

RESSALVA

Atendendo solicitação da autora, o texto completo desta dissertação será disponibilizado somente a partir de 17/06/2026.

**PROGRAMA DE
PÓS-GRADUAÇÃO EM
MATEMÁTICA**

**Aplicações do Método de Imputação Múltipla
MICE em Dados Longitudinais**

Thais Vieira dos Santos

INSTITUTO DE GEOCIÊNCIAS E CIÊNCIAS EXATAS

RIO CLARO
2026

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Thais Vieira dos Santos

Aplicações do Método de Imputação Múltipla MICE em Dados Longitudinais

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Mestre em Matemática.

Orientadora
Profa. Dra. Selene Maria Coelho Loibel

Rio Claro/SP
2026

D724a dos Santos, Thais Vieira
Aplicações do Método de Imputação Múltipla MICE em Dados Longitudinais / Thais Vieira dos Santos. -- Rio Claro, 2026
80 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Instituto de Geociências e Ciências Exatas, Rio Claro
Orientadora: Selene Maria Coelho Loibel

1. Multiple imputation (Statistics). 2. Estudos longitudinais. 3. Ausencia de dados (Estatística). I. Título.

Impacto potencial desta pesquisa

Este trabalho apresenta aplicações em estudos longitudinais da Imputação Múltipla com o método intitulado *Multivariate Imputation by Chained Equations* (MICE). Na atual literatura há uma demanda dessas aplicações. Assim, esse estudo pode contribuir uma vez que apresenta soluções para o problema de dados longitudinais incompletos.

Potential impact of this research

This essay presents applications in longitudinal studies of Multiple Imputation with the method entitled *Multivariate Imputation by Chained Equations* (MICE). In the current literature, there is a demand for these applications. Furthermore, this work can contribute since it presents solutions to the problem of incomplete longitudinal data.

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Thais Vieira dos Santos

APLICAÇÕES DO MÉTODO DE IMPUTAÇÃO MÚLTIPLA MICE EM
DADOS LONGITUDINAIS

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista “Júlio de Mesquita Filho”, como parte dos requisitos para obtenção do título de Mestre em Matemática.

Comissão Examinadora

Prof. Dra. Selene Maria Coelho Loibel
IGCE/UNESP/Rio Claro (SP)

Prof. Dr. José Silvio Govone
IGCE/UNESP/Rio Claro (SP)

Prof. Dr. Marcos Tadeu dos Santos
SEDUC/Rio Claro (SP)

Conceito: Aprovado

Rio Claro (SP), 17 de dezembro de 2025

À minha família.

Agradecimentos

Durante a graduação, a possibilidade de me tornar mestre parecia distante. Vivia um dia de cada vez, esforçando-me para superar as dificuldades que me faziam desacreditar do meu potencial matemático. Encontrei o equilíbrio ao me realizar na educação básica e descobrir a plenitude no ato de ensinar. Com esse propósito, iniciei o mestrado e hoje realizo o sonho de contribuir com a formação de estudantes do ensino básico.

Para a conquista deste título, contei com o amparo divino e o apoio fundamental de pessoas a quem dedico minha profunda gratidão.

Agradeço a Deus, que me concedeu forças para seguir e iluminou minhas escolhas até aqui.

Aos meus pais, Almir e Maria, por todo o apoio, cuidado, paciência e dedicação. Jamais esquecerei as noites e madrugadas em que meu pai me buscava na rodoviária, nem do suporte essencial que ambos me deram para que eu pudesse conciliar o trabalho, os estudos para concursos e o mestrado. Sem o apoio deles, eu certamente não teria suportado uma rotina tão exaustiva.

À minha irmã Nathalia, pelo apoio constante, pela torcida e por ouvir meus desabafos nos momentos difíceis. À minha irmã Cintia, pela torcida sempre presente. À minha avó Lindalva, que mesmo à distância me apoia com suas rezas.

À minha orientadora, Profa. Dra. Selene Maria Coelho Loibel, expressei minha profunda gratidão pela orientação, pela paciência e pelo incentivo constante ao longo de todo este processo. Seus ensinamentos foram fundamentais para a realização deste trabalho e para o meu amadurecimento acadêmico.

Ao PGMAT (Programa de Pós-Graduação em Matemática) e seus docentes, pela excelência no ensino. Agradeço, em especial, à Profa. Dra. Carina, pela paciência e compreensão.

Aos professores da minha banca examinadora: à Profa. Marta, pelas orientações fundamentais no exame de qualificação; ao Prof. Dr. Silvio, por suas valiosas contribuições tanto na qualificação quanto na defesa deste trabalho; e ao Prof. Dr. Marcos, por sua participação e contribuições essenciais na banca de defesa.

