# Chapter 6

## Research Errors Encountered in Data Handling

# 6 Research Errors Encountered in Data Handling

**Felix Kutsanedzie[1]; Ofori Victoria[2]; Selassie Kwasi Diaba[3]; Anthony Deku[1]**

[1]Accra Polytechnic, Accra, Ghana
[2]Agricultural Engineering Department, KNUST, Ghana
[3]Anglican University College of Technology, Nkorazan, Sunyani–Ghana

## Abstract

The human beings are not perfect hence in every human engagements there are elements of errors that are encountered and must be handled in order that results are not drastically affected. Likewise in carrying out a research, there are so many errors that can be encountered in the taking of measurements which eventually affect the reliability and validity of data collected. Statistics or numbers are used for estimation and the extent to which they would perfectly represent the facts depends on plethora of factors such as the instruments, the influence of the researcher and the environment. This chapter concentrates on the various errors that are encountered by a researcher in the conduct of research and how to note them and the measures to adopt in reducing or treating them such that they do not easily influence the results of findings unduly. This thus prevents false interpretations for consumers of research information, and denting the reputation of the researcher in question.

## Keywords

Errors, Scores, Estimation, Interval, Confidence Level

## 6.1 Introduction

Research is defined as a planned enquiry which utilizes the scientific method to develop and test hypotheses and also find solutions that are generally practicable in addressing the myriads of problems in a society. Therefore the two important purposes of research are to develop and test hypotheses, and to apply findings to practically solve problems.

The developing and testing hypotheses is meant to expand the frontiers of knowledge, thus allowing students, researchers, lecturers and other academics to tap from this knowledge pool. Also by applying research findings to solve problems make the world better for all. However, the knowledge contributed by a researcher should be proven to be working and can stand the test of time. Therefore whatever results of findings the researcher comes out with must be foolproof else it would lead to deceiving all those who tap from it as well as those who apply it, which consequentially can have devastating effects. This is why the data collected must be error tight or if not possible must cater for error management in order to serve the empirical characteristic of research.

## 6.2 Types of Scores

Usually in the conduct of research, data is collected and the outcome of the research is dependent on the data obtained. However in collecting data, errors can occur through measuring, recording, use of faulty instruments, researcher's influence, etc. Whenever an instrument is used in taking a measurement, the

value obtained is just an estimation of the variable being measured. Thus the taking of such a measurement once and then using it to conclude can result in grave errors being made, hence the need for replication and averaging of values of variables. The following scores in terms of measurements must be explained to enhance understanding of the errors to be considered in this chapter:

*True score*

The true score is referred to as the actual value of the measurement being taken and it is the value that is expected to be exact estimation of the variable. This cannot easily be obtained by taking only one measurement and it can be a value that lies within a range of measured values of variables. For instance if a researcher is taking the temperature of the body, how will he or she know by taking the reading once that, the value obtained must be the true score? What if the value is taking for the first time and it read 37 ℃, and then the second time 39 ℃, third time 35 ℃, fourth time 30 ℃ and finally the fifth time 40 ℃? Which will be the true score of the variable? This presupposes that the true value is hidden within 30 ℃ and 40 ℃ with the two numbers inclusive. Thus the reason for the instrument not giving the exact true value is due to error.

*Obtained / achieved score*

*Obtained / achieved score* refers to the value obtained when an instrument is used for the measurement of a variable. In this case once a thermometer is used to take the temperature of a body and it reads 37 ℃, this becomes obtained or achieved score. What this means is that for the obtained or achieved scores there are always the element of error because the possibility of another value of the same variable being obtained in the second instance of taking the measurement is ruled out which needs to be checked.

*Error score*

Now the error score is the score associated with error, and it is the error that is associated with the obtained score which prevents getting the true score. Assuming a weight of a body is taking on a scale and it records 60kg and the same body without being tampered with is weighed a minute later with the same scale and records 85kg. What could have happened with this drastic weigh changed in the same body having not been tampered with in anyway? This difference can be attributed to error.

Having explained these types of scores, it is now ripe to introduce the various types of estimation of values of variables recorded; how to account for the errors associated with the data obtained in the conduct of a research study; and the interpretations that can be given to the data for the consumption of the end-users of the findings.

## 6.3 Types of Estimations

There are three main types of estimations: point estimation; interval estimation, and confidence lnterval level estimation. Each of these estimations enable the people to know the errors associated with the data from which the findings emanate in order to know the confidence to repose in them.

*Point Estimation*

With regards to handling data by point estimation, the value of the variable recorded when measurement is taken is considered as the true score. It is assumed that the true score is equal to the obtained or achieved score, disregarding the possibility of error. Thus if a researcher handles data this way, it shows that the possibility of error occurring in the value has not been

accounted for and therefore obtained score is considered as exact, when in actual fact cannot be so. The findings from such researches must be applied cautiously knowing that error has not been catered for.

