SAMPLE SIZE ESTIMATION USING KREJCIE AND MORGAN AND COHEN STATISTICAL POWER ANALYSIS: A COMPARISON

Chua Lee Chuan
Jabatan Penyelidikan

ABSTRACT

In most situations, researchers do not have access to an entire statistical population of interest partly because it is too expensive and time consuming to cover a large population or due to the difficulty to get the cooperation from the entire population to participate in the study. As a result, researchers normally resort to making important decisions about a population based on a representative sample. Hence, estimating an appropriate sampling size is a very important aspect of a research design to allow the researcher to make inferences from the sample statistics to the statistical population. The power of a sample survey lies in the ability to estimate an appropriate sample size to obtain the necessary data to describe the characteristics of the population. With that as the rationale, this article was written to make comparison between two commonly used approaches in estimating sampling size: Krejcie and Morgan and Cohen Statistical Power Analysis. It also highlights the significance of using Cohen's formula over Krejcie and Morgan's for higher accuracy to base decisions on research findings with confidence.

INTRODUCTION

For most studies that require data from a wide and diverse population size, rarely do researchers cover the whole population. The normal practice is to draw a sample from the target population. Salant and Dillman (1994) defined a sample as a set of respondents selected from a larger population for the purpose of a survey. The main reason to sample is to save time and money. Furthermore, it is generally not necessary to study all possible cases to understand the phenomenon under consideration (Ary, Jacobs, & Razavieh, 1996). The most important thing taken into consideration is that the sample drawn from the population must be representative so that it allowed the researcher to make inferences or generalisation from the sample statistics to the population understudied (Maleske, 1995). If the sample size is too low, it lacks precision to provide reliable answers to research questions investigated. If the sample size is too large, time and resources could be wasted often for minimal gain. Therefore, the power of a sample survey actually lied in the ability to obtain the necessary information from a relatively few respondents to describe the characteristics of the entire population.

DETERMINING THE SAMPLE SIZE

For any research, the sample size of any study must be determined during the designing stage of the study. However, before determining the size of the sample that needed to be drawn from the population, a few factors must be taken into consideration. According to Salant and Dillman (1994), the size of the sample is determined by four factors: (1) how much sampling error can be tolerated; (2) population size; (3) how varied the population is with respect to the characteristics of interest; and (4) the smallest subgroup within the sample for which estimates are needed.

Using the above methods as a guideline, the following section aims to compare two approaches in determining the sample size of a population of 500 people using (a) Krejcie and Morgan (1970) and (b) Cohen Statistical Power Analysis.

Krejcie and Morgan

Estimation of sample size in research using Krejcie and Morgan is a commonly employed method. Krejcie and Morgan (1970) used the following formula to determine sampling size:

s	=	$X^2NP\ (1-P) / d^2\ (N-1) + X^2P(1-P)$
s	=	required sample size
X^2	=	the table value of chi-square for one degree of freedom at the desired confidence level
N	=	the population size
P	=	the population proportion (assumed to be .50 since this would provide the maximum sample size)
d	=	the degree of accuracy expressed as a proportion (.05)

Based on Krejcie and Morgan's (1970) table for determining sample size, for a given population of 500, a sample size of 217 would be needed to represent a cross-section of the population. However, it is important for a researcher to consider whether the sample size is adequate to provide enough accuracy to base decisions on the findings with confidence. Therefore, in order to find out if the sample size recommended by Krejcie and Morgan (1970) is sufficient, the next section aims to illustrate the estimation of sampling size using Cohen's (1988) statistical power analysis.

Cohen Statistical Power Analysis

According to Cappelleri and Darlington, (1994), Cohen Statistical Power Analysis is one of the most popular approaches in the behavioural sciences in calculating the required sampling size. According to Cohen (1998), in order to perform a statistical power analysis, five factors need to be taken into consideration:

1. significance level or criterion
2. effect size
3. desired power
4. estimated variance
5. sample size

Cohen (1988) statistical power analysis exploits the relationships among the five factors involved in statistical inferences. For any statistical model, these relationships are such that each is a function of the other four. Taking that into consideration, it means that if sample size is to be determined, it can be estimated for any given statistical test by specifying values for the other four factors: (1) significance level, (2) effect size, (3) desired power and (4) estimated variance. When Cohen's statistical power analysis is used to determine the sample size, the objective of the analysis is to calculate an adequate sampling size so as to optimise as opposed to maximising sampling effort within the constraint of time and money. Optimising sampling efforts will avoid situations where lack of subjects is considered giving rise to inconclusive inference-making. Contrary, maximising sampling efforts occur when the collection of data goes beyond the required level to achieving significant results, thereby, limited resources are wasted.

