Missing Data

Independent Study Activity 1

Orville D. Hombrebueno
odhombrebueno@nvsu.edu.ph

February 8, 2020

Perform the following tasks relative to the reality or inevitability of "missing data" or "missing values".

1. Explain the nature of missing data or missing values and the importance of coding and mechanically treating missing values. Elaborate on the serious implications of not coding missing values to the integrity of a research.

   The importance of dealing with missing was discussed by Ho (2014) in the following:

   > Encountering missing data in research is part and parcel of data analysis. The problem arises for a number of reasons; examples include subjects dropping out of an experimental condition because of boredom/fatigue; refusal to answer a particular question on a survey questionnaire; death; and refusal to participate in the post-test of a longitudinal study. Whatever the reasons, analysis of a "missing data set" can be problematic if the pattern of missing data is non-random. Any statistical results based on such data would be biased to the extent that the variables included in the analysis are influenced by the pattern of non-randomness of the missing data. According to Tabachnick & Fidell (2001), if only a few data points (5%) are missing in a random pattern from a large data set, the problems are less serious and almost any procedure for handling missing values yields similar results. Thus, when screening data, it is important to identify the pattern of missing data (p.23-24).

   Missing data when properly coded can be easily summarized. This summary will contribute to understanding the data gathering procedure and the respondents. This insight can back-up findings in the study. Further more, one can easily understand the limitation of the study when coding and treatment of missing data is properly done and discussed in the study. Otherwise, one might be claiming a finding that has really no basis since the coding and handling of missing data is not properly done.

2. Explain the following labels for missing data/value (e.g. MANN = Missing and not needed, when the characteristic of interest or condition is absent and not necessary).

   2.1. NA

   Not answered. This is a label for an item that has no entry. The respondent did not answer the item.

   2.2. NAP

   Not appropriate. This is the label for items that have entries but are not appropriate for the item. Example: the item is asking for monthly income but the respondent entered his age.

   2.3. NAV

   2.4. DK

   Don't know. The respondent placed "I don't know." as answer to an item.

   2.5. DC

   Don't care. The respondent placed "I don't care." as answer to an item.

   2.6. Refused

   The respondent refused to answer question since the information is confidential.

   2.7. NO

   2.8. NOP

   2.9. DR

   2.10. NS

   2.11. Illigible

   2.12. Patterns of Missing Data

   2.12.1. Missing Completely at Random (MCAR)

   2.12.2. Missing at Random (MAR)

   2.12.3. Missing Not at Random (MNAR)

   To be honest I am not aware of these labels: NAP, NAV, DK, DC, Refused, NO, NOP, DR, NS, Illegible. I don't even know how to use them since I have not yet done research wherein I gather data through questionnaires. I usually use secondary data for research and usually use NA for missing data. NA simply means it is missing.

3. Describe specific instances in your own particular discipline or area of research interest to illustrate (5) different types of the above missing values of data that may be gathered (e.g. MANN: In surveys about facilities for photography, darkroom is no longer necessary as photos can be easily developed using digital printing).

   NA: In any field when using questionnaires to gather data, some respondent simply don't enter a value for the item for some unknown reason.

   MANN: In surveys about

4. Propose three distinct ways of controlling or preventing the occurrence and undesirable sources of "missing" data/values.

5. Differentiate among the following in mechanical and statistical procedures dealing with cases with missing data:

   5.1. Listwise Deletion

   5.2. Pairwise Deletion

   5.3. Mean Replacement

6. In what special instances of missing data can sampled units be subjected for further relevant case analysis instead of listwise or pairwise elimination, or value imputation?

Bibliography

Ho, R. (2014). *Handbook of univariate and multivariate data analysis with ibm spss*. CRC Press.

Horber, E. (2019). *Surveys and missing values*. http://www.unige.ch/ses/sococ/cl/stat/action/diagmiss.act.surveys.html

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Allyn and Bacon.