Measuring effectiveness of IT surveys through statistical sampling scheme

Dr. Pankaj Nagar*
Mr. A.K.Gupta**
Dr. V.P.Gupta***

Abstract

In order to measure IT effectiveness, most firms will need to administer a survey to staff, suppliers and customers /end users .To do this, and considerable knowledge and skill relating to survey design, sample choice and so on is required If handled correctly, a survey can provide detailed information about system effectiveness, but conducted poorly, the survey may be useless and a complete waste of time and money. It is not possible to describe a single, concise way in which to conduct a survey using questionnaires. There are many different approaches to a survey and the appropriateness of the approach is entirely dependent on the particular circumstances being addressed. Furthermore survey design is still regarded as more of an art than a science. It is generally agreed that each study using questionnaires is unique.The sampling scheme is one of the important issue to be considered.

Introduction

All surveys require the selection of those individuals who are to provide the information. This set of individuals is called the sample. The sample comes from much larger group of individuals or objects called the target population. The target population referred to as the population in the sequel is that group about which it is intended to make generalized statements based on the sample findings .The sample is ideally chosen so that no significant differences exist between sample and population in any important characteristics.

Sampling has problems,however,it must be ensured that the sample is representative of the whole population, or the result will be biased and therefore will not be applicable to the whole population. Sampling also has problems of variability. Even if great care is taken to avoid bias, the sample will never be exactly like population. If another sample is chosen in exactly the same way, it would be different. This places a limit on how accurate the sample can be, and therefore how accurately statements can be made about the population.

Choice of Sampling Frame

The sampling frame is a comprehensive list of individuals or objects from which the sample is to be drawn. In practice, the findings of a simple up to date list are highly unlikely; combining more than one list can possibly help improve matters.

**Department of Statistics,University of Rajasthan, Jaipur,Rajasthan(India), pnagar121@gmail.com*
***Assistant Professor, I.M.S.-Ghaziabad, (U.P.)-India, sarthakcc@gmail.com*
****Professor, K.I.E.T.-Ghaziabad, (U.P.)-India, vpguptajp@yahoo.com*

Types of Samples

Sampling techniques fall into two broad categories, namely non-probability samples and probability samples.For a non-probability sample there is no way of estimating the probability of an individual having been included in the sample. Such a sample can occur when individuals are included in the sample on a ‘first arrive first questioned basis’ and as a consequence it is not possible to answer that the sample is representative. Examples of non probability samples include convenience sample judgment samples, and quota samples. In probability sampling each individual has a known not necessarily equal, probability of being selected. Example of probability sampling includes simple random sampling, systematic sampling, stratified sampling, cluster sampling and multistage sampling. Probability samples can be rigorously analyzed by means of statistical techniques whereas for non probability samples this is not possible.

Probability Samples

In obtaining a probability sample, use is made of some random procedure for the selection of the individuals or objects .This is done so as to remove the possibility of selection bias. In simple random sampling each member of the population has an equal chance of being selected. Numbering individuals in the sampling frame and then selecting from these by some random procedure can achieve this.

A systematic sample is selected from the sampling frame of size N in the following manner. Having decided what size sample n is to be selected from the sampling frame, calculate: [N/n] where [ ] denotes the largest integer I < = N/n

Now select a random number I,say, in the interval 1 <= i <=I

The sample size n then consists of ith; $(i+1)^{th}$; $(i+2I)^{th}$; and so on, up to the $(I+(n-1)I)^{th}$ item from the sampling frame.

Should there be some pattern present in the sampling frame. Then such samples will be biased. For example. a systematic sample from the daily sales of a supermarket could result in picking out sales figures for Saturdays only.

In stratified sampling the population is subdivided into homogeneous groups, called strata, prior to sampling. Random samples are then drawn from each of the strata and aggregate of these forms the stratified sample.

In Cluster Sampling, the population is made up of groups, called clusters, where the clusters are naturally formed groups such as companies or locational units.

Size of Sample

Determination of the sample size is a complex problem. Factors which need to be taken into consideration include type of sample , variability in the population , time costs, accuracy of estimates required and confidence with which generalizations to the population are made.

There exists formula for computing sample size, which are based on sound scientific principles. In practice, the sample size resulting from the application of the formula is not slavishly adhered to and is frequently ignored. In fact the sample size chosen tends to be one that fits in with company policy or is/are regarded as credible because it has been used by others conducting similar studies in the past.

Statistical determination of Sample Size

This section describes two situations encountered in practice, namely, how to determine the sample size for estimating a population mean to a specified margin of error, or accuracy with a specified level of confidence; and how to determine the sample size needed to estimate a population proportion (or percentage) to a specified margin of error or accuracy with in a specified level of confidence. These formulas only apply for probability samples taken from a very large population where the sample will be less than 10 % of the population.