In order to determine an adequate sample size, the values of significance level, effect size, power and estimated variance have to be pre-determined.

The statistical level of significance for most studies in the teaching field is often fixed at alpha = .05. Alpha is the probability of wrongly rejecting the null hypothesis, thus committing Type I error. Assigning a less stringent alpha would increase the risk of false rejection or 'crying wolf' (Eagle, 1999), casting doubts on the validity of the results. However, if the alpha is too conservative, evidence from the findings might fail to reject the null hypothesis in the presence of substantial population effect. Therefore, setting the alpha at .05, is considered the most conventional level of significance, which is normally used in the field of education. (Ary, et al., 1996).

The next factor to be determined is the effect size. Effect size generally means the degree to which the phenomenon is present in the population or the degree to which the null hypothesis is false (Cohen, 1988). It essentially measures the distance or discrepancy between the null hypothesis and a specified value of the alternative hypothesis. Each statistical test has its own effect size index. All the indexes are scale free and continuous ranging from zero upwards (Cohen, 1992). For any statistical test, the null hypothesis has an effect size of zero. For example, in using the product-moment correlation to test a sample for significance, the effect size index is r, and H_0 posits that r = 0. For multiple regression, the effect size index is f^2 and H_0 posits that $f^2 = 0$.

Effect size can be measured using raw values or standardised values. Cohen has standardised effect sizes into small, medium and large values depending on the type of statistical analyses employed. The effect sizes to test the significance of product-moment correlation coefficient, r, are, .10, .30, and .50, for small, medium and large respectively. For regression analysis, the effect size index, f^2 for small, medium and large effect sizes are $f^{2\,=}$ .02, .15, and .35 respectively. The smaller the effect size, the more difficult it would be to detect the degree of deviation of the null hypothesis in actual units of response. Cohen (1992) proposed that a medium effect size is desirable as it would be able to approximate the average size of observed effects in various fields. Cohen (1992) also argued that a medium effect size could represent an effect that would likely be "visible to the naked eye of a careful observer" (p156).

Next to determine is the statistical power. The power of a statistical test is defined as the probability that a statistical significance test will lead to the rejection of the null hypothesis for a specified value of an alternative hypothesis (Cohen, 1988). Power analysis has the ability to reject the null hypothesis in favour of the alternative when there is sufficient evidence from a collected sample that a value of a parameter from the population of interest is different from the hypothesised value (High, 2000). Putting it simply, it is the probability of correctly rejecting the null hypothesis given that the alternative hypothesis is true.

In statistical parlance, power is expressed as $1-\beta$, where β is the probability of wrongly accepting the null hypothesis when it is actually false or failure to reject null hypothesis that is false. This is known as committing Type II error. The value can range between zero to one.

According to Thomas and Juanes (1996), power analysis is a critical component in designing experiments and testing results. However, computing power for any specific study can be a difficult task. High (2000) argued that when low power is used in a study, the risk of committing Type II error is higher, that is, there is little chance of detecting a significant effect, which can give rise to an indecisive result. Stating it differently, the effect is there but the power is too low to detect it. However, if the power is set too high, a small difference in the effect is detectable, which means that the results are significant, but the size of the effect is not practical or of little value. In addition, a larger power would result in a demand for N that is likely to exceed the resources of the researcher (Cohen, 1992). To avoid these problems, Cohen (1992) suggested fixing the power at .80 (β = .20), which is also a convention proposed for general use. However, this value is not fixed. It can be adjusted depending on the type of test, sample size, effect size as well as the sampling variation.

The fourth and last factor to determine is standard deviation, which is often used for estimating the variation in the response of interest. This value can be obtained, either from previous studies or pilot studies. However, when standardised measures are dimensionless quantities, the sampling variance is already implicitly incorporated. Such standardised measures include the d-values or correlation coefficients and as such the value of variance is not required (Thomas & Krebs, 1997). Therefore if study aims to look at the correlation of variables, this value is not needed for calculating the sample size of the study.

