

5-1-2009

Which Is the Best Parametric Statistical Method For Analyzing Delphi Data?

Hiral A. Shah

St. Cloud State University, hashah@stcloudstate.edu

Sema A. Kalaian

Eastern Michigan University, skalaian@emich.edu

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Shah, Hiral A. and Kalaian, Sema A. (2009) "Which Is the Best Parametric Statistical Method For Analyzing Delphi Data?," *Journal of Modern Applied Statistical Methods*: Vol. 8 : Iss. 1 , Article 20.

DOI: [10.22237/jmasm/1241137140](https://doi.org/10.22237/jmasm/1241137140)

Which Is the Best Parametric Statistical Method For Analyzing Delphi Data?

Hiral A. Shah
St. Cloud State University

Sema. A. Kalaian
Eastern Michigan University

This study compares the three parametric statistical methods: coefficient of variation, Pearson correlation coefficient, and F-test to obtain reliability in a Delphi study that involved more than 100 participants. The results of this study indicated that coefficient of variation was the best procedure to obtain reliability in such a study.

Key words: Survey Research, Reliability, Delphi Technique, Decision-making.

Introduction

The Delphi Technique is a method for systematic solicitation and collation of judgments on a particular topic through a set of carefully designed sequential questionnaires interspersed with summarized information and feedback of opinions derived from earlier responses (Delbecq, Van de Ven, & Gustafson, 1975). The Delphi technique can be considered as an important tool to bring the knowledge and intuition of a group of qualified individuals to bear upon the future possibilities in a given field. Therefore, the technique can be used at a micro-level to arrive at a qualitative forecast which may vary from past trends in an organization.

The Delphi process consists of a series of rounds of questionnaires. The first round is characterized by exploration of the subject under discussion, wherein each individual contributes with information that he/she believes is pertinent. The second round involves the process of reaching an understanding of how the group views the issue (i.e., what group members mean by relative terms such as importance, desirability

or feasibility). The Delphi rounds of questionnaires should continue until a predetermined level of consensus is reached or no new information is gained (Ludwig, 1997; Linstone & Turoff, 1975; Delbec, Van de Ven & Gustafson, 1975). In most instances it is found that three iterations are sufficient, and not enough new information is gained to warrant the cost of more iterations (Ludwig, 1997).

Parametric statistical methods such as the coefficient of variation (CV) and the F-test have been used in Delphi studies with sample size below 50. The CV is a statistical measure of the deviation of a variable from its mean. The F-test is performed to determine the ratio of squares of two variances or, in other words, to test if the standard deviations of two populations are equal.

English and Kernan (1976, cited in Yang 2003) used the coefficient of variation (CV) to determine the stopping rule. If the magnitude of CV for an item was found to be too large (e.g., greater than 0.8), the corresponding statement was needed to be modified and required an additional round(s) of questionnaire administration.

Yang (2003) suggested using the F-test to compare two variances. The F-value is determined by the ratio of the variances of item scores among panelists between the two successive rounds. If no significant difference is found in the F-test, the questionnaire item will be dropped from further rounds. Questionnaire items where significant between-round difference in variances is found are retained in a

Hiral A. Shah is an Assistant Professor in Engineering Management. Email: hashah@stcloudstate.edu. Sema. A. Kalaian is a Professor in Technology Studies. Email: skalaian@emich.edu.

subsequent round. Yang (2003) described this method as being suggested by Jolson and Rossow (1971) with the problem being that assumptions made for the F-test may be seriously violated when using data collected from the Delphi rounds.

This study compared the three parametric statistical methods: coefficient of variation (CV), Pearson correlation coefficient, and F-test to obtain consensus and reliability in a Delphi study using data from Shah (2004) and Shah and Kalaian (2006) to find out the best method that fits the study that involves a large number of participants.

The data used in this study was obtained from research conducted by Shah (2004). The purpose of this study was to gather data using Delphi technique to discover and describe what experts in the field consider important to know in the discipline of engineering management, and use that information to update the curriculum for Eastern Michigan University's Engineering Management masters program. The Delphi panel in the study consisted of 194 panelists. These panelists were asked to rate the competency areas on a 5-point Likert-type scale and provide qualitative comments through mailed questionnaires. The following criterion of importance was assigned to the responses provided on the questionnaire, along with an example of how to respond: 5 = of very high importance, 4 = of high importance, 3 = of medium importance, 2 = of low importance, 1 = of very low importance.