Às gestões e colegas das escolas onde atuei e atuo – E.E. Profa. Julieta Farão, E.E. Prof. Arthur Chagas Junior, CEU EMEF Tatiana Belinky, CEU EMEI Profa. Erika de Souza Brito Matos e CIEJA Rolando Boldrin – pela compreensão e apoio à minha jornada dupla entre o magistério e a pesquisa.

À gestão de 2022 da Moradia Estudantil da Unesp de Rio Claro. Agradeço a existência desse espaço, essencial para minha vida acadêmica, onde conheci pessoas valiosas que levo no coração: Nathalia, Valeria e Wilson. Agradeço também aos moradores da casa 12 por toda a acolhida e pela flexibilização do cotidiano diante da minha ida semanal para a realização das disciplinas.

Expresso minha gratidão às minhas amigas Carolina, Mayara, Luciana e Daniele, pela torcida e pela compreensão nos períodos em que estive ausente. Um agradecimento especial à Viviane, por todas as conversas e pelos pertinentes conselhos compartilhados.

Aos colegas Raul e Gabriela, pela contribuição direta neste mestrado.

Por fim, agradeço aos meus colegas de mestrado pelo apoio mútuo, estendendo um agradecimento especial à Renata e à Sara, pela parceria e auxílio.

A vida se mostra a cada passo.
Saikawa Roshi

Resumo

Dados longitudinais estão presentes nos mais variados estudos, sendo comum a presença de dados ausentes nesse contexto. O principal objetivo deste trabalho é mostrar as vantagens do uso da Imputação Múltipla de dados ausentes, com o método denominado *Multivariate Imputation by Chained Equations* (MICE). Duas aplicações com dados longitudinais reais foram realizadas com esse fim. O primeiro estudo contou com um conjunto de dados completos sobre a qualidade da água do Rio Piracicaba e, com a retirada aleatória de uma parte dos dados, foram aplicados os diversos métodos de imputação usuais e a imputação múltipla. O conjunto de dados original e os conjuntos imputados com todos os métodos foram analisados e os resultados comparados. No segundo estudo, sobre parâmetros hematológicos e bioquímicos em serpentes da espécie *Crotalus durissus terrificus*, somente a imputação múltipla foi aplicada para obtenção dos perfis médios das variáveis.

Palavras-chave: Imputação Múltipla. Dados longitudinais. Dados ausentes.

Abstract

Longitudinal data are featured in the most varied studies, and the presence of missing data in this context is common. The main objective of this dissertation is to present the advantages of using Multiple Imputation of missing data, with the method named Multivariate Imputation by Chained Equations (MICE). Two applications with real longitudinal data were made for this purpose. The first study had a complete set of data on the water quality of the Piracicaba River and, with the random removal of a part of the data, the various usual imputation methods and multiple imputation were applied. The original dataset and imputed sets with all methods were analyzed, and the results compared. In the second study, on hematological and biochemical parameters in snakes of the species *Crotalus durissus terrificus*, only multiple imputation was applied to obtain the average profiles of the variables.

Keywords: Multiple Imputation. Longitudinal data. Missing data.

Lista de Figuras

4.1	Padrões de ausência de dados	25
5.1	Imputação pela média	31
5.2	Imputação por regressão	32
5.3	Imputação por regressão estocástica	33
5.4	Imputação múltipla - Primeiro conjunto completo	34
5.5	Dispersão dos dados observados e de cada um dos 5 conjuntos de dados imputados	35
6.1	Esquema de Imputação Múltipla	37
7.1	Padrão de ausência dos dados	46
7.2	Imputação por média - Todas variáveis	47
7.3	Condutividade e Sólidos	47
7.4	Regressão com Imputação por média	49
7.5	Imputação múltipla - Todas variáveis	51
7.6	Perfis médios - Contagens de hemácias em indivíduos tratados	56
7.7	Perfis médios - Contagens de leucócitos em indivíduos tratados	57
7.8	Perfis médios - Contagens de trombócitos em indivíduos tratados	58
7.9	Perfis médios - Concentrações de hemoglobina em indivíduos tratados	59
7.10	Perfis médios - Concentrações de glicose em indivíduos tratados	60
7.11	Perfis médios - Concentrações de fósforo em indivíduos tratados	61
7.12	Perfis médios - Concentrações de ácido úrico (AU) em indivíduos tratados	62
7.13	Perfis médios - Concentrações de colesterol em indivíduos tratados	63
7.14	Perfis médios - Contagens de hemácias em indivíduos não tratados	64
7.15	Perfis médios - Contagens de leucócitos em indivíduos não tratados	65
7.16	Perfis médios - Contagens de trombócitos em indivíduos não tratados	66
7.17	Perfis médios - Concentrações de hemoglobina em indivíduos não tratados	67
7.18	Perfis médios - Concentrações de glicose em indivíduos não tratados	68
7.19	Perfis médios - Concentrações de fósforo em indivíduos não tratados	69
7.20	Perfis médios - Concentrações de ácido úrico (AU) em indivíduos não tratados	70
7.21	Perfis médios - Concentrações de colesterol em indivíduos não tratados	71

