

Missing Values in Repeated Measurement Designs

Sanchita Khan (STAT-24)

Sumedha Guha (STAT-05)

Shramana Guin (STAT-06)

Suryadeep Ghosh (STAT-28)

Department of Statistics

Third Year, Semester-V

Contents

1. What Is Missing Data?
2. Missing Data in Clinical Trials.
3. Reasons for Missing Data.
4. Effects of Missing Data.
5. Missing Data Mechanism.
6. Types of Missing Data.
7. General Approaches to Missing Data.
8. References.
9. Acknowledgement.

What Is Missing Data?

- ▶ In simple words, missing data are defined as values that are not observed and would be meaningful if they are observed.

What Is Missing Data?

- ▶ In simple words, missing data are defined as values that are not observed and would be meaningful if they are observed.
- ▶ Missing data can be anything from missing sequence, incomplete feature, files missing, information incomplete, data entry error etc. Most datasets in the real world contain missing data.

Missing Data in Clinical Trials

- ▶ Clinical trials comparing different treatments on participants often involve repeated measures of the outcome over time.

Missing Data in Clinical Trials

- ▶ Clinical trials comparing different treatments on participants often involve repeated measures of the outcome over time.
- ▶ It is common in clinical trials for studies comparing the efficacy of several noncurative treatments because they can effectively remove within subject variances.

Missing Data in Clinical Trials

- ▶ Clinical trials comparing different treatments on participants often involve repeated measures of the outcome over time.
- ▶ It is common in clinical trials for studies comparing the efficacy of several noncurative treatments because they can effectively remove within subject variances.
- ▶ However, by the very nature of taking repeated measurements, they also suffer from the common problem of missing values on one or more occasions.

Reasons For Missing Data

- ▶ Patients are in very critical conditions.
- ▶ Patients wants to change the treatment.
- ▶ Missing due to the breakdown of machines .
- ▶ Failed in continuing the follow up.
- ▶ Failed to answer some questionnaires.
- ▶ Patients are cured or died before study.
- ▶ Investigators forgot to collect the data.
- ▶ Family migrated.
- ▶ Patients' profile might be missing.

Effects of Missing Data

- ▶ Power.

Effects of Missing Data

- ▶ Power.
- ▶ Bias.

Effects of Missing Data

- ▶ Power.
- ▶ Bias.
- ▶ Inaccurate results.

Missing Data Mechanism

- ▶ Understanding the reasons why data are missing is important for handling the remaining data correctly.

Missing Data Mechanism

- ▶ Understanding the reasons why data are missing is important for handling the remaining data correctly.
- ▶ In most studies unobserved data are treated as “missing” in the sense that their underlying values that could have been observed but were not.

Missing Data Mechanism

- ▶ Understanding the reasons why data are missing is important for handling the remaining data correctly.
- ▶ In most studies unobserved data are treated as “missing” in the sense that their underlying values that could have been observed but were not.
- ▶ Depending on how the data have become missing, treating unobserved data simply as missing can lead to serious estimation bias.

Missing Data Mechanism

- ▶ Understanding the reasons why data are missing is important for handling the remaining data correctly.
- ▶ In most studies unobserved data are treated as “missing” in the sense that their underlying values that could have been observed but were not.
- ▶ Depending on how the data have become missing, treating unobserved data simply as missing can lead to serious estimation bias.
- ▶ The most appropriate way to handle missing data or incomplete data depends on how the data points became missing.

Types of Missing Data

- ▶ Little and Rubin defined three unique types of missing data mechanism, distinguishing among the different situations according to how the data became missing and incomplete.

Types of Missing Data

- ▶ Little and Rubin defined three unique types of missing data mechanism, distinguishing among the different situations according to how the data became missing and incomplete.
- ▶ They are,

Types of Missing Data

- ▶ Little and Rubin defined three unique types of missing data mechanism, distinguishing among the different situations according to how the data became missing and incomplete.
- ▶ They are,
 1. MCAR: Missing completely at random.