This study consisted of three rounds of questionnaires. The sample comprised of individuals who belonged to any of these four categories: (a) Professor/instructor of Engineering Management, (b) Industry Professional, (c) Author of published text/papers/articles related to the breadth of Engineering Management discipline, and (d) Certified Engineering Manager/Certified Enterprise Integrator. Moreover, the competency areas were also grouped into four categories, namely: (a) Technical, (b) R&D/Design, (c) System/Organization/Project Management, and (d) Human Issues.

Methodology

The Round 1 Delphi questionnaire was sent to 707 subjects. Based on the information obtained from Round 1, Delphi panel members were selected and the Round 2 Delphi questionnaire was developed. In the second round, an analysis of the group's modal response and percentage concurrence for each degree of importance from the first round was provided to the Delphi panel for reference. Specific comments to a particular competency provided by the Round 1 subjects were reported in the Round 2 Delphi questionnaire. A space for comments was provided after every competency area for the respondents to respond to the comments made by other panel members from Round 1 or to give their own comment. Additional competency areas suggested by the Round 1 respondents were added to the existing list of competencies.

Panel members were asked to consider respondent comments and the percents of concurrence obtained from Round 1, rate each competency area on a five-point Likert-type scale, and explain their choice if it was two or more categories away from the Round 1 respondent's modal rating. An example of how and where to record their responses and comments was also provided. Additional comments made by Round 1 respondents were reported in Round 2 questionnaire for their reference. Space for additional or general comments was provided at the end of questionnaire.

The Round 3 Delphi questionnaire was developed using Round 2 results and was administered in the same manner as Round 2. Based on the category in which the Delphi panel members categorized themselves, a six-digit (rCodexxx) alphanumeric code was assigned to each of them. The first digit - r - represented the Delphi round (2 or 3) to which they responded; the code represented the category to which they belonged to in the form of letters A-for authors, C-for Certified Engineering Manager/Certified Enterprise Integrator, I-Industry professionals, and P-Professors teaching Engineering Management; and xxx represented the panel member's assigned number. Round 3 was also sent to the individuals who participated in Round 1 but did not participate in Round 2.

BEST PARAMETRIC METHOD FOR ANALYZING DELPHI DATA

Because the codes could not be assigned to the panelists who did not participate in Round 2, an additional sheet was sent to these panelists asking them to participate in the final Round 3 and also to checkmark the category to which they belonged.

The Round 3 Delphi questionnaire that was sent to the panel members who participated in Round 2 had individual codes. Moreover, additional questions were asked on the front page of the questionnaire asking the Delphi panel members to: Rate the overall importance of the results of this to the discipline of Engineering Management as a guide for others for curriculum development, rate the overall quality of this study, rate their own level of expertise in the field of Engineering Management, and additional space was provided to comment on the importance/quality of this study and suggestions for possible improvements. Table 1 shows the participation in the study and the response rates at the end of each round of Delphi study.

Table 1: Response Rates from Three Rounds of Delphi Study

| Delphi Round | Number Sent | Number Received | Response Rate (%) |
|--------------|-------------|-----------------|-------------------|
| 1 | 707 | 194 | 27.4% |
| 2 | 194 | 148 | 76.3% |
| 3 | 194 | 136 | 70.1% |

Data Analysis using Parametric Statistical Methods

The data was entered for each of the rounds using SPSS software. Due to missing values for one or more competency areas in several cases, those cases were excluded from the study. Thus, the sample size for this study was 52. The mean and standard deviation corresponding to each of the competency areas in Rounds 1, 2 and 3 were calculated using SPSS and Microsoft Excel software. Because coding for each panel member was applied from Round 2, the data obtained from Rounds 2 and 3 of the Delphi study could be corresponded case-wise. Hence, for this study, Rounds 2 and 3 will be considered for analysis purposes.

Coefficient of Variation

The Coefficient of Variation (CV), which is the ratio of standard deviation (σ) of a competency area to its corresponding mean (μ) among the panelists, was calculated using the formula:

$$CV = \sigma / \mu. \quad (1)$$

The CV was obtained for Rounds 2 and 3, and in order to determine if additional rounds were required, the absolute difference was calculated by subtracting the CV obtained from Round 3 from that obtained from Round 2. A small CV value was an indication that the data scatter or variation compared to the mean was small. A large CV value compared to the mean was an indication that the amount of variation was large.

As shown in Table 2, the absolute value of the difference in CV between Rounds 2 and 3 was less than 0.2, which can be considered to be a minor difference according to Dajani (1979, cited in Yang, 2003). Though negative values of difference was obtained for competency areas such as: Information systems, Linear programming, Materials engineering, Metrology-Measurement Science, Six sigma black belt certification and others, the absolute difference was still less than 0.2. Hence, it can be assumed that stability was reached for each of the competency areas and no further rounds of Delphi were required.

