**To check if the data is missing at Random**

Is my variable missing in random? (MAR)

Pattern analysis

Chisquare test of independence missing data vs other variables

Descriptive stats with and without missing value is it same?

**If the data is Missing Completely At Random**

Multiple Imputation with observed data

MCAR

IPTW\*

Imputation with auxiliary data

Complete Case Analysis

Single Imputation

**Descriptive stats with and without missing value is it same?**

1. Compare the descriptive stats of available data with and without missing value
2. If the distribution is similar, then the data is missing at random

**Chi square test of independence missing data vs other variables**

1. Check if the missing data is related or independent to observed data

**Pattern analysis**

1. Visualize the missing data pattern: You can visualize the missing data pattern
2. A heat map can help you identify the missing data pattern by displaying the distribution of missing values across the variables

After the above exploration if the data is in fact missing at random. The following will be done.

**Complete Case Analysis**

1. In this method we discard the missing values from the dataset
2. Cons:
   1. Reduced the statistical power of the analysis
   2. If the data is not completely missing at random, this could reduce the generalizability of the model

**Single Imputation**

1. In this method the missing values are imputed with single value from the observed data either mean or median
2. Median is preferred over mean if we observe outliers in the data as it is less sensitive to outliers

**Multiple imputation with observed data**

1. We’ll perform regression or classification models to predict the missing values using the observed data as predictors

**Imputation with auxiliary data**

1. Using the domain expertise, knowledge from previously published researches to find a value to impute the missing value

**IPTW (Inverse Probability Treatment Weighing)**

1. First propensity score will be calculated for the data points in reference to the observed data
2. Then the model weights the patient who is like that of the patient missing data higher to counter the missing data
3. The weights are further used in the analysis