Data handled in this way when variables are measured once and the values taken as exact without replicated to establish variations demonstrate point estimation. For example using a thermometer to take the temperature of a body once by means of contact at one end of the body and then using the value obtained as the true score. This implies that the researcher is ignoring the fact that there can be researcher fault or calibrating problems of the instruments.

*Interval Estimation*

When interval estimation is used, the value obtained from the measurement of a variable is taken as the obtained or achieved scores and not the true score. It is assumed that the true score lies within a range of the obtained or achieved scores. It presupposes that in other to find the true score more than one measurement must be taken for errors to be accounted for, for the range within or the interval within which the true score lies to be established.

In interval estimation measurement of variables are replicated so that more than one obtained or achieved scores are obtained for the mean of the scores, variations and the standard error to be calculated to account for the error in the data. Assuming the interval estimation is to be used for estimating the temperature of a body say B. A thermometer is used to take the temperature of the body B at various points of contact with the body B as $T_1$, $T_2$, $T_3$, $T_4$, $T_5$ ….etc. $T_1$, $T_2$, $T_3$, $T_4$ and $T_5$ become the obtained or achieved score of the measurement. Let say their respective values are 20, 22, 21, 24 and 26. Once the obtained or achieved scores are identified, the researcher must proceed to find

the range within which the true score lies and then account for the error in the data collected as regard their measurement.

To find where possible, the true score lies, the mean of the obtained or achieved scores is calculated as follows:

$$\bar{x}\ (obtained\ scores) = \frac{T_1 + T_2 + T_3 + T_4 + T_5}{n}$$

$$\bar{x}\ (obtained\ scores) = \frac{20 + 22 + 21 + 24 + 26}{5}$$

$$\bar{x}\ (obtained\ scores) = \frac{113}{5} = 22.6$$

*$\bar{x}$=mean of obtained or achieved scores, $T_1$ ……. $T_5$ represent the obtained scores, n= number of temperature readings taken.*

Thus the mean of the obtained scores is 22.6. Now the standard error of the mean is computed to establish the range within which the true score lies from the mean.

Confidence Interval Level Estimation

The confidence interval level estimation is more like the interval estimation but unlike the latter, it takes into consideration the certainty or the confidence level ascribed to the true score of being found within the established interval. The interval estimation helps to establish the interval or the range within which the true scores lies with respect to the mean of the obtained or achieved scores but in this case the researcher cannot assign any certainty or probability of the true scores being found within the established interval. Like the interval obtained in the interval estimation example given earlier, what confidence level will the researcher have that the true score lies within such an interval? Could one be 100% or 99% or 20% sure that the true score lies within the interval or what? The confidence interval level estimation allows one to know to what level

of confidence to repose in the data collected and thus goes beyond the information that can be provided by interval estimation.

The confidence level interval estimation takes into consideration the assigning of probability or calculating the chances of the true scores being found within the established interval determined from the obtained or achieved scores. For instance when a lecturer examines students via examination over a total of 100 scores, the obtained scores are likely to range between 0 and 100 with both values inclusive. So the interval within which the true scores of the students would lie is from 0 to 100, and the probability of the true scores of the student being found in this range would be 100%, this percentage thus becomes the confidence level.

Likewise when a die is tossed the interval within which the obtained scores would lie are from 1 to 6, however if the true score is expected to be an even number, then the confidence level that can be assigned to the true scores of being found within the established interval is the probability of even number appearing when the die is tossed, which is $\frac{3}{6} = 0.5 = 50\%$. This implies that the research is 50% certain that the true score would be found within the established interval.

## Bibliography

[1] Dong, Y., Peng, C. Y. J. (2013). *Principled missing data methods for researchers. Springer Plus*.

[2] Ercan, I., Ocakoglu, G., Sigirli, D, Ozkaya, G. (2012) Assessment of Submitted Manuscripts in Medical Sciences According to Statistical Errors. Turkiye Klinikleri *J Med Sci* 32(5): 1381-1387.

[3] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. (1998). *Multivariate data analysis*. Prentice Hall, New Jersey..

[4] Little, R. J. (2002). Rubin DB. *Statistical analysis with missing data*. John Wiley and Sons, New York.

[5] Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc*. 83(404): 1198-202. .

[6] Little, R., Rubin D. B. (1987). *Statistical analysis with missing data*. John A. Wiley & Sons, Inc., New York. .

[7] Simpson, E. H. (1951). The interpretation of interaction in contingency tables. J Royal Stat Soc Series B (Methodological), 1951: p. 238-241.

[8] Steyerberg, E. W. (2009). *Clinical prediction models*. Springer, New York. .

[9] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3): 581-92.

[10] Vittinghoff, E., Shiboski, S., McCulloch, C. E. (2005). *Regression methods in biostatistics*. Springer, New York.

[11] McGuigan, S. (1995) The use of statistics in the British Journal of Psychiatry. *Br J Psychiatry* 167(5): 683-688.

[12] He, J., Jin, Z., Yu, D. (2009) Statistical reporting in Chinese biomedical journals. *Lancet* 373(9681): 2091-2093.