F-test to compare Two Variances

The F value for each competency area was obtained by calculating the ratio of the variances (σ^2) of item scores among panelists between Rounds 2 and 3. Hence,

$$F\text{-Ratio} = \frac{\sigma^2 \text{Round } 3}{\sigma^2 \text{Round } 2} \quad (2)$$

It is important to note that the degrees of freedom have not been taken into consideration in the F-test as they are already a part of variances. When no significant difference in the F-test is obtained, the questionnaire item will be dropped from further rounds.

The F-ratio of 1 implies that the variance of Round 3 is equal to the variance of

Table 2: Results of the Three Parametric Procedures from Round 2 to Round 3

| Statistic | Absolute difference in CV = CV(R2) – CV(R3) | F-ratio = Var(R3/R2) | Pearson’s r |
|----------------------------------|--|-------------------------|-------------|
| Mean | 0.025 | 0.789 | 0.397 |
| Median | 0.025 | 0.746 | 0.416 |
| Minimum Value | 0.070 | 0.000 | -0.240 |
| Maximum Value | 0.130 | 2.070 | 0.730 |
| % Reliability Obtained | 100% | 79% | 83% |
| Skewness Value Using Z scores | 0.080 | 0.093 | -0.429 |

Note: R3=Round 3, R2=Round 2

Round 2. Hence, a F-ratio less than or equal to 1 is desirable. The results from the F-test suggested that 79% of the competency areas had F-ratios less than or equal to 1 (see Table 2), indicating that stability was established in Round 3.

Pearson’s Product-Moment Correlation

Correlation is a technique used to determine the relationship between two quantitative, continuous variables. A correlation is often called a bivariate correlation to designate a simple correlation between two variables, as opposed to relationships between more than two variables (George & Mallery, 2005). A correlation, also known as Pearson’s Product-Moment Coefficient of Correlation, or the Pearson r, is one such measure of the strength of the association between two variables. George and Mallery (2005) stated, “although the Pearson r is predicted on the assumption that the two variables involved are approximately normally distributed, the formula often performs well even when assumptions of normality are violated or when one of the variables is discrete” (p. 124). A correlation value of +1.00 indicates a perfect, positive correlation, whereas, a correlation of zero

indicates no relationship between the two variables. A negative correlation indicates a relation in which one variable tends to increase as the other variable tends to decrease. The closer a correlation coefficient is to zero, the weaker the relationship between the two variables.

The correlation value, r, was obtained for each competency area using SPSS software. If the correlation coefficient for a particular competency area varied significantly from zero and was very high, it indicated that the ratings of panel members on the competency area were stable and less fluctuating.

The Pearson’s correlation coefficient obtained, indicated that there was a negative relationship for competency areas: Management of technology, Communications, Customer issues, and People and teamwork. Values of these coefficients were closer to zero, indicating a weaker tendency of increase in value of one competency with the decrease of value in the subsequent round. Thus, the panel members who responded lower in Round 2 for these competency areas, tended to respond higher in Round 3. The relationship was found to be weak and hence it was an indication that stability was obtained in Round 3. The results from Pearson’s correlation indicated that 83% of the

BEST PARAMETRIC METHOD FOR ANALYZING DELPHI DATA

competency areas had correlation values, which were either greater than or equal to zero (Table 2). Thus, it can be implied that there was a good correlation between the competency areas in Round 2 and Round 3.

Results

As the results of all the three parametric procedures used to obtain reliability in the Delphi study indicated similar results, it was important to determine the best procedure among the three. Hence, further analysis was performed on the results of the three parametric procedures: CV, F-ratio, and Pearson's r . Because the values of the three procedures were on a different scale, transformation of the values to similar scales for all the three procedures was completed using z scores (a measure of the distance in standard deviations of a sample from the mean). The z transformation is calculated as $(X - \mu)/\sigma$; where X is the observation, μ is the mean and σ is the standard deviation of the observations. A positive z score indicates that an

observation is greater than the mean whereas a negative z score indicates that an observation is below the mean.

A box plot comparing the z scores of the three parametric procedures for the 76 competency areas contained three outliers: case numbers 32, 38 and 69. As the outliers tended to skew the normal distribution, these cases were deleted and a box plot was derived. Figure 1 shows the box plots comparing the three parametric procedures without outliers and Figure 2 shows the histogram obtained from the data.

Because skewness is a measure of symmetry of the distribution, a positive value shows the distribution is positively skewed and a negative value shows that the data is negatively skewed. A comparative look at the values of skewness for all the three parametric procedures as shown in Figure 2 and Table 2 was the procedure to determine the best parametric procedure. Coefficient of variation had a smaller positive value of skewness (0.080) compared to Pearson r (-0.429), and F-ratio (0.093).

