Using Friedman Test for Creating Comparable Group Results of Nonparametric Innovation Competence Data

Pasi Porkka Jari Jussila Anu Suominen
Industrial Engineering and Management, Tampere University of Technology, Pori unit
(E-mail: pasi.porkka@tut.fi, jari.j.jussila@tut.fi, anu.suominen@tut.fi)

Abstract The paper employs competence self-evaluation approach in order to develop human resources. Usually the data in human sciences is nonnumeric and therefore nonparametric. This reduces the number of valid statistical methods that are available for analysis. Thus the paper presents a statistically valid Friedman test for achieving group results from nonparametric data. It is implied that the traditional statistical methods used in analyzing nonparametric data lead to false conclusions. Therefore, the paper sets up an appropriate method for analyzing innovativeness of human resources. The paper draws a conclusion that self-evaluation combined with a valid statistical method offers a useful instrument, but a human must have the major role in analyzing the results and for deciding the education and development methods needed.
Key words competence data, innovation, self-evaluation, statistical methods, friedman test

1 Introduction

When human beings and their behavior are researched, some of the studied aspects are nonnumeric. Actually there are less numerically measureable aspects than there are nonnumeric, e.g. linguistic ones. The problem with nonnumeric variables is that traditional statistical methods are not usable to them. For example, two linguistic answers like “quite strong” and “strong” can not be summed up. Also comparison of results of two different people is difficult and achieving reliable group results is also of a problem.

Usually the inputs to nonnumeric variables are turned into numbers, since numbers are easier to store with computers. This transformation brings along the temptation to use mathematical methods to the received numbers. However, traditional statistical methods are mathematically and statistically hardly ever valid to such data.

The non-validity claim arises from the data itself. When humans evaluate something, the inherent method is linguistic. However, linguistic variables are nonparametric by nature and therefore not comparable.

In this paper a valid test for this kind of nonparametric data is shown. The Friedman test (Friedman 1937, 1939) gives also some advantages over the not-so-valid traditional tests.

2 Specific Features of Nonnumeric and Nonparametric Data

Typical methods for gathering data in human sciences are questionnaires, evaluations and interviews. The focus in human sciences is usually in some aspect of human life, which can be described in words but not measured with parameters (see Friedman 1937, p. 675). Answers to linguistic questions can hardly ever be parametrical. Here, the parametrical means that there is an implicit functional relationship between different answers. If answers are parametrical, then there is relation between them, and they are for example comparable. If two different answers are nonparametric, then there is no valid way to compare them. One cannot distinguish, which is greater than the other; they are mainly someone’s opinions.

The focus of human studies can be related to feelings, needs, wishes or relations between humans. All these aspects are nonparametric. Therefore all methods used to study these aspects, produce data, which is also nonnumeric or nonparametric. Only valid statistical methods are nonparametric. Nonparametric statistics are the branch of statistics that studies data measurable on an ordinal or nominal scale, to which arithmetic operations cannot be applied (Conover 1999).

It is important to comprehend how some answers are achieved. The method of achieving may define the set of valid methods also for analysis. Sometimes the focus of the study is measurable and the data can be achieved by measurements. For example the height of a man is measurable and different measures are comparable. If there are two men, whose heights are measured to be 185cm and 183cm, we can say that the first one is taller. However, if two different people evaluate the heights of these same men, without measuring them, their answers would not be comparable. They could easily evaluate the

heights incorrectly.

Another important aspect with linguistic variables is that all people have their own degree of scale. For example, what is the height after which people are considered to be tall? If the respondent is 200cm tall, his answer would most likely differ from an answer of 150cm tall respondent. People have their own inherent way of scaling various things. The scale differs from man to man, but in a questionnaire form there is only one scale.

This is the reason why two different (but seemingly equal) answers, to the question about some evaluated, but not measured feature, are not comparable, nor available to arithmetic operations. The respondents may have made errors on their evaluation and/or the respondent's degrees of scale are different.

The problem here arises from the data itself. From the statistics point of view the data is of nominal scale. The nominal scale means that numbers (the values) are used merely as a means of separating the answers into different classes or categories (Conover, 1999). The different answers of a single respondent are comparable, but answers of different respondents are not. The ability to compare answers of one respondent allows us to conclude who is strongest in respondent's family (Table 1.). However, when analyzing group results, the traditional statistic functions (sums, means, etc.) are not applicable.

