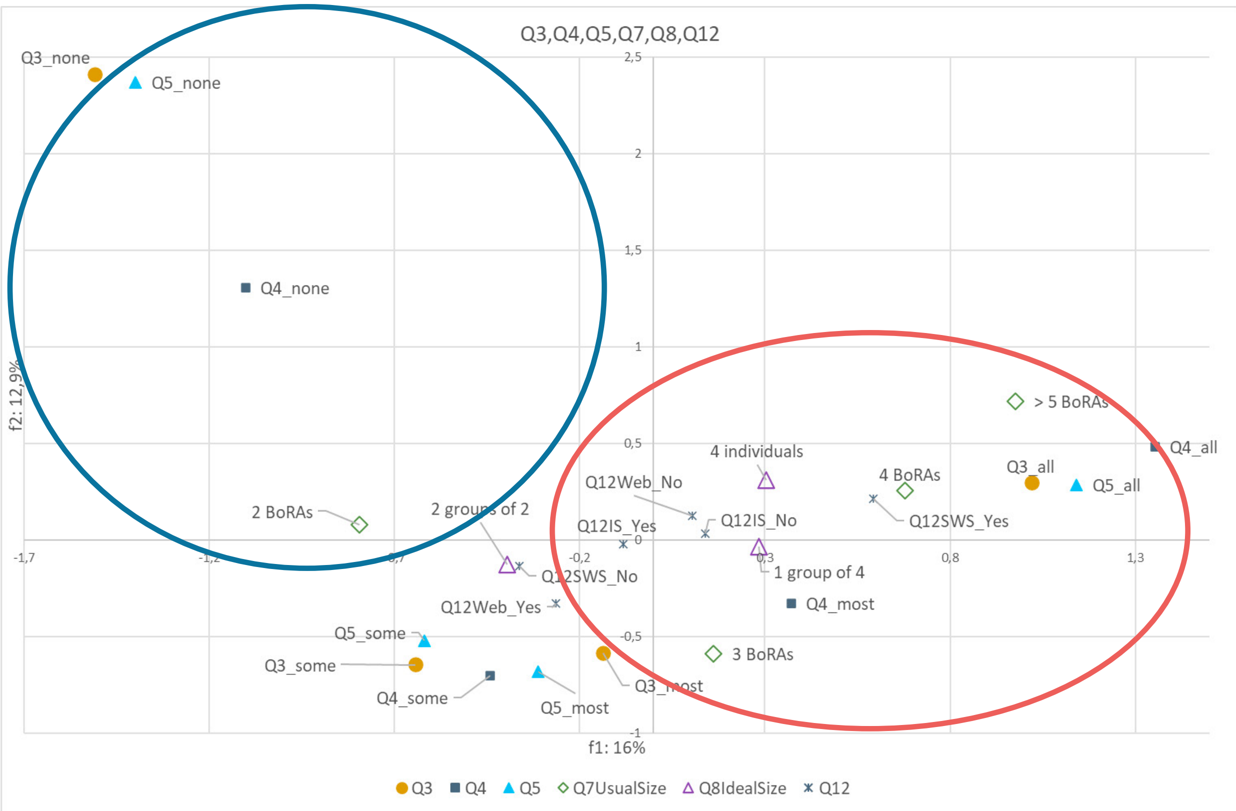

Fig. 2. MCA biplot: large-size (Q3), tight-deadline (Q4), complexity (Q5), usual size of groups (Q7), ideal size of groups (Q8), type of project (Q12)

4. Conclusion

Guidelines for team size in digital projects are given by many authors (a review is given in [10]). However, they are rarely based on empirical data. Even focusing on requirements elicitation activities, which are necessary for the analysis of the problem in all kinds of projects (not only digital ones), suggestions for forming groups sound like rules of thumb.