Figure 1: Box Plot Comparing the Z-Scores of the Three Parametric Procedures: Coefficient of Variation (CV), F-test, and Pearson's r

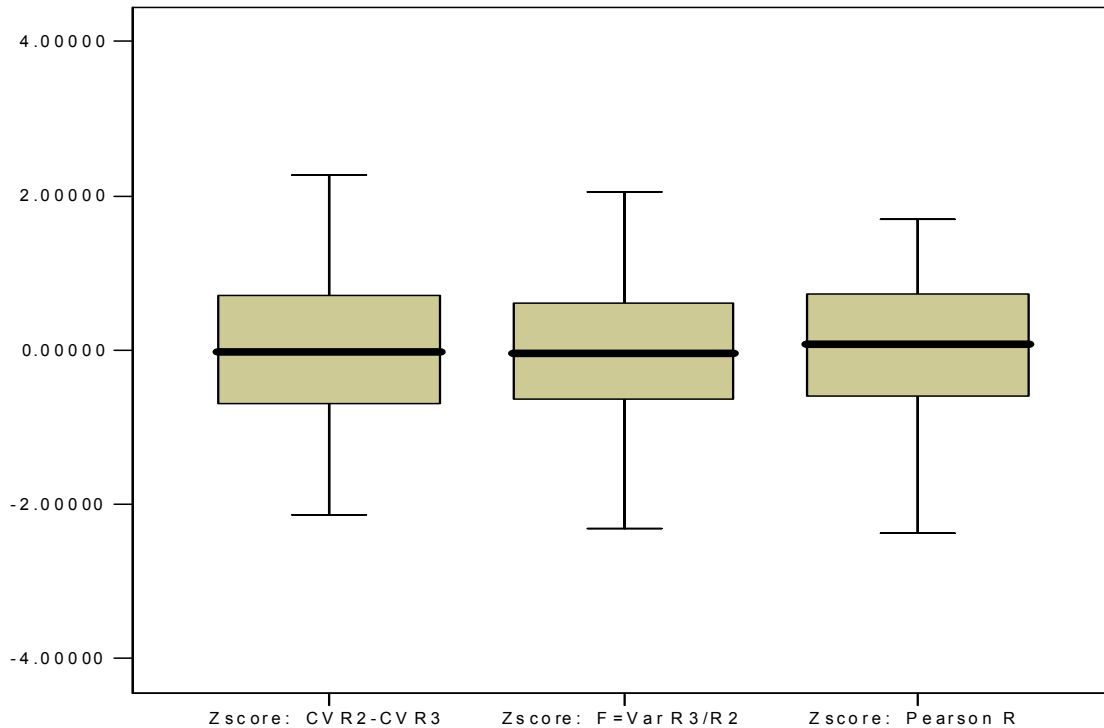
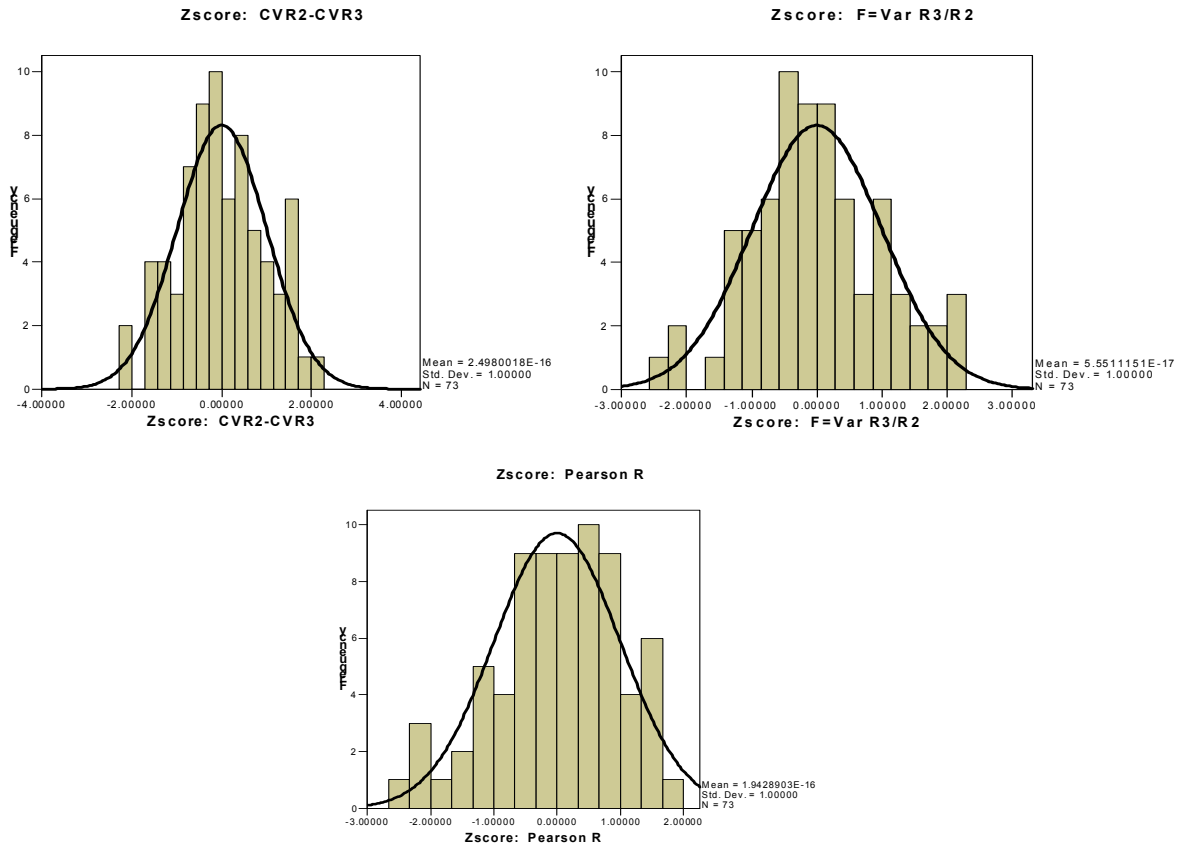