Usually in human studies, there is more interest in the group results, than answers of a single respondent. Researchers are eager to get some explaining power from group results to their hypotheses. However, the use of the traditional statistics may lead astray and lead to non-valid results. The next small example describes this problem: Let there be five families. One child of each family evaluates the strength of his/her parents. The answers are shown in Table 1.

Table 1 Fictitious Answers to the Strongest Parent in a Family Survey

Respondent	Strength of Father	Strength of Mother
A	0,2	0,1
B	0,3	0,3
C	0,2	0,1
D	0,3	0,2
E	0,3	**0,9**
SUMS	1,3	**1,6**

If we sum the answers, to get group results, it looks like mothers are stronger than the fathers. However, there is only one family (E's family), where mother is stronger. The respondent E's personal degree of scale is quite different than with the others. In three families the father is considered to be stronger. The most important question that arises is: "Are mothers really stronger in the group of respondents?"

Therefore, we suggest that the answers should be evaluated with some valid method. There are several suitable statistical methods for nonnumeric data. In our study case there are several related variables in the ontologies. The most powerful test for several related samples, where the number of different treatments is more than six, is the Friedman test (Conover, 1999).

3 Innovation Competence Data

Competence can be defined as underlying characteristic of a person, that may be a motive, trait, skill, aspects of one's self-image or social role, or a body of knowledge which he or she uses (Spencer & Spencer 1993, p. 21). Innovation competence refers here to those underlying characteristics of a person that enable generation of innovations in an organization (c.f. Amabile, 1997). As a result of literature survey a total of 27 competences were identified as important for innovation (Suominen et al. 2008). The scope of this paper is in the individual competences and how they are evaluated in terms of creative tension within two different organizations.

In our research, we have gathered data by self-evaluation. We have used a previously developed self-evaluation questionnaire program Evolute (Kantola 2005). The program offers the means for self-evaluation over the internet. The respondents answer several statements (linguistic variables) and the program saves answers to these statements in a numeric way. Each statement has a relation to some variable of our interest. Typically there are 60-120 statements and 15-30 variables in these research projects. If these variables are studying humans, they are called competences. If the focus of the research is on organizations, then the variables are called system variables. The set of statements and corresponding competences or system variables form the conceptual model, i.e. ontology to be studied. In future in this paper the word competence may refer to any kind of a variable.

Competences are the explanatory variables, which explain the target of the study. They are non-related, which means that they exist independently of each other. The value of any variable is not depending of the value of any other variable. This specification gives us the possibility to compare the values of a single respondent's variables, and rank his/her answers.

To study an ontology and its competences with self-evaluation, there are 2-5 statements specified to each competence. The statements are the questions. As an answer, each statement in the program provides an area, marked with labels, where the respondent gives the answer by clicking the mouse on the point of the area of his or her choosing, shown as lines. To each statement, there are four labels that provide the domain for the answer. The answer can be given to anywhere in the area, not just to the spot besides the label. The labels can be, for example (in Figure 1.), "strongly agree", "agree", "disagree" and "strongly disagree".