In this paper, we proposed applying a multivariate technique (MCA) to explore the relationships between the factors investigated in a ReqElic survey focused on team size. Among the many possible combinations of the answers to the questionnaires, we focused on the two that are connected to the most frequently given guidelines, i.e., to adopt small groups and that the size of the group should be adapted to the characteristics of the projects. The aim was not to provide an exhaustive description of the MCA technique, but rather to demonstrate its potential as a statistical tool for analyzing the relationships between multiple variables involved in requirements elicitation, without assuming an a priori model. The MCA enables multidimensional data to be visualized in two-dimensional diagrams; however, interpreting these diagrams is not straightforward. MCA biplots visualize some interesting associations between qualitative variables characterizing teamwork in requirements elicitation and confirm guidelines suggested by the literature. These results are limited by the small number of respondents to the questionnaire, and to (b) the long period of time during which they were collected [10]. Due to the size of the dataset, the reported results are highly exploratory and limit the strength of inferences that can be made, as can be seen from the variable contributions in the MCA biplots. The long time it took to reach the 92 filled-in forms is indicative of the difficulties researchers face in obtaining empirical data on digital system design and, in particular on requirements elicitation. Adding more statistical techniques such as MCA to the toolbox is then relevant to exploit the available data.

Acknowledgements

Thanks to Pietro Marzani and Pierluigi Novi Inverardi for the MCA, and to the co-authors of the previous papers on the size of groups in requirements elicitation.

References

[1] DeMarco, Tom, and Lister, Tim (2013) *Peopleware: Productive Projects and Teams*. 3rd ed., Addison-Wesley Professional.

[2] Pohl., Klaus. (2025) *Requirements Engineering: Fundamentals, Principles and Techniques*, 2nd ed., Springer.

[3] Pressman, Roger S., and Maxim, Bruce. (2020) *Software Engineering: A Practitioner's Approach*, 9th ed., McGraw Hill Education.

[4] Lewin, Kurt, and Lewin, Gertrude W. (Ed.) (1948) *Resolving social conflicts: selected papers on group dynamics (1935–1946)*. New York, Harper and Brothers.

[5] Isenberg, Daniel J. (1986) “Group polarization: A critical review and meta-analysis.” *J. Pers. Soc. Psychol*. **50** (**6**): 1141–1151.

[6] Mich, Luisa, Sakhnini, Victoria, and Berry, Daniel M. (2023) “To group or not to group? Group sizes for requirements elicitation.” *Inf. Softw. Technol.* **160**, (Aug 2023). https://doi.org/10.1016/j.infsof.2023.107229

[7] Warner, Rebecca M. (2012) *Applied Statistics: From Bivariate Through Multivariate Techniques: From Bivariate Through Multivariate Techniques.* Sage Publications, Thousand Oaks, CA, USA

[8] Benzécri, Jean-Paul. (1992) *Correspondence Analysis Handbook*. Dekker, New York.

[9] Hair, Joseph F., Hult, G. Thomas. M., Ringle, Christian M., Sarstedt, Marko, Danks, Nicholas P., and Ray, Soumya. (2021) “An Introduction to Structural Equation Modeling.” In: *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. Springer, Cham. https://doi.org/10.1007/978-3-030-80519-7_1

[10] Mich, Luisa, Sakhnini, Victoria, and Berry, Daniel M. (2021) *Raw data from 92 responses from questionnaire*. https://doi.org/10.5281/zenodo.7617278.

[11] Clausen, Sten-Eric. (1998) *Applied Correspondence Analysis. An introduction*. SAGE Publication.

[12] Greenacre, Michael. (2007) *Correspondence Analysis in Practice*. 2nd ed. Chapman & Hall/CRC, London, UK.

[13] Pearson, Karl. (1901). “On lines and planes of closest fit to systems of points in space.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **2** (**11**): 559-572. https://doi.org/10.1080/14786440109462720

[14] Salkind, Neil. (Ed.) (2007) *Encyclopedia of Measurement and Statistics*. Sage Publications, Thousand Oaks, CA, USA