Figure 2: Histogram Obtained From the Z-Scores of the Three Parametric Procedures: Coefficient of Variation (CV), F-test, and Pearson's r



Conclusion

In summary, the results of the three parametric procedures indicated: 100% of the competency areas in Round 3 obtained stability and hence reliability was achieved by the coefficient of variation method; 79% of the competency areas had F-ratios less than or equal to 1, which indicated that stability has been established in Round 3; and 83% of the competency areas had Pearson r correlation values either greater than or equal to zero, depicting a good correlation between Round 2 and Round 3. As all the three parametric procedures were a good indication of obtaining reliability in a Delphi study a z scores were calculated and box plot was graphed.

The values of the skewness obtained from the descriptive values of the box plots, it was found that the coefficient of variation (CV) had a smaller positive value of skewness (0.080)

compared to Pearson's r (-0.429), and F-ratio (0.093). From these values, it could be concluded that the coefficient of variation was the best procedure to obtain reliability in a Delphi study that included more than 100 participants. The second best procedure to obtain reliability in a Delphi study is F-ratio and the third one is Pearson's r. As the literature related to Delphi procedure describes, it can be further confirmed that stability is obtained at the third round of Delphi and hence, three rounds of questionnaire are enough in a Delphi study.

References

Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Glenview, IL: Scott Foresman and Company.

BEST PARAMETRIC METHOD FOR ANALYZING DELPHI DATA

Dunham, R. B. (1996). *The Delphi technique*. Retrieved September 20, 2004 from http://www.slais.ubc.ca/resources/research_methods/group.htm#delphi

George, D. & Mallery, P. (2005). *SPSS for Windows step by step: A simple guide and reference, 12.0 update*. Boston, MA: Pearson Education.

Jones, C. M. (1994). *The component skills of workplace literacy and the utilization of computer assisted instruction to achieve it*. Unpublished doctoral dissertation, Kent State University.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Boston, MA: Addison-Wesley.

Ludwig, B. (1997). Predicting the future: Have you considered using the Delphi methodology? *Journal of Extension*, 35(5). <http://www.joe.org/joe/1997october/tt2.html>.

Riggs, W. E. (1983). *The Delphi technique: An experimental evaluation*. *Technological Forecasting and Social Change*, (23), 89-94.

Rowe, S. E. (2001). *Development of a test blueprint for the National Association of Industrial Technology certification exam*. Doctoral dissertation, Iowa State University, Ames, IA.

Shah, H. A. (2004). *A Delphi study to develop engineering management curriculum at Eastern Michigan University*. Unpublished Master's Thesis, Eastern Michigan University, Ypsilanti, MI.

Tillman, T. (1989). *A Delphi study to identify fundamental competency areas for Certification Testing of Manufacturing Technologists and entry-level manufacturing engineers*. Unpublished Doctoral thesis, Purdue University.

Yang, Y. N. (2003). *Convergence on the guidelines for designing a web-based art-teacher education curriculum: A Delphi study*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.