Lista de Tabelas

3.1	Estrutura para dados longitudinais	22
3.2	Tabela no formato <i>wide</i>	23
3.3	Tabela no formato <i>long</i>	23
5.1	Dados de qualidade do ar	28
7.1	Dados de qualidade da Água - Rio Piracicaba	44
7.2	Dados de qualidade da Água - Rio Piracicaba	45
7.3	Estatísticas com o conjunto de dados completo	45
7.4	Regressão com o conjunto de dados completo	46
7.5	Estatísticas com a imputação por média	48
7.6	Regressão com a imputação por média	48
7.7	Estatísticas com <i>Listwise</i>	50
7.8	Regressão com <i>Listwise</i>	50
7.9	Estatísticas com <i>Pairwise</i>	50
7.10	Regressão com <i>Pairwise</i>	50
7.11	Regressão com <i>IM</i>	51
7.12	Intervalos de Confiança e Amplitudes - Dados Completos	52
7.13	Intervalos de Confiança e Amplitudes - Dados Imputados pela média	52
7.14	Intervalos de Confiança e Amplitudes - Dados Imputados por <i>Listwise</i>	52
7.15	Intervalos de Confiança e Amplitudes - Dados Imputados por <i>Pairwise</i>	52
7.16	Intervalos de Confiança e Amplitudes - Dados Imputados por IM	53
7.17	Contagens de hemácias em indivíduos tratados	56
7.18	Contagens de leucócitos em indivíduos tratados	57
7.19	Contagens de trombócitos em indivíduos tratados	58
7.20	Concentrações de hemoglobina em indivíduos tratados	59
7.21	Concentrações de glicose em indivíduos tratados	60
7.22	Concentrações de fósforo em indivíduos tratados	61
7.23	Concentrações de ácido úrico (AU) em indivíduos tratados	62
7.24	Concentrações de colesterol em indivíduos tratados	63
7.25	Contagens de hemácias em indivíduos não tratados	64
7.26	Contagens de leucócitos em indivíduos não tratados	65
7.27	Contagens de trombócitos em indivíduos não tratados	66
7.28	Concentrações de hemoglobina em indivíduos não tratados	67
7.29	Concentrações de glicose em indivíduos não tratados	68
7.30	Concentrações de fósforo em indivíduos não tratados	69
7.31	Concentrações de ácido úrico (AU) em indivíduos não tratados	70

7.32	Concentrações de colesterol em indivíduos não tratados	71
A.1	Dados de qualidade de ar	77

Sumário

1	Introdução	16
2	Conceitos Fundamentais de Inferência Estatística	18
2.1	Inferência Estatística	18
2.1.1	Inferência Clássica	18
2.1.2	Inferência Bayesiana	18
3	Caracterização dos Estudos Longitudinais	21
3.1	Conceitualização	21
3.1.1	Exemplo de um conjunto de dados longitudinais com covariáveis . .	21
3.2	Formatos para organização de dados longitudinais	22
4	Dados Ausentes	24
4.1	Padrão e quantificação de dados ausentes	24
4.2	Mecanismos que geram dados ausentes	25
5	Imputação de Dados Ausentes	27
5.1	Histórico das ideias para imputação	27
5.2	Métodos usuais de imputação para valores ausentes	27
5.2.1	<i>Listwise deletion</i>	28
5.2.2	<i>Pairwise deletion</i>	29
5.2.3	Imputação pela média	30
5.2.4	Imputação por regressão	31
5.2.5	Imputação por regressão estocástica	32
6	Imputação Múltipla (IM)	36
6.1	Histórico dos métodos de Imputação Múltipla	36
6.2	Imputação Múltipla	36
6.2.1	<i>MICE</i>	39
6.2.2	<i>Predictive Mean Matching - PMM</i>	41
7	Aplicações com dados reais	42
7.1	Comparação de métodos de imputação - Aplicação para dados de qualidade de água doce (Rio Piracicaba - SP)	42
7.1.1	Resultados	44
7.2	Aplicação para dados de serpentes	54
7.2.1	Resultados	55
7.2.2	Perfis médios do Grupo 1 - Indivíduos tratados com vermífugo . . .	56

7.2.3	Perfis médios do Grupo 2 - Indivíduos não tratados com vermífugo	64
8	Conclusões	72
	Referências	73
A	Apêndice A	76
A.1	Dados de qualidade do ar	77

1 Introdução

Em qualquer experimento ou estudo que envolve uma análise estatística é importante conhecer e verificar quais são os tipos de dados que o estudo envolve. Há, então, dados denominados como longitudinais que estão presentes em diversas áreas e com isso, faz-se necessário considerar suas especificidades.