Using the factors mentioned above to estimate sample size, the next section aims to illustrate the use of the Cohen Statistical Power Analysis to calculate an adequate sample size. However, before the sample size is estimated, researchers need to predetermined factors pertaining to alpha size, effect size and power. Additionally, it is also important for researchers to know the underlying objectives of the study and how the data will be analyzed to achieve the objectives. This is because, the sampling size varies according to the type of statistical tests performed on the data gathered.

For instance, the factors pre-determined in order to estimate an adequate sample size for a study are, the alpha level is set at .05, the effect size is medium and the power is set at .80. For illustrative purposes, two statistical tests will be used to analyse the data of a study, such as, Pearson Product Moment Correlation and Multiple regression analysis.

Using the predetermined values and the two statistical tests as guidelines, the next section will illustrate on how to calculate a suitable sample size.

ILLUSTRATIVE EXAMPLES

The sample size for Pearson Product Moment Correlation Analysis and Multiple Regression Analysis can be easily determined using Cohen statistical power analysis.

Pearson Product Moment Correlation Analysis

If a study aims to find out the degree of the relationship (non-directional) between a dependent variable and ten independent variables, with a predetermined effect size of $r = .30$ (medium), a significant alpha = .05 and a statistical power of .80, the desired sample size to test these relationships as indicated in Table 3.4.1 is 85 (Cohen, 1992). This means that 85 respondents are sufficient to perform this statistical analysis.

Regression Analysis

If the study also aims to investigate the contribution of each of the ten predictor variables towards the variance of a dependent variable. This investigation required the use of multiple regression analysis. To estimate the sample size for the regression analysis segment of the research, Cohen's formula takes into consideration the number of k independent variables used in the analysis. Calculation can performed on the maximum of 10 independent variables (u = 10). With the specified power of .80, a medium effect size of $f^2 = .15$, a significant alpha of .05, Cohen's statistical power analysis formula to calculate the sample size needed for this analysis is

$$N = \lambda / f^2$$

This formula required the determination of unknown lambda value, λ, which is then needed to find the necessary sample size, N. However, the lambda value depended on the degree of freedom of the denominator of the F ratio, v.

$$v = N - u - 1$$

To account for this problem, a trial value of v is taken from Table 9.4.2 (Cohen, 1988) to obtain the lambda value which is needed to compute N. If the computed N implied v substantially differed from the trail value, the computation of the new v value has to be used.

For a trail value of v = 120, $\lambda = 17.4$ (Table 9.4.2, Cohen, 1988). Substituting λ into the sample size formula ($N = \lambda / f^2$), gives N = 17.4/.15 = 116, which implied that v = 116-10-1 = 105.

However a more accurate value for N required reiteration by interpolating between lambda values for v = 60 and v = 120 where

$$\lambda = \lambda_L - \frac{1/v_L - 1/v}{1/v_L - 1/v_U} (\lambda_L - \lambda_U)$$

λ_L = lambda value when v = 60
λ_U = lambda value when v = 120
v_L = lower v value
v_U = upper v value

When v = 60, $\lambda = 18.7$, when v = 120, $\lambda = 17.4$. Substituting λ into the formula, the exact λ value =

$$\lambda = 18.7 - \frac{1/60 - 1/105}{1/60 - 1/120} (18.7 - 17.4) = 17.58$$

Therefore N = 17.58/.15 = 117. The result showed that reiteration by interpolating for λ between v = 60 and v = 120 did not significantly change the previous value. Therefore, no further iteration is necessary and the originally computed N = 116 is maintained.

COMPARISON OF ESTIMATED SAMPLING SIZE

Based on the above calculation using Krejcie and Morgan (1970), the estimated sampling size for a population of 500 is 217. However, the estimated sampling size calculated using Cohen (1992) differs according to the type of statistical tests employed by the researcher. The sample size that is required for a correlational study is 85 while a multiple regression analysis requires 116. This indicates that the sampling size can range from a minimum of 85 for performing correlation analysis to a maximum of 217 as recommended by Krejcie and Morgan (1970).