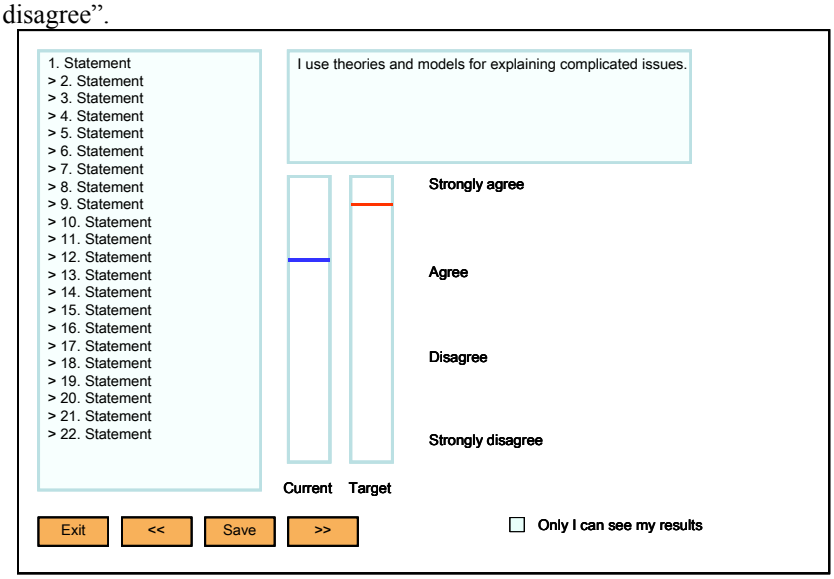

Figure 10 Answering to a Statement; View from a Self-evaluation Webpage

For each statement the respondent should estimate both the current and the target values for that statement (Figure 1). It is also possible to change the value. The current is the value of the statement at the moment. The target value represents the respondent's wish of the statement's value in the future. The difference of target and current can be considered as the creative tension of the statement (Senge, 1990). The creative tension is the most important motivation factor that pushes forward people in their personal development. We use the creative tension as the proof of motivation. If there is big creative tension for some competence, then people are motivated to improve that competence. Therefore it is also wise, from employer's point of view, to arrange education or development measures to improve those competences.

Since the statements are linguistic variables, the values given to them are personally dependent. Although the respondent draws a line as an answer to a statement, the program transforms it to a numerical value between 0 and 1. These numerical values are then summed together with fuzzy logic to gain one single value for corresponding competence. The values of different competences are comparable only within one person. Also values of same competence of two different people are not comparable. For example, two different people are asked how strong their fathers are. There is no way of an outsider, not personally knowing the fathers, to be able to say which father is actually stronger. Even if we give a numeric value for their answers, they are still not comparable, since we do not know the respondents personal scale of degrees. It is also mathematically not valid to sum answers of different respondents together. How can you sum "quite strong" and "extremely strong" together; what is their sum result?

4 The Friedman Test

In our data the competences, or any other variables used, are called treatments. One person's answers to these variables are gathered in a block. The data can be arranged into a table where columns are treatments and on a row there are always one person's answers. There are two assumptions the data must fulfill. Firstly, the variables must be mutually independent, that is that the results within one block cannot influence results in another block. Secondly, the observations may be ranked according to some

criteria (Conover, 1999). In the previous fictitious example it is possible to rank the strongest parent in each family.

The Friedman test continues by ranking all treatments within a block. If there are ties, then ranking is divided within all same values. If there were three persons to be evaluated and all had different strengths, then the rankings would be 1, 2 and 3. If there were two equally strong and one weaker, then the rankings would be 1.5, 1.5, and 3. The fictitious example (from Table 1.) continues and in table 2 the rankings have been calculated. Note respondent B, where there was a tie and both have same ranking.

Table 2 Fictitious Answers are Substituted with Rankings within a Block.

Respondent	Strength of Father	Strength of Mother	SUMS
A	2	1	3
B	1,5	1,5	3
C	2	1	3
D	2	1	3
E	1	2	3
SUMS	8,5	6,5	15

Now in the group results, it can be seen that the fathers seem to be stronger than mothers, which actually was the case in the group. In general case each treatment (or competence) will get the sum of all rankings it had in all blocks. These sums are now comparable. It can be determined which competence was top-rated.

The Friedman test gives also some extra benefits compared to the sums. Firstly with Friedman test one can calculate, what is the minimum difference those sums must have, after which they are regarded unequal. With this value the competences can be grouped into clusters. The clustering is great help in further analysis. This difference may also be calculated with different approximate sizes.

The transformation of the values into rankings also removes the twisting affect, that the results are not comparable. Now, since the sum of all answers of each respondent is always the same, the group results are comparable. As it can be seen (in the Table 2.), the sum of answers of one respondent is always 3. Then those different respondent's results can be compared directly; additionally also different group results, with different amount of respondents, since the sum of means will always be 3.

