# Sample Size Required For Statistical Techniques

## ABSTRACT

**Purpose** – This paper aims to present the guidelines given in the literature as to the appropriate sample size for the various statistical techniques (factor analysis, regression analysis, and structural equation modelling).

**Design/methodology/approach** – Data for the research were collected through secondary sources.

**Conclusions** – Sample size estimation depends on the nature of research and statistical technique to be employed in research. Most of the statistical techniques are sample size sensitive. (a) The chi-square is sensitive to sample size; its significance becoming less reliable with sample sizes above 200 or less than 100 respondents. In large samples, differences of small size may be found to be significant, whereas in small samples even sizable differences may test as non-significant. (b) For factor analysis appropriate sample sizes depend upon the numbers of items available for factor analysis; for 10 items a sample size of 200 is required; for 25 250; for 90 items 400 and for 500 items a sample size of 1000 deemed necessary. (c) For multiple regression analyses the desired level is between 15 to 20 observations for each predictor variable. (d) Sample size for conjoint studies generally ranges from about 150 to 1,200 respondents; for non-comparative group a sample size of 300 respondents seems reasonable while for comparative groups 200 respondents for each group are required. (e) For SEM at least 15 cases per measured variable or indicator are needed.

**Research limitations/implications** – Exposure of author is limited to social scinces only therefore application of these findings may be more generalizeable for social scinces. Seconldy this research paper only provide the guidlines for sample size estimation needed for statistical techniques rather than mathematical computation.

**Practical implications** – Young social scientists and doctoral candidates are always enquitsive about the adequacy of sample size for their research where they might be applying sophisticated multivariate techniques.

**Originality/value** – Although there are other studies providing a numerical solution to the problem; this paper answers to the demanding question from less-experienced researchers who would like to have some rule of thumb to decide upon. By contributing to the body of knowledge in this area, this research adds significant value.

**Key words:**
Sample Size, Factor Analysis, Regression Analysis, Structural Equation Modelling

## INTRODUCTION

Sample size estimation is the most difficult task of research study employing statistical techniques. Most of the methods to estimate the sample size are based on statistics, and requires computations. Expereinced researchers developed heuristics for quickly determining sample sizes based on experience, rules-of-thumb, and budget constraints. This paper offers reliable literature for successful and meaningful sample-size determination.

## LITERATURE REVIEW

There are a number of different guidelines given in the literature as to the appropriate sample size needed for a typical type of research and required for the various statistical techniques i.e., Chi-Square, factor analysis, multiple regression analysis, conjoint analysis and structural equation modelling.

### *Type of Research*

One of the most important dimensions to consider while estimating sample size is the type of research. For exploratory research or problem identification research minimum acceptable sample is 500 respondents [range 1000-2500]. For confirmatory studies or problem solving research a sample size of 200 respondents is adequate [range 300-500].

### *Sample Size for Chi-Square $\chi^2$*

Despite numerous ambiguities associated with interpreting $\chi^2$, the value of the statistic itself holds the most promise for the development of an index of fit for which the sampling distribution is known (Hoyle and Panter, 1995). It also forms the basis for nested model comparison and these values must be accompanied by the values of degrees of freedom, and sample size. However, $\chi^2$ is sensitive to sample size; as the sample size increases (generally above 200), the $\chi^2$ test statistic has a tendency to indicate a significant probability level. In contrast, as the sample size decreases (generally below 100), the $\chi^2$ test statistic is prone to indicate non-significant probability levels (Schumacker, 1996).

The chi-square is sensitive to sample size; its significance becoming less reliable with sample sizes outside this range. This makes it more likely to reject the model in structural equation modelling due to the larger the sample size. In large samples, differences of small size may be found to be significant, whereas in small samples even sizable differences may test as non-significant (Fan, Thompson, & Wang, 1999; Marsh, Balla, & McDonald, 1988).

### *Sample Size for Factor Analysis:*

Different guidelines are available for judging the adequacy of the sample size for factor analysis. One guidline suggest 50 as very poor; 100 as poor; 200 as fair; 300 as good; 500 as very good, and 1,000 as excellent numbers for sample size estimates (Comrey & Lee (1992). Another guidline suggest that atleast 200 respondents must be sampled in order to attain a stable solution through factor analysis (Thompson, 2004). Another reseaccrh suggest that adequate sample size is not simple as listed above, and it needs to be quite large (e.g., 400 or greater) to produce undistorted results (MacCallum, Widaman, Zhang, & Hong, 1999). Appropriate sample sizes depend upon the numbers of items available for factor analysis as shown in Table below (Meyers, Gamst & Guarino, 2006).

| Appropriate sample size for factor analysis | |
| --- | --- |
| **Number of items** | **Sample Size** |
| 10 | 200 |
| 25 | 250 |
| 90 | 400 |
| 500 | 700-1000 |

*Sample Size for Regression Analysis***:**
For multiple regression analyses the desired level is between 15 to 20 observations for each predictor variable (Hair, Black, Babin, Anderson, Tatham, 2006). However, if a stepwise procedure is employed, the recommended level must be increased to acquire a reasonable level of generalisability from the results. For example the predictor variables are 30, which require a sample size of 600, but raising this to a sample size of 1,000 will considered adequate as further analyses may require step-wise regression.