Estimating an adequate number of respondents is critical to the success of a research. According to High (2000), the size of the study sample is critical to producing meaningful results. When there were too few subjects, it might be difficult to detect the effect or phenomenon understudied, thus providing inconclusive inference-making. On the other hand, if there were too many subjects, even trivially small effect could be detected, but the findings would be of insignificant value, wasting valuable time and resources.

Most studies are conducted using Cohen's (1988) statistical power analysis as the guideline for estimating the desired sample size. A few reasons justified the use of this analysis. First, Cohen is not only concerned about the magnitude with regards to the statistical test results and its accompanying ρ value (as most researchers are) but also the existence of the phenomenon understudied by considering additional factors such as population effect size and the statistical power. In most research, significance testing is heavily preferred to confidence interval estimation (Cohen, 1992). They failed to consider the importance of effect size and the statistical power, which has been established in the preceding section. Considering all these factors as suggested by Cohen (1988) would lead to more meaningful results than results that have been inferred from the observed p-value. Furthermore, lacking of controversies among methodologies on the importance of Cohen's (1988) statistical power analysis and the availability of ample resources for estimating sample sizes in research designs using power analysis, this analysis has achieved high reliability for determining an appropriate sample size.

Based on the above justifications, the sample size calculated using the formula derived from Cohen's Statistical power analysis would be more meaningful and acceptable. A sample size of 217 as recommended by Krejcie and Morgan (1970) would be too large a number. Conducting a study, which involved too many subjects than what is deemed necessary would mean that valuable time and resources were not used efficiently and economically.

THE FINAL SAMPLE SIZE

The sample size can be increased to N = 120, slightly more than the recommended size. This number can be rounded up (from 116) to allow the researcher to execute Cohen's (1988) table for further analysis of the power level. A sample size of 120 would be sufficient to answer research objectives using, both, correlation analysis and multiple regression analysis.

With a new sample size of N = 120, it is necessary to estimate the power value for the two statistical analyses. This is essentially important to ensure that the predetermined power value of .80 be achieved, which according to Cohen (1998), is the probability that a statistical significance test can gather enough evidence to correctly reject the null hypothesis in favour of the alternative hypothesis. Since the value of power varied with the type of statistical analysis performed, still maintaining the predetermined statistical criterion of .05 and medium effect size (r = .30), the power value for correlation analysis is increased from .80 to .92 (Cohen, 1988: Table 3.3.5).

As for multiple regression analysis, the calculation of power also required the values of the predetermined factors. Based on the sample size of 120 and the predetermined statistical criterion $\alpha = .05$, medium effect size, $f^2 = .15$, the calculated power for multiple regression analysis is .807.

Based on the calculated power values for the two statistical analyses, with a sample size of 120, the values ranged from .80 to .92. These reported values achieved the minimum proposed value of .80 from Cohen (1988).

Considering the seriousness of type I and type II errors and the cost of obtaining data, this sample size is adequate and manageable. A sample size of 120 is adequate as it has the ability to detect an effect at the desired power equal to a minimum of . 80 or even larger. Reducing the sample size would reduce the power value to below .80, which would be undesirable. As a result, 120 respondents can be randomly selected from the target population to participate in this study.

REFERENCES

Ary, D., Jacobs, L. C., & Razavieh, A. (1996). *Introduction to research in education*. Orlando, Florida: Harcourt Brace College Publishers.

Cappelleri, J. C., & Darlington, R. B. (1994). The power analysis of cutoff-based randomized clinical trials. *Evaluation Review*, 18(2).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin*, 112(1), 155-159.

Eagle, P. C. (1999). *Power: A Primer and other take home messages*. http://www.mpl-pwrc.usgs.gov/powcase/primer.html. Accessed on 1 December 2002.

High, R. (2000). Important factors in designing statistical power analysis studies. *Computing News*, Summer issue, 14-15.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30, 607-610.

Maleske, R. T. (1995). *Foundations for gathering and interpreting behavioral data*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Salant, P., & Dillman, D. A. (1994). *How to conduct your own survey*. New York: John Wiley & Sons, Inc.

Thomas, L., & Juanes, F. (1996). The importance of statistical power analysis: An example from animal behaviour. *Animal Behaviour*, 52, 856-859.

Thomas, L., & Krebs, C. J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America*, 78(2), 126-139.