5 Using Friedman test for Innovation Data

The empiric evidence contains twelve people from a telecommunication company (Company A) and ten people from a finnish university (Company B). The two test groups ran a self-evaluation test of the ontology of innovative competence (Jussila et al 2008). The creative tensions of those two groups were compared. Firstly, a traditional way, with statistically false method with averages of the given data are shown (Table 3.)

Table 3 Averages of Creative Tensions of two Different Companies .

	COMPANY A		COMPANY B
	n=12		n=10
0,26	Self-control	0,31	Absorptive capacity
0,25	Absorptive capacity	0,28	Professional and technical expertise
0,23	Change orientation	0,27	Self-confidence
0,22	Analytical thinking	0,27	Intuitive thinking
0,21	Conceptual thinking	0,26	Understanding others
0,20	Understanding others	0,24	Analytical thinking
0,17	Self-development	0,22	Communication
0,16	Flexibility	0,21	Accurate self-assessment
0,16	Stress tolerance	0,20	Conceptual thinking
0,15	Professional and technical expertise	0,20	Self-control
0,15	Trustworthiness	0,19	Self-development
0,13	Intuitive thinking	0,19	Flexibility
0,13	Conflict management	0,18	Stress tolerance
0,12	Seeking information	0,15	Relationship building
0,11	Relationship building	0,14	Initiative

0,11	Achievement orientation
0,11	Accurate self-assessment
0,10	Independence
0,10	Teamwork and cooperation
0,09	Self-confidence
0,08	Divergent thinking
0,08	Responsibility
0,08	Initiative
0,07	Imagination
0,06	Leveraging diversity
0,04	Communication
0,02	Risk orientation
3,56	SUM

0,13	Leveraging diversity
0,12	Imagination
0,11	Seeking information
0,11	Change orientation
0,09	Conflict management
0,09	Teamwork and cooperation
0,08	Trustworthiness
0,07	Achievement orientation
0,06	Responsibility
0,05	Divergent thinking
0,04	Independence
0,00	Risk orientation
4,25	SUM

As it clearly can be seen, the total sums are different, so comparison of the values of different companies is out of the question. It is also difficult to say which competencies belong to the best group and which to the worst group. How great should the difference be in order to be considered significant?

In Friedman test, the answers of one respondent are ranked. Then all rankings of one competence are summed to gain group results. If the test is carried out with the same data as is used in previous example (Table 3.), two tables are resulted (Table 4). Into the sub-tables the minimum difference from the actual given values have been calculated. Because the rankings are different, also the minimum difference values are different. Both sub-tables are divided into three clusters according to this minimum difference value. The upper third of the table consists of the competences, where the respondents have the highest creative tension. From the statistics point of view all competences in the cluster are equally important. However, in the analysis one can naturally consider the higher values to be more important. The lowest third consists of those competences the respondents have given smallest values. Also the sums are now equal in both tables, and therefore directly comparable.

As it can be seen from the table the two columns representing the rankings of creative tensions of two different organizations (Company A and Company B) have been clustered differently. This is due to the fact that the division to the clusters is dependable of the data at hand. The upper part of the column represents those competencies that have ranked high with their creative tensions. Similarly the lower part of the column represents those competencies that have ranked low with their creative tension. Furthermore, the middle part has ranked with mediocre rankings.

Table 4 Rankings of Creative Tensions of two Different Companies .

COMPANY A		COMPANY B	
	n=12, α=0.05, min.diff. 5,71		n=10, α=0.05, min.diff 5,79
21,83	Self-control	21,70	Absorptive capacity
20,83	Absorptive capacity	21,20	Intuitive thinking
20,25	Change orientation	21,10	Professional and technical expertise
19,17	Conceptual thinking	19,50	Self-confidence
18,83	Analytical thinking	19,30	Understanding others
18,08	Understanding others	18,30	Communication
17,50	Stress tolerance	18,10	Analytical thinking
16,63	Flexibility	17,80	Accurate self-assessment
15,67	Trustworthiness	16,80	Flexibility
15,67	Self-development	16,70	Self-development
14,92	Professional and technical expertise	16,40	Self-control
14,83	Conflict management	16,05	Stress tolerance
14,63	Intuitive thinking	16,00	Conceptual thinking
13,00	Relationship building	14,40	Relationship building
12,29	Seeking information	13,90	Initiative
12,13	Teamwork and cooperation	11,80	Leveraging diversity
11,79	Achievement orientation	11,20	Conflict management
11,63	Accurate self-assessment	11,15	Seeking information
11,46	Self-confidence	10,85	Change orientation
11,29	Divergent thinking	10,70	Imagination
11,17	Initiative	10,15	Responsibility
10,67	Imagination	9,00	Teamwork and cooperation
10,00	Independence	8,30	Achievement orientation
9,88	Responsibility	8,15	Trustworthiness
8,83	Leveraging diversity	7,90	Divergent thinking
7,79	Risk orientation	7,05	Independence
7,25	Communication	4,50	Risk orientation
378,00	SUM	378,00	SUM