Types of Missing Data

- ▶ Little and Rubin defined three unique types of missing data mechanism, distinguishing among the different situations according to how the data became missing and incomplete.
- ▶ They are,
 1. MCAR: Missing completely at random.
 2. MAR: Missing at random.

Types of Missing Data

- ▶ Little and Rubin defined three unique types of missing data mechanism, distinguishing among the different situations according to how the data became missing and incomplete.
- ▶ They are,
 1. MCAR: Missing completely at random.
 2. MAR: Missing at random.
 3. MNAR: Missing not at random.

Missing Completely at Random

- ▶ **Missing completely at random (MCAR)** is defined as when the probability that the data are missing is not related to either the specific value which is supposed to be obtained or the set of observed responses, i.e MCAR situation assumes that the reasons why the data are missing is completely unrelated to the data (both covariates and response).

Missing Completely at Random

- ▶ **Missing completely at random (MCAR)** is defined as when the probability that the data are missing is not related to either the specific value which is supposed to be obtained or the set of observed responses, i.e MCAR situation assumes that the reasons why the data are missing is completely unrelated to the data (both covariates and response).
- ▶ MCAR is an ideal but unreasonable assumption for many studies performed in the field of anesthesiology.

Missing Completely at Random

- ▶ **Missing completely at random (MCAR)** is defined as when the probability that the data are missing is not related to either the specific value which is supposed to be obtained or the set of observed responses, i.e MCAR situation assumes that the reasons why the data are missing is completely unrelated to the data (both covariates and response).
- ▶ MCAR is an ideal but unreasonable assumption for many studies performed in the field of anesthesiology.
- ▶ However, if data are missing by design, because of an equipment failure or because the samples are lost in transit or technically unsatisfactory, such data are regarded as being MCAR.

Missing Completely at Random

- ▶ An example of MCAR is a weighing scale that ran out of batteries. Some of the data will be missing simply because of bad luck.

Missing Completely at Random

- ▶ An example of MCAR is a weighing scale that ran out of batteries. Some of the data will be missing simply because of bad luck.
- ▶ Another example is when we take a random sample of a population, where each member has the same chance of being included in the sample. The (unobserved) data of members in the population that were not included in the sample are MCAR.

Missing Completely at Random

- ▶ An example of MCAR is a weighing scale that ran out of batteries. Some of the data will be missing simply because of bad luck.
- ▶ Another example is when we take a random sample of a population, where each member has the same chance of being included in the sample. The (unobserved) data of members in the population that were not included in the sample are MCAR.
- ▶ MCAR causes enlarged standard errors due to the reduced sample size, but does not cause bias.

Missing Completely at Random

- ▶ An example of MCAR is a weighing scale that ran out of batteries. Some of the data will be missing simply because of bad luck.
- ▶ Another example is when we take a random sample of a population, where each member has the same chance of being included in the sample. The (unobserved) data of members in the population that were not included in the sample are MCAR.
- ▶ MCAR causes enlarged standard errors due to the reduced sample size, but does not cause bias.
- ▶ Thus, power may be lost in the design, but the estimated parameters are not biased by the absence of the data.

Missing at Random

- ▶ MAR means the propensity or tendency for a data point to be missing is not related to the missing data, but it is related to some of the observed data.

Missing at Random

- ▶ MAR means the propensity or tendency for a data point to be missing is not related to the missing data, but it is related to some of the observed data.
- ▶ That is when the missingness is not random, but where missingness can be fully accounted for by variables where there is complete information.

Missing at Random

- ▶ MAR means the propensity or tendency for a data point to be missing is not related to the missing data, but it is related to some of the observed data.
- ▶ That is when the missingness is not random, but where missingness can be fully accounted for by variables where there is complete information.
- ▶ MAR is a much broader class than MCAR.

Missing at Random

- ▶ MAR means the propensity or tendency for a data point to be missing is not related to the missing data, but it is related to some of the observed data.
- ▶ That is when the missingness is not random, but where missingness can be fully accounted for by variables where there is complete information.
- ▶ MAR is a much broader class than MCAR.
- ▶ For example, when we take a sample from a population, where the probability to be included depends upon some known property.

