



1.3.4: Missing Data and Sampling Weights (Asynchronous-Online)

COMING SUMMER 2025

Module “1.3.4: Missing Data and Sampling Weights” will be posted prior to the In-Person Workshops in Summer 2025.

Session Objectives

1. Identify and summarize missing data.
2. Learn methods to handle missing data according to variable type.
3. Use a survey sampling weight to generate more representative descriptive and inferential statistical values (brief intro)
4. Discuss potential bias when removing missing observations without careful examination.

Key points:

1. R packages that support missing data examination
2. Mean/median imputation for continuous variables
3. What to do with missing observations for categorical variables
4. Ways to examine potential differences between complete and missing observations in association between certain independent and dependent variables
 - What to do if such association significantly differs between complete and missing observations
5. R packages for complex survey data (e.g., survey package)
 - R codes to generate weighted descriptive statistics and contingency tables, as well as to develop weighted linear models

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.



- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and Teun van den Brand. 2024. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.