*Sample Size for Conjoint Analysis*
Sample size for conjoint studies generally ranges from about 150 to 1,200 respondents and it largely depends on the purpose of research (Orme, 2010). For robust quantitative research where subgroup comparasion is not the purposes, at least 300 respondents are required to make a meaninigfull insight. For investigational work and developing hypotheses about a market, between thirty and sixty respondents may do. If the purpose of research is to compare groups of respondents and detect significant differences, sample size to accommodate a minimum of about 200 per group would be required. Therefore, for a study planning to divide respondents into as many as four groups (i.e., through cluster analysis) it would be wise to include, at a minimum, $4 \times 200 = 800$ respondents (Orme, 2010).

*Sample Size for Canonical Correlation*
The sample size depends on reliability of variables. For variables with reliability around 0.8 about 10 cases are needed for every variable. For very high reliability higher a much lower ratio of cases to variables is acceptable (Tabachnick & Fidell, 1996).

*Sample Size for Structural Equation Modelling***:**
Sample size plays an important role in the estimation and interpretation of SEM results (Hair, Black, Babin, Anderson, Tatham, 2006). In general the literature suggests that sample sizes for structural equation models commonly run in the 200 to 400 range for models with 10 to 15 indicators. At least 100 cases are required for SEM and preferably 200 (Loehlin, 1992). With more than 10 variables, sample sizes under 200 generally cause parameter estimates to be unstable and the tests of statistical significance to lack power. Furthermore, Loehlin explained a rule of thumb i.e., the sample size should be at least 50 more than eight times the number of variables in the model. Another rule of thumb, suggested by Stevens (2002), is to have at least 15 cases per measured variable or indicator, while Bentler and Chou (1987) describe the minimum requirements as five cases per parameter estimate. For example a study that uses a maximum of 50 variables in any single model which requires a sample size of about 450 as per the guidelines provided above.

For smaller sample size or excessive kurtosis in SEM, the researcher should report the results in maximum likelihood (ML) estimation method as ML performs reasonably well under a variety of less-than-optimal analytic conditions (Hoyle & Panter (1995).

## CONCLUSIONS

Sample size estimation depends on the nature of research and statistical technique to be employed in research. Most of the statistical techniques are sample size sensitive. (a) The chi-square is sensitive to sample size; its significance becoming less reliable with sample sizes above 200 or less than 100 respondents. In large samples, differences of small size may be found to be significant, whereas in small samples even sizable differences may test as non-significant. (b) For factor analysis appropriate sample sizes depend upon the numbers of items available for factor analysis; for 10 items a sample size of 200 is required; for 25 250; for 90 items 400 and for 500 items a sample size of 1000 deemed necessary. (c) For multiple regression analyses the desired level is between 15 to 20 observations for each predictor variable. (d) Sample size for conjoint studies generally ranges from about 150 to 1,200 respondents; for non-comparative group a sample size of 300 respondents seems reasonable while for comparative groups 200 respondents for each group are required. (e) For SEM at least 15 cases per measured variable or indicator are needed.

## RESEARCH LIMITATIONS

Exposure of author is limited to social scinces only therefore application of these findings may be more generalizeable for social scinces. Seconldy this research paper only provide the guidlines for sample size estimation needed for statistical techniques rather than mathematical computation.

## PRACTICAL IMPLICATIONS

Young social scientists and doctoral candidates are always enquitsive about the adequacy of sample size for their research where they might be applying sophisticated multivariate techniques.

## ORIGINALITY/VALUE

Although there are other studies providing a numerical solution to the problem; this paper answers to the demanding question from less-experienced researchers who would like to have some rule of thumb to decide upon. By contributing to the body of knowledge in this area, this research adds significant value.

## References

Bentler, P.M., Chu, C.P., (1987) Practical issues in structural modelling, *Sociological Methods and Research*, 16, 78- 117

Burns, A.C., & Bush, R.F., (2006). *Marketing Research*, 4th Ed.  Upper Saddle River, N.J.: Pearson/Prentice Hall.

Comrey, A. L., &  Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Fan, X., Thompson, B. &  Wang, L. (1999). Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling* (6), 56-83.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., (2006) *Multivariate Data Analysis,* 6th Edition, Prentice Hall, New Jersey.

Hoyle, R.H. (Ed.) (1995). *Structural Equation Modelling. Concepts, Issues, and Applications*, Thousand Oaks, CA: Sage Publications.

Hoyle. R.H. &  Panter, A.T. (1995). *Writing About Structural Equation Models.* In Hoyle, R.H. (Ed.) (1995). *Structural Equation Modelling. Concepts, Issues, and Applications*, Thousand Oaks, CA: Sage Publications.

Loehlin, J.C., (1992) *Latent Variable Models: An introduction to factor, path and structural analysis*, 2nd ed, Hillsdale, New Jersey

MacCallum, R.C., Widaman, K.F., Preacher, K.J., & Hong, S. (2001) Sample Size in Factor Analysis: The Role of Model Error.  *Multivariate Behavioral Research*;  36(4), 611-637

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness of fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin* 103, 391-410.

Meyers, L.S., Gamst, G., Guarino, A.J., (2006) *Applied Multivariate Research: Design and Interpretation*, Sage Publications

Schumacker, R.E, (1996). Editor's Note, *Structural Equation Modelling*. 3 (1), 1-3.

Tabachnick, B.G., & Fidell, L.S., (1996). *Using Multivariate Statistics.* 3rd Ed. New York: Harper-Collins.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Orme, B. (2010) *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. 2nd Ed, Madison, Wis.: Research Publishers LLC.