Missing at Random

- ▶ MAR means the propensity or tendency for a data point to be missing is not related to the missing data, but it is related to some of the observed data.
- ▶ That is when the missingness is not random, but where missingness can be fully accounted for by variables where there is complete information.
- ▶ MAR is a much broader class than MCAR.
- ▶ For example, when we take a sample from a population, where the probability to be included depends upon some known property.
- ▶ Another example of MAR is that males are less likely to fill in a depression survey but this has nothing to do with their level of depression, after accounting for maleness.

Missing at Random

- ▶ The practical problem with the MAR mechanism is that there is no way to confirm that the probability of missing data on Y is solely a function of other measured variables

Missing at Random

- ▶ The practical problem with the MAR mechanism is that there is no way to confirm that the probability of missing data on Y is solely a function of other measured variables
- ▶ For example, suppose that Hispanic children with poor reading skills have higher rates of missingness on the reading achievement test.

Missing at Random

- ▶ The practical problem with the MAR mechanism is that there is no way to confirm that the probability of missing data on Y is solely a function of other measured variables
- ▶ For example, suppose that Hispanic children with poor reading skills have higher rates of missingness on the reading achievement test.
- ▶ This situation is inconsistent with an MAR mechanism because there is a relationship between reading achievement and missingness, even after controlling for ethnicity

Missing at Random

- ▶ The practical problem with the MAR mechanism is that there is no way to confirm that the probability of missing data on Y is solely a function of other measured variables
- ▶ For example, suppose that Hispanic children with poor reading skills have higher rates of missingness on the reading achievement test.
- ▶ This situation is inconsistent with an MAR mechanism because there is a relationship between reading achievement and missingness, even after controlling for ethnicity
- ▶ However, the researcher would have no way of verifying the presence or absence of this relationship without knowing the values of the missing achievement scores.

Missing at Random

- ▶ The practical problem with the MAR mechanism is that there is no way to confirm that the probability of missing data on Y is solely a function of other measured variables
- ▶ For example, suppose that Hispanic children with poor reading skills have higher rates of missingness on the reading achievement test.
- ▶ This situation is inconsistent with an MAR mechanism because there is a relationship between reading achievement and missingness, even after controlling for ethnicity
- ▶ However, the researcher would have no way of verifying the presence or absence of this relationship without knowing the values of the missing achievement scores.
- ▶ Consequently, there is no way to test the MAR mechanism or to verify that scores are MAR.

Missing at Random

- ▶ The practical problem with the MAR mechanism is that there is no way to confirm that the probability of missing data on Y is solely a function of other measured variables
- ▶ For example, suppose that Hispanic children with poor reading skills have higher rates of missingness on the reading achievement test.
- ▶ This situation is inconsistent with an MAR mechanism because there is a relationship between reading achievement and missingness, even after controlling for ethnicity
- ▶ However, the researcher would have no way of verifying the presence or absence of this relationship without knowing the values of the missing achievement scores.
- ▶ Consequently, there is no way to test the MAR mechanism or to verify that scores are MAR.
- ▶ This represents an important practical problem for missing data analyses because maximum likelihood estimation and multiple imputation (the two techniques that methodologists currently recommend) assume an MAR mechanism

Missing Not at Random

- ▶ **Missing not at random (MNAR)** (also known as nonignorable nonresponse) is data that is neither MAR nor MCAR (i.e. the value of the variable that's missing is related to the reason it's missing).

Missing Not at Random

- ▶ **Missing not at random (MNAR)** (also known as nonignorable nonresponse) is data that is neither MAR nor MCAR (i.e. the value of the variable that's missing is related to the reason it's missing).
- ▶ To extend the previous example, this would occur if men failed to fill in a depression survey because of their level of depression.

Missing Not at Random

- ▶ **Missing not at random (MNAR)** (also known as nonignorable nonresponse) is data that is neither MAR nor MCAR (i.e. the value of the variable that's missing is related to the reason it's missing).
- ▶ To extend the previous example, this would occur if men failed to fill in a depression survey because of their level of depression.
- ▶ The cases of MNAR data are problematic.