Sample size to estimate the mean

Suppose you wish to estimate the true average of a system's response time . In order to estimate this, a random sample of response times is taken and the average of these used to estimate the system's actual mean response time. The question

$\sum 2$ now addressed is, what size of sample is needed to be 95% confident that the sample mean will be with in E Units of the true mean, where the unit of measurement of E can be in, say seconds or minutes ? E is therefore the accuracy required from the estimate. The sample size is given by:

$$N = (3.84 * \sum * \sum)/(E*E)$$

Where $\sum$ the population standard deviation response times and 3.84 is the constant derived from the normal distribution ensuring 95% confidence level. In practice is inevitably unknown and will have to be estimated. This can be done by using response times for a pilot sample of size np, say, in the sample standard deviation formula:

$$S = \sqrt{1/(np-1)\sum(xi - (\sum x/n))2}$$

Where xi is the np pilot response time and x is the numerical average of the sample response times.

A simpler approach, often used, is to estimate $\sum$ from the range of the pilot sample values.This is done according to the following formula:

S = (max (xi) – min (xi))/4 = range (xi)/4

Should it be required to estimate the mean to the same accuracy E as before, but now with a confidence level of 99% then the sample size is given by

N= (6.66)/ (E*E)

Where 6.66 is the constant derived from the normal distribution ensuring a 99% confidence level?

Sample size to estimate a percentage

Suppose you wish to estimate the actual percentage P, say, of your customers who purchase software from a competing company .Suppose further that you require to know what sample size is needed to be 95% confident that the estimate of P resulting from the sample will be with in E% of the actual percentage, P:

n= 3.84P (100-P)/ (E*E)

The cobeat in this case is that P is not known, as it is the parameter being estimated. In practice the value of P used in the above formula can be estimated in a number of ways. It can be estimated subjectively, or from a pilot sample or taken to be 50 %.

The latter results in the most conservative sample size estimate.

For a 99% confidence level :

N = 6.66 P(100-P)/(E*E)

Where P can be estimated as described above.

Sample Size Correction Factor

As previously stated the above formulae hold strictly only should the target population be infinite, and will provide good approximations should the calculated sample size n be small relative to the target population size N. By small we understand the sample size to be 10% or less of the population size. That is, n/N*100 < 10%

In situations where the sample size (n), as determined by the formulae above, exceeds 10% of the population size (N), n has to be adjusted downwards by applying a sample size correction factor . In this case the required sample size n is given by

n' = n* (N/N+n-1)

Where (N/N+n-1) is the sample size correction factor.Thus use of the sample size n will provide the desired accuracy E. The need for correction often arises in practice. For Example, it is likely to occur should a firm decide to conduct an internal survey among staff using the computer network. In accuracy all that needs to be done is to apply the sample size calculation under the infinite population size assumption and then should ((n/N)*100) be greater than 10% the calculated sample size n has to be multiplied by the sample size correction factor.

Another situation that arises is the need to calculate the accuracy E, say associated with a specific confidence level given a sample size n. For estimating the percentage accuracy associated with 95% confidence and sample size n' :

E' = $\sqrt{(3.84P(100-P)/n')*((N-n')/(N-1))}$

And for estimating the accuracy of the mean associated with 95% confidence and sample size n :

E' = $\sqrt{((3.84*\sum*\sum)/n')}*\sqrt{\ }((N-n')/(N-1))$

Conclusion

Survey design is very much an art and invariably results in economic considerations forcing the researchers to sacrifice what they ideally would require for what is practical in terms of time and money available. It must be accepted that no survey will be found to be perfect. The key to a successful survey is the care taken in carrying out the time consuming and costly upfront work. This includes tasks such as clearly defining the purpose and objectives of the study, the running of focus groups, analyzing transcripts of the focus group meetings, conducting fairly open ended interviews with appropriate persons, and the development and thorough testing of the questionnaire.

References

- Agresti A (1990), *Categorical Data Analysis*, John Wiley & Sons, Inc.
- Bannister F and Remenui D(2000), "Acts of Faith : instinct , value and IT Investment decisions", *Journal of Information Technology,* Vol.15, No.3.
- Berelson A(1980), CITED in Krippendorf K,*Content Analysis*, Sage Publications, Thousand Oaks, CA.
- Brynjolfsson E (1993), "The productivity paradox of information technology: review and assessment", *Communications of the ACM* , December.
- Brynjolfsson E and Hitt L (1995), "Information technology as a factor of production: the role of differences among firms", *Economics of Innovation and New Technology*, Vol. 3 No. 4.
- Economist (1991), "IT Investment", 24 August, pp. 75.
- Ezingeard J N and Race P (1996), "A Survey of current practice of value assessment in the management of manufacturing information and data systems", *Proceedings of the 2nd International Conference on Management Integrated Manufacturing*, 26-28 June, Leicester University.
- Hochstrasser B (1992), "Justifying IT Investments", in Proceedings of the conference on *Advanced Information Systems : The new technologies in today's business environment*, Imperial College, London.

- Malhotra N K (2004), *Marketing Research : An applied orientation*, 4th Edition, Prentice Hall, N J.
- Manly B F J (1994) , *Multivariate Statistical Methods: A Primer,* 2nd Edition, Chapman & Hall, London.

****