# What is Missing Data?*

**What should we do about it?**

Fatimah Yunusa

March 5, 2024

## 1 Introduction

In the data acquisition process, no matter how efficient and careful we are, there is always a high possibility that we will have missing data.Missing data poses a challenge when cunducting statistical analysis because it often impacts reliability and validity. This Mini-essay delves into the nature of missing data, its importance, the types, causes of missing data, consequences of missing data and how to properly handle missing data.

### 1.1 What is missing data?

Missing data is when there are missing observations in the data set. It is variables that were not obtained for different observations.

### 1.2 Why is it important to us?

Missing data is important to us because it adds further uncertainty to our statistical analysis. We must think about the possibility of missing data because it also enables us to think about other missing variables that we have not accounted for or provided measures of.

*https://github.com/fatimahsy/What-is-Missing-Data-.git

## 2 Types of Missing Data

When we identify missing data, the first thing we ough to do is try and figure out what type of missing data it is. Identifying the type enables us to understand how to deal with it. There are three main classifications of missing data:

-Missing Completely At Random -Missing At Random -Missing Not At Random

Data are Missing Completely At Random when observations do not show up in the data set across all observations.This is a case that rarely happens but in cases like that,there is less concern relating to summary statistics or inference.

When observations are Missing at Random, they are missing from the data set in relation to the other variables that are included in the data set. This case accounts for data that has some sort of response bias embedded into it.

When observations are Missing Not At Random, this means that the probability of missing data is connected to the values that have not been observed. So the missing data has a relation to the variable or variables being measured. For example if in a medical study, patients with severe side effects are less likely to return for follow-up visits, the missing data on side effects is not random but related to the severity of those side effects.

## 3 What Causes Missing Data?

Data might be missing for many reasons. Data might be missing because of a lack of responses from participants, errors in data entry, failure of research equipment, flaws in the design of the study, deliberate omission of data from respondents, data processing errors and many other unprecedented events.

## 4 Consequences of ignoring Missing Data

Missing data can have several consequences and this is why it is important to identify and deal with them.They can introduce bias to the results, increase variability, lead to wrong inferences, reduce external validity and lead to incorrect conclusions.

## 5 Handling Missing Data

Dealing with missing data requires careful consideration fo many factors.There are many tactics we can use to deal with these missing variables. These include:

- dropping the observation that has missing values

- using multiple imputation
- General data imputation
- Creating a dummy variable
- Using predictive Models
- Assigning weights to different observations

## 6 Conclusion

In conclusion, missing data plays a big role in our statistical analysis. It can potentially lead us to presenting biased results and misleading conclusions. It is important that the researchers understand the different types of missing data and how to handle them. They must decide which method of handling is most appropriate for their data type.