Missing Not at Random

- ▶ **Missing not at random (MNAR)** (also known as nonignorable nonresponse) is data that is neither MAR nor MCAR (i.e. the value of the variable that's missing is related to the reason it's missing).
- ▶ To extend the previous example, this would occur if men failed to fill in a depression survey because of their level of depression.
- ▶ The cases of MNAR data are problematic.
- ▶ The only way to obtain an unbiased estimate of the parameters in such a case is to model the missing data.

Missing Not at Random

- ▶ **Missing not at random (MNAR)** (also known as nonignorable nonresponse) is data that is neither MAR nor MCAR (i.e. the value of the variable that's missing is related to the reason it's missing).
- ▶ To extend the previous example, this would occur if men failed to fill in a depression survey because of their level of depression.
- ▶ The cases of MNAR data are problematic.
- ▶ The only way to obtain an unbiased estimate of the parameters in such a case is to model the missing data.
- ▶ The model may then be incorporated into a more complex one for estimating the missing values.

General approaches to Missing Data

- ▶ Analysis techniques have to deal with incomplete /missing data as it is useful especially in case of experiments involving humans as data in these cases is expensive and significant.

General approaches to Missing Data

- ▶ Analysis techniques have to deal with incomplete /missing data as it is useful especially in case of experiments involving humans as data in these cases is expensive and significant.
- ▶ Three broad categories of methods for handling missing data are,

General approaches to Missing Data

- ▶ Analysis techniques have to deal with incomplete /missing data as it is useful especially in case of experiments involving humans as data in these cases is expensive and significant.
- ▶ Three broad categories of methods for handling missing data are,
 1. Deletion.

General approaches to Missing Data

- ▶ Analysis techniques have to deal with incomplete /missing data as it is useful especially in case of experiments involving humans as data in these cases is expensive and significant.
- ▶ Three broad categories of methods for handling missing data are,
 1. Deletion.
 2. Imputation.

General approaches to Missing Data

- ▶ Analysis techniques have to deal with incomplete /missing data as it is useful especially in case of experiments involving humans as data in these cases is expensive and significant.
- ▶ Three broad categories of methods for handling missing data are,
 1. Deletion.
 2. Imputation.
 3. Augmentation.

General approaches to Missing Data

- ▶ Analysis techniques have to deal with incomplete /missing data as it is useful especially in case of experiments involving humans as data in these cases is expensive and significant.
- ▶ Three broad categories of methods for handling missing data are,
 1. Deletion.
 2. Imputation.
 3. Augmentation.
- ▶ For any missing data methods to be effective, it is important for them to incorporate a specific missing data mechanism in its development.

Data deletion

- ▶ Data deletion methods such as list-wise and pair-wise deletion are efficient ways of dealing with missing data as long as missing data are MCAR.

1. Complete Case Analysis (Listwise Deletion)

- ▶ When some variables are not recorded for some of the units, a simple expedient, is simply to discard the data for any case that has one or more missing values and to analyze only the units with complete data.

1. Complete Case Analysis (Listwise Deletion)

- ▶ When some variables are not recorded for some of the units, a simple expedient, is simply to discard the data for any case that has one or more missing values and to analyze only the units with complete data.
- ▶ Example: Here we are deleting the incomplete units of data.

1. Complete Case Analysis (Listwise Deletion)

- ▶ When some variables are not recorded for some of the units, a simple expedient, is simply to discard the data for any case that has one or more missing values and to analyze only the units with complete data.
- ▶ Example: Here we are deleting the incomplete units of data.

Gender	8 th grade math test score	12 th grade math score
F	45	.
M	.	99
F	55	86
F	85	88
F	80	75
.	81	82
F	75	80
M	95	.
M	86	90
F	70	75
F	85	.

Advantages

- ▶ It is easy to implement and may be satisfactory with small amounts of missing .Restricting the analyses to the complete cases eliminates the need for specialized software and complex missing data handling techniques.

Advantages

- ▶ It is easy to implement and may be satisfactory with small amounts of missing .Restricting the analyses to the complete cases eliminates the need for specialized software and complex missing data handling techniques.
- ▶ listwise deletion can produce unbiased estimates of regression slopes under any missing data mechanism, provided that missingness is a function of a predictor variable and not the outcome variable (Little, 1992). This relatively esoteric scenario is the only situation in which listwise deletion is likely to outperform maximum likelihood estimation and multiple imputation with missing not at random (MNAR) data.

Disadvantages

- ▶ It can lead to serious biases, however, and it is not usually very efficient, especially when drawing inferences for subpopulations. because , it requires MCAR data and can produce distorted parameter estimates when this assumption does not hold.

Disadvantages

- ▶ It can lead to serious biases, however, and it is not usually very efficient, especially when drawing inferences for subpopulations. because , it requires MCAR data and can produce distorted parameter estimates when this assumption does not hold.
- ▶ listwise deletion is potentially very wasteful, particularly when the discarded cases have data on a large number of variables. Deleting the incomplete data records can produce a dramatic reduction in the total sample size.

Disadvantages

- ▶ It can lead to serious biases, however, and it is not usually very efficient, especially when drawing inferences for subpopulations. because , it requires MCAR data and can produce distorted parameter estimates when this assumption does not hold.
- ▶ listwise deletion is potentially very wasteful, particularly when the discarded cases have data on a large number of variables. Deleting the incomplete data records can produce a dramatic reduction in the total sample size.
- ▶ These shortcomings led to investigators to develop the following second approach .

2. Available Case Analysis (pairwise deletion)

- ▶ This uses all available data to estimate parameters of the model.

2. Available Case Analysis (pairwise deletion)

- ▶ This uses all available data to estimate parameters of the model.
- ▶ When a researcher looks at univariate descriptive statistics of a data set with missing observations, he or she is using available case analysis, examining the means and variances of the variables observed throughout the data set.

Advantages

- ▶ When the data are MCAR, when the remaining observations are representative of the originally identified data set shows that available case analysis provides consistent estimates when variables are moderately correlated in regression models.

Disadvantages

- ▶ The primary problem with pairwise deletion is that it requires MCAR data and can produce distorted parameter estimates when this assumption does not hold.

Disadvantages

- ▶ The primary problem with pairwise deletion is that it requires MCAR data and can produce distorted parameter estimates when this assumption does not hold.
- ▶ When calculating the correlation coefficient $r = \frac{\hat{\sigma}_{XY}}{\sqrt{\hat{\sigma}_X \cdot \hat{\sigma}_Y}}$, variable case analysis often might lead to implausible values of r , such as estimating correlations outside the range of -1.0 to 1.0 . Errors in estimation occur because of the differing number of observations used to estimate components of r .

Imputation

- ▶ Imputation is the process of replacing missing data with substituted values.

Imputation

- ▶ Imputation is the process of replacing missing data with substituted values.
- ▶ An advantage of the approach is that one can use standard statistical software to analyze the resulting completely filled data.

Imputation

- ▶ Imputation is the process of replacing missing data with substituted values.
- ▶ An advantage of the approach is that one can use standard statistical software to analyze the resulting completely filled data.
- ▶ Two popular method of imputation are **Single imputation** and **Multiple imputation**.

Imputation

- ▶ Imputation is the process of replacing missing data with substituted values.
- ▶ An advantage of the approach is that one can use standard statistical software to analyze the resulting completely filled data.
- ▶ Two popular method of imputation are **Single imputation** and **Multiple imputation**.
- ▶ We may be familiar with filling missing values with the sample mean value (mean imputation).

Imputation

- ▶ Imputation is the process of replacing missing data with substituted values.
- ▶ An advantage of the approach is that one can use standard statistical software to analyze the resulting completely filled data.
- ▶ Two popular method of imputation are **Single imputation** and **Multiple imputation**.
- ▶ We may be familiar with filling missing values with the sample mean value (mean imputation).
- ▶ Alternatively, missing data imputation can be slightly more sophisticated, using regression predictions to fill in missing cases (regression imputation).

Imputation

- ▶ Imputation is the process of replacing missing data with substituted values.
- ▶ An advantage of the approach is that one can use standard statistical software to analyze the resulting completely filled data.
- ▶ Two popular method of imputation are **Single imputation** and **Multiple imputation**.
- ▶ We may be familiar with filling missing values with the sample mean value (mean imputation).
- ▶ Alternatively, missing data imputation can be slightly more sophisticated, using regression predictions to fill in missing cases (regression imputation).
- ▶ These methods, known as single imputation techniques.

Imputation

- ▶ Here the imputed value is treated as the true value, ignoring the fact that no imputation method can give the exact value.

Imputation

- ▶ Here the imputed value is treated as the true value, ignoring the fact that no imputation method can give the exact value.
- ▶ Single imputation does not reflect the uncertainty about the prediction of the missing values.

Imputation

- ▶ Here the imputed value is treated as the true value, ignoring the fact that no imputation method can give the exact value.
- ▶ Single imputation does not reflect the uncertainty about the prediction of the missing values.
- ▶ Here we use multiple imputation.

Imputation

- ▶ Here the imputed value is treated as the true value, ignoring the fact that no imputation method can give the exact value.
- ▶ Single imputation does not reflect the uncertainty about the prediction of the missing values.
- ▶ Here we use multiple imputation.
- ▶ Multiple imputation uses a predicted value for a given subject and time point using statistical modeling of available data.

Imputation

- ▶ Here the imputed value is treated as the true value, ignoring the fact that no imputation method can give the exact value.
- ▶ Single imputation does not reflect the uncertainty about the prediction of the missing values.
- ▶ Here we use multiple imputation.
- ▶ Multiple imputation uses a predicted value for a given subject and time point using statistical modeling of available data.
- ▶ The processes and results of data augmentation are quite similar to those of MI. But they are complex in nature.

Use of Maximum likelihood

- ▶ Maximum likelihood also plays a central role in missing data analysis.

Use of Maximum likelihood

- ▶ Maximum likelihood also plays a central role in missing data analysis.
- ▶ The main aim is to analyze the full, incomplete data set using maximum likelihood estimation.

Use of Maximum likelihood

- ▶ Maximum likelihood also plays a central role in missing data analysis.
- ▶ The main aim is to analyze the full, incomplete data set using maximum likelihood estimation.
- ▶ This method does not impute any data, but rather uses each cases available data to compute maximum likelihood estimates.

Use of Maximum likelihood

- ▶ Maximum likelihood also plays a central role in missing data analysis.
- ▶ The main aim is to analyze the full, incomplete data set using maximum likelihood estimation.
- ▶ This method does not impute any data, but rather uses each cases available data to compute maximum likelihood estimates.
- ▶ The maximum likelihood estimate of a parameter is the value of the parameter that is most likely to have resulted in the observed data.

Use of Maximum likelihood

- ▶ Maximum likelihood also plays a central role in missing data analysis.
- ▶ The main aim is to analyze the full, incomplete data set using maximum likelihood estimation.
- ▶ This method does not impute any data, but rather uses each cases available data to compute maximum likelihood estimates.
- ▶ The maximum likelihood estimate of a parameter is the value of the parameter that is most likely to have resulted in the observed data.
- ▶ The starting point for a maximum likelihood analysis is to specify a distribution for the population data.

Use of Maximum likelihood

- ▶ Maximum likelihood also plays a central role in missing data analysis.
- ▶ The main aim is to analyze the full, incomplete data set using maximum likelihood estimation.
- ▶ This method does not impute any data, but rather uses each cases available data to compute maximum likelihood estimates.
- ▶ The maximum likelihood estimate of a parameter is the value of the parameter that is most likely to have resulted in the observed data.
- ▶ The starting point for a maximum likelihood analysis is to specify a distribution for the population data.
- ▶ Suppose we have n independent observations ($i = 1, \dots, n$) on k variables ($y_{i1}, y_{i2}, \dots, y_{ik}$) and no missing data. Then the likelihood function is
$$L = \prod_{i=1}^n f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta)$$

Use of Maximum Likelihood

- Now if there are m observations with complete data and $n - m$ observations with data missing in y_1 and y_2 , the likelihood function for the full data set becomes

$$L = \prod_{i=1}^m f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) \prod_{i=m+1}^n f_i^*(y_{i3}, \dots, y_{ik}; \theta) \text{ where}$$
$$f_i^*(y_{i3}, \dots, y_{ik}; \theta) = \sum_{y_1} \sum_{y_2} \text{ or } \int \int f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) dy_1 dy_2$$

$y_1 \text{ and } y_2$

where the observations are ordered such that the first m have no missing data and the last $n - m$ have missing data.

Use of Maximum Likelihood

- Now if there are m observations with complete data and $n - m$ observations with data missing in y_1 and y_2 , the likelihood function for the full data set becomes

$$L = \prod_{i=1}^m f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) \prod_{i=m+1}^n f_i^*(y_{i3}, \dots, y_{ik}; \theta) \text{ where}$$
$$f_i^*(y_{i3}, \dots, y_{ik}; \theta) = \sum_{y_1} \sum_{y_2} \text{ or } \int \int f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) dy_1 dy_2$$

$y_1 \text{ and } y_2$

where the observations are ordered such that the first m have no missing data and the last $n - m$ have missing data.

- This likelihood can then be maximized to get ML estimates of θ .

Use of Maximum Likelihood

- ▶ Now if there are m observations with complete data and $n - m$ observations with data missing in y_1 and y_2 , the likelihood function for the full data set becomes

$$L = \prod_{i=1}^m f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) \prod_{i=m+1}^n f_i^*(y_{i3}, \dots, y_{ik}; \theta) \text{ where}$$
$$f_i^*(y_{i3}, \dots, y_{ik}; \theta) = \sum_{y_1} \sum_{y_2} \text{ or } \int \int f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) dy_1 dy_2$$

$y_1 \text{ and } y_2$

where the observations are ordered such that the first m have no missing data and the last $n - m$ have missing data.

- ▶ This likelihood can then be maximized to get ML estimates of θ .
- ▶ Researchers in the social and the behavioral sciences routinely assume that their variables are normally distributed in the population.

Use of Maximum Likelihood

- ▶ Now if there are m observations with complete data and $n - m$ observations with data missing in y_1 and y_2 , the likelihood function for the full data set becomes

$$L = \prod_{i=1}^m f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) \prod_{i=m+1}^n f_i^*(y_{i3}, \dots, y_{ik}; \theta) \text{ where}$$
$$f_i^*(y_{i3}, \dots, y_{ik}; \theta) = \sum_{y_1} \sum_{y_2} \text{ or } \int \int f_i(y_{i1}, y_{i2}, \dots, y_{ik}; \theta) dy_1 dy_2$$

$y_1 \text{ and } y_2$

where the observations are ordered such that the first m have no missing data and the last $n - m$ have missing data.

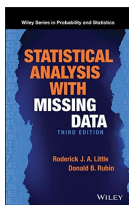
- ▶ This likelihood can then be maximized to get ML estimates of θ .
- ▶ Researchers in the social and the behavioral sciences routinely assume that their variables are normally distributed in the population.
- ▶ Maximum likelihood estimators is generally most powerful when the number of complete pairs of observations is moderately large (larger than 20) and the correlation is in the range of 0.3–0.9.

References

- ▶ Statistical Analysis With Missing Data, Third Edition, Roderick J.A. Little, Donald B. Rubin.

References

- ▶ Statistical Analysis With Missing Data, Third Edition, Roderick J.A. Little, Donald B. Rubin.

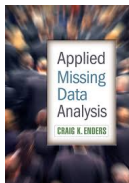


- ▶ Missing Values in Repeated Measurement Designs, Keumhee Cho Carriere, Rong Huang, Yuanyuan Liang, Xiaoming Sheng.



References

- ▶ Applied Missing Data Analysis, Craig K. Enders



Acknowledgement

- ▶ We would like to express our special thanks of gratitude to our respected **Professor Arunangshu Biswas** who gave us the golden opportunity to do this wonderful presentation on the topic **Missing Values in Repeated Measurement Designs**, which also helped us in doing a lot of Research and we came to know about so many new things. We are really thankful to him.