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

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, JooYoung Seo, Ken Brevoort, and Olivier Roy. 2024. *Gt: Easily Create Presentation-Ready Display Tables*. https://gt.rstudio.com.

Kassambara, Alboukadel. 2023. *Ggpubr: Ggplot2 Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/.

Meyer, David, Achim Zeileis, and Kurt Hornik. 2006. "The Strucplot Framework: Visualizing Multi-Way Contingency Tables with Vcd." *Journal of Statistical Software* 17 (3): 1–48. https://doi.org/10.18637/jss.v017.i03.

Meyer, David, Achim Zeileis, Kurt Hornik, and Michael Friendly. 2023. *Vcd: Visualizing Categorical Data*. https://CRAN.R-project.org/package=vcd.

Mock, Thomas. 2024. *gtExtras: Extending Gt for Beautiful HTML Tables*. https://github.com/jthomasmock/gtExtras.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. https://patchwork.data-imaginist.com.

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

Schloerke, Barret, Di Cook, Joseph Larmarange, Francois Briatte, Moritz Marbach, Edwin Thoen, Amos Elberg, and Jason Crowley. 2024. *GGally: Extension to Ggplot2*. https://ggobi.github.io/ggally/.

Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with r, Plotly, and Shiny*. Chapman; Hall/CRC. https://plotly-r.com.

Sievert, Carson, Chris Parmer, Toby Hocking, Scott Chamberlain, Karthik Ram, Marianne Corvellec, and Pedro Despouy. 2024. *Plotly: Create Interactive Web Graphics via Plotly.js*. https://plotly-r.com.

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.

Zeileis, Achim, David Meyer, and Kurt Hornik. 2007. "Residual-Based Shadings for Visualizing (Conditional) Independence." *Journal of Computational and Graphical Statistics* 16 (3): 507–25. https://doi.org/10.1198/106186007X237856.