Dados longitudinais podem ser definidos como aqueles que são coletados para acompanhar as mudanças de uma mesma variável ao longo do tempo e tais mudanças podem ser comparadas. Na última década, diversos trabalhos analisaram dados longitudinais, principalmente em contextos das áreas biológicas, como por exemplo o de Schober e Vetter (2018) que apresenta três abordagens para dados longitudinais no contexto de pesquisa médica.

Segundo Cao *et al.* (2022) e Rosato *et al.* (2021) a falta de dados é comum em estudos longitudinais e a presença de dados ausentes pode afetar a validade dos resultados e o poder do estudo. Essa não-resposta se faz presente por diversos fatores, sendo um deles a possibilidade, por exemplo, de informações perdidas ao longo das entrevistas da pesquisa. Outra possibilidade referente a não-resposta acontece quando os participantes da pesquisa se recusam a responder algumas questões.

De acordo com Huque *et al.* (2020) para que não haja resultados viciados, é necessário que métodos de imputação sejam adequados para os dados longitudinais. Nesse contexto, a Imputação Múltipla (IM) surge como uma opção viável e é definida como a geração de múltiplas cópias de conjuntos de dados completos, onde os valores ausentes são substituídos por valores imputados amostrados de sua distribuição preditiva a posteriori, considerando os dados observados.

Muitos trabalhos utilizam esse método como por exemplo, o artigo de Huque *et al.* (2020) no qual foram aplicados sete métodos de Imputação Múltipla (IM) para dados ausentes em configurações longitudinais. Nesse artigo, após a imputação com os diversos métodos, o modelo linear misto foi utilizado na análise final.

Outro trabalho que apresenta esse mesmo contexto é o de Cao *et al.* (2022) que compara quatro métodos de imputação múltipla *Fully Conditional Specification* (FCS). O método FCS gera imputações, iterando densidades condicionais de tal modo que utiliza um modelo de regressão univariado apropriado para cada variável com valor ausente.

O objetivo principal desse trabalho foi o estudo dos métodos de imputação frequentemente utilizados e a verificação das vantagens em se utilizar o método de imputação FCS na maioria das aplicações encontradas na literatura.

O objetivo secundário foi aplicar os diversos métodos de imputação para dois conjuntos de dados reais: dados de qualidade da água doce do Rio Piracicaba e dados de parâmetros hematológicos e bioquímicos em serpentes da espécie *Crotalus durissus terrificus*. A partir dessas aplicações foi possível comparar os métodos.

Neste trabalho de dissertação foi aplicado o método *Fully Conditional Specification* (FCS) em dois tipos de formatos de dados longitudinais: *wide* e *long*. dados longitudinais. A saber: no capítulo 2 foram apresentados os conceitos fundamentais de estatística, o capítulo 3 contém a caracterização dos estudos longitudinais com exemplificação e explanação dos formatos possíveis em que dados longitudinais podem ser organizados, o capítulo 4 apresentou a definição de dados ausentes, com o padrões e mecanismos de perda de dados, no capítulo 5 foi explicitada a imputação de dados ausentes com exemplos de métodos mais simplificados, no capítulo 6 foi apresentada a conceitualização de imputação múltipla e suas vantagens, no capítulo 7 duas aplicações em dados longitudinais foram apresentadas e por fim, seguiu o capítulo de conclusões.

8 Conclusões

Neste trabalho foram apresentados os diversos métodos de imputação de dados ausentes, desde os mais simples até o método de imputação múltipla. Foram aplicados os métodos a três conjuntos de dados. A primeira aplicação foi inteiramente reproduzida com todos os cálculos refeitos no software R, baseando-se nos dados de qualidade do ar de um exemplo do livro “*Flexible Imputation of Missing Data*”. Esse exemplo mostra os problemas quando se utiliza os métodos mais simples: ocorre a falsa diminuição ou aumento da variabilidade total, há a imputação de valores fora do suporte das variáveis, há vícios nos estimadores, entre outras desvantagens.

Na segunda aplicação, sobre qualidade da água do Rio Piracicaba, um conjunto de dados completo foi utilizado para a comparação dos métodos de imputação. Foram retirados aleatoriamente alguns valores em cada variável e em seguida esses valores foram imputados pelos diversos métodos. Um modelo de regressão foