

Sampling adequacy and research



A Journal of
the Bombay
Management
Association

Prof. (Mrs.) K. R. Sinimole

Faculty,
St. Francis Institute of Management & Research
(SFIMAR), Borivali

Volume : 11

E-REVIEW: November 09

Dear Members,

Hope you found reading our e-review containing an exhaustive article on 'Tourism' by Dr. M. R. Menon, Dean, Research, Aicar Business School, Neral insightful.

We are happy to enclose an article by Prof. (Mrs.) K. R. Sinimole, Faculty, St. Francis Institute of Management & Research (SFIMAR), Borivali on the theme of how sampling errors can be devastating for good research.

Warm greetings for the season...

Best regards,



Abhijit Phadnis
Chairman, Editorial Board

EDITORIAL BOARD

Chairman

Mr. Abhijit Phadnis

Members

Dr. Pradeep B. Kakkad

Mr. S. Mohan

Dr. Anil Naik

Mr. Gladwyn A. Pinto

Note: The views expressed by the author/s in BMA E-review are their own. The editorial committee neither endorses nor necessarily agrees with the views.

Sampling Error: An unseen demon capable of killing any good research.

Smt. Sinimole K.R.

Abstract

This paper discusses about the unseen demon capable of killing any good research. Quality of data has always been the essence of quality of the research. Statisticians and research analysts can draw conclusion only on the basis of the data they have. To collect the right kind of data from the right sources is a challenging task. In the technological era it is easier to contact a large number of audiences. However, this tends to increase the sampling error as well. If not controlled, it will lead to wrong analysis.

“The world is littered with statistics, and the average person is bombarded with five statistics a day. Statistics can be misleading and sometimes deliberately distorting. There are three kinds of commonly recognized untruths: Lies, damn lies and statistics.”

- Mark Twain

“On almost every occasion when we release a new survey, someone in the media will ask, "What is the margin of error for this survey?" There is only one honest and accurate answer to this question -- which I sometimes use to the great confusion of my audience -- and that is, "The possible margin of error is infinite. Polls are the worst way of measuring public opinion and public behavior, or of predicting elections -- except for all of the others.”“ (The Article “How the Media Confuse and Mislead Readers and Viewers” by Humphrey Taylor)

Sampling error is any type of bias that is attributable to mistakes made either in the selection process for sampling units or in determining the sample size. Sampling errors arise from estimating a population characteristic by looking at only a portion of the population rather than the entire population. Sampling errors

- generally decreases as the sample size increases (but not proportionally)
- depends on the variability of the characteristic of interest in the population
- can be accounted for and reduced by an appropriate sample plan
- can be measured and controlled in probability sample surveys.

Let us go through the points one by one.

Perhaps the most important thing to check for is sample size and margin of error. It is often the case that with small samples, a change in one sample or one data item can completely change the results. Many people are surprised by the small size of well-known surveys. For example, polls that try to predict voting patterns are taken from sample sizes ranging from 1,000 to 2,000 people, with samples of about 1,000 people being the most common. Ratings for television programs are estimated from approximately 2,000 viewers. This small sample represents the television preferences of a total population of millions households! Despite a widely-held perception that such polls are reliable; some statisticians question their accuracy because of the small sample size. **This is surely a classic case of a little knowledge being a dangerous thing.**

In general, the greater the difference between **the population units**, the larger the sample size required to achieve a specific level of reliability. For example, if you were to conduct a survey on work environments for a population where the income varies from 10,000 /- to 15,000/- , you would use a smaller sample size to achieve the same level of reliability than you would use for a population of equal size for which income varies from 3,000 /- to 1,000,000/-.

Before analyzing data statistically, it is important to consider if the data was **collected appropriately**.

- Where did the data come from?
- Who ran the survey? Do they have an ulterior motive for having the result go one way?

- How was the data collected? What questions were asked? How did they ask them? Who was asked?
- The sample design (for telephone surveys, how the numbers were selected and how the individuals are selected within the household);
- The non-availability problem (are people who are available different on the variables we are measuring than the people who are not available?);
- The refusal problem (is the refusal rate different on the particular variable we are measuring?);
- Question wording;
- Question order; Deliberate, or unconscious, lying or false reporting by respondents;
- Inappropriate or inadequate weighting of the data.

All of these variables have been shown in various studies to have been sources of error and sometimes of quite substantial error.

The statistical analysis will only likely be a minor part of the total expense of a properly conducted experiment, so time, effort, and money spent ensuring the data are collected appropriately is certainly well spent. The computer adage **Garbage In, Garbage Out** or **GIGO** is rather apropos.

The **medium used** (mail, phone, personal interview) is important.

Surveys are a very popular method of data collection for social issues. Mail surveys tend to have a lower response rates which will distort and hence flaw a sample. Although telephone surveys may be relatively efficient and inexpensive, the more time consuming and correspondingly expensive personal interview allows more detailed and complex data to be collected. For example, in a country like India, mail surveys are not reliable, since internet penetration in India is low as compared to other countries. By conducting such surveys, the authorities behind it are conveying a wrong message to the 99.9% of the population, that is, the non participants. Same is the case with SMS rating.

The next point to be considered is whether the **sample is representative** of the population. For instance, a technology firm discovered that 40% of all sick days were taken on a Friday or a Monday. They immediately clamped down on sick leave before they realised their mistake. Forty per cent

represents two days out of a five day working week and therefore is a normal spread, rather than a reflection of swathes of feckless opportunists trying to extend their weekends. Similarly if you are conducting a survey to find out the government's success in agricultural operations, through SMS rating, then that sample will definitely not be a representative of the original population.

The next point to be considered is the sample plan. The preceding section has covered the most common problems associated with statistical studies. The desirability of a sampling procedure depends on both its vulnerability to error and its cost. However, economy and reliability are competing ends, because, to reduce error often requires an increased expenditure of resources. Sampling error can be controlled by exercising care in determining the method for choosing the sample. Sampling bias on the other hand may be minimized by the wise choice of a sampling procedure.

How **nonresponse** error is handled.

Lindner, Murphy, and Briers suggested that procedures for handling nonresponse issues be implemented when less than 85% response rate is achieved. To further reduce the threat of nonresponse error, it is recommended that a minimum response rate of 50% be achieved (L. E. Miller, personal communication, December 12, 2001; Fowler, 2001; Babbie, 1990).

Indian scenario

Consider television programmes like Indian Idol or Sa Re Ga Ma. "The popularity of the **Indian idol** finals could be thought of with the number of SMS that reached some where near to 3 crores." (www.indianidol.tblog.com). Even the participants are taking SMS ratings! How many of the viewers really wanted to participate in it? Even though they wanted to participate how many of them can? Mobile penetration in India is too low for such ratings (3 crores are negligible when comparing with total Indian population). Consider another case, where surveys are done by various web sites about the success of Mumbai's disaster management system. What does it mean? Have all those who have really suffered participated in the survey? How the result is interpreted? The questions "**How was the data collected? What questions were asked? How did they ask them? Who was asked?**" are relevant here in this context.

Now the importance of the questions "**Who ran the survey? Do they have an ulterior motive for having the result go one way?**" "

Some individuals might deliberately use a biased sample to produce misleading results. A survey on the effects of passive smoking, sponsored by a major tobacco manufacturer, is hardly likely to be impartial, but on the other hand neither is one carried out by a medical firm with a vested interest in promoting health products.

If a survey on road accidents claims that cars with brand X tyres were less likely to have an accident, check who took part. The brand X tyres may be new and only fitted to new cars, which are less likely to be in accidents anyway.

Even rating scales are used to manipulate the results. Rating scales are used quite frequently in survey research and there are many different kinds of rating scales. A typical rating scale asks subjects to choose one response category from several arranged in hierarchical order. Either each response category is labeled or else only the two endpoints of the scale are "anchored." Unfortunately, there are many ways that a rating scale can be biased. **Dishonest researchers can, of course, manipulate the outcome of their research, if they wish.** For example, An unethical researcher interested in manipulating results could place the desired response on the left side of the scale. For example, instead of rating the product as "excellent" to "very poor" the scale would start with "very poor" and end with "excellent." Although we have no way of determining which scale is more valid, we can be reasonably certain that the latter scale will produce more negative evaluations than the former scale. A researcher can influence results by shrewd selection of the endpoints to anchor the scale. Pollack, Friedman, and Presby (1990) found that a scale anchored at the endpoints with the adjectives "superior" and "terrible" will not produce the same results as one anchored at the endpoints with the weaker adjectives, "very good" and "very bad." Respondents seem to be reluctant to choose extreme descriptors for their response.

The researcher has several ethical responsibilities to the respondents in the sampling process. **It is unethical and misleading to treat such surveys as unbiased surveys and to project the result to a target population.**

So the questions **“Who ran the survey? Do they have an ulterior motive for having the result go one way?”** are very much important.

Based on all points discussed here, one can draw upon a conclusion that “Well designed, well conducted surveys work. Their overall record is pretty good.”

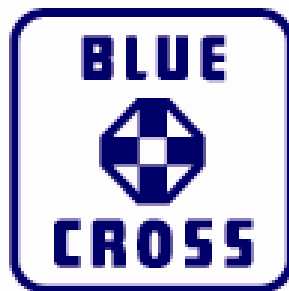
However it all depends on the way the data is collected. The erroneous methodology of collecting the data may allow the demon of sampling error to creep in and destroy the otherwise good research.

References

1. Naresh K. Malhotra , Satyabhushan Dash, Marketing research An applied orientation fifth edition, pearson Education
2. Richard I Levin, David S. Rubin, Statistics for Management Seventh Edition, Prentice-Hall of India
3. Robert E. Stevens, Bruce Wrenn, Morris E. Ruddick, Philip K. Sherwood, The marketing Research Guide Viva books Private Lmted.
4. T.N. Sriviatava, Shialaja Rego Statistics for management, McGraw Hill Education (India) Ltd.

Web References

1. www.bbc.co.uk
2. www.pollingreport.com
3. <http://home.ubalt.edu>
4. <http://www.qualtrics.com>
5. <http://nhits.ornl.gov>



**WORLD CLASS
QUALITY MEDICINES
AT REASONABLE PRICES**

BLUE CROSS LABORATORIES LTD.
PENINSULA CHAMBERS, P.O. BOX 16560, LOWER PAREL, MUMBAI-400 013,
INDIA.