6 Results

The table of these two organization's competence data analyzed with Friedman test, give interesting results to be studied further. Out of the total number of 27 competencies, in both organizations there are 7 corresponding competencies that are clustered with the high ranked creative tensions (Absorptive capacity, Analytical thinking, Change orientation, Conceptual thinking, Flexibility, Self-control, Stress tolerance and Understanding others). Similarly, there are 6 corresponding competencies that are clustered with the low ranked creative tensions (Achievement orientation, Independence, Divergent thinking, Responsibility, Risk orientation and Teamwork and cooperation). The further analysis of the meaning of the similarities in the clusters remains to be done in the future. However, the first great significance of the Friedman test is that it allows this further analysis to be carried out in the first place. Otherwise the analysis would be sheer speculation and conjecture. The second significance is the fact that this Friedman test can be carried out to a rather small group of respondents; however, naturally the group results more reliable the greater the respondent group is.

The result of the study suggests that Friedman test is a appropriate method for producing comparable empirical test results of nonparametric data regarding individual competencies for group comparison purposes.

7 Conclusions

The Friedman test is a useful method of analysis for nonparametric data, especially in human competence research. The method allows producing test results of groups, comparable with each other, regardless of the size of the group, even with small respondent numbers.

Our claims in this article have been as follows:

First, nonparametric data, especially group result of creative tensions of individual competence produced with self-evaluation, should not be accumulated and analyzed with traditional statistical methods.

Second, with Friedman test nonparametric data, especially self-evaluated group data of creative tensions of those individual competences can be analyzed with statistically valid methods.

Third, two test groups producing competence data via self-evaluation, even with different number of respondents, can be analyzed with this statistically valid method of Friedman test; even with small groups with small number of respondents.

Therefore, finally, we conclude that group results of linguistic data produced with self-evaluation, although raising questions of invalidity by its nature can be analyzed with statistically valid method of Friedman test in order to compare those group results. However, the further analysis of the gathered results, the meaning of the clustered rankings, still have to carried out by a human.

References

[1] Amabile, T.M. (1997) Motivating Creativity in organizations: on doing what you love and loving what you do. California Management Review, Vol. 40, No. 1. pp. 39-58.

[2] Conover, W. J., 1999. Practical nonparametric statistics. New York, John Wiley & Sons.

[3] Friedman, M. 1937. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. Journal of the American Statistical Association, Vol. 32, No. 200. pp. 675-701.

[4] Friedman, M. 1939. A Correction: The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. Journal of the American Statistical Association, Vol. 34, No. 205. pp. 109-109.

[5] Kantola, J. 2005. Ingenious Management. Tampere University of Technology. Publication 568.

[6] Jussila, J., Suominen, A., Vanharanta, H. 2008. Competence to Innovate? In: Karwowski, W. & Salvendy, G. (eds.) 2008 AHFE International Conference, 14-17 July 2008, Caesars Palace, Las Vegas, Nevada, USA, Conference Proceedings.

[7] Suominen A., Jussila, J., Porkka, P. Vanharanta, H. 2008. Interrelations of development needs between innovation competence and innovation culture? Proceedings of The 6th International [8] Conference on Manufacturing Research (ICMR08) Brunel University, UK, 9-11th September 2008.

Senge, P. 1990. The Fifth Discipline, New York, Currency Doubleday.

[9] Spencer, L.M., Spencer, S.M. 1993. Competence at work: models for superior performance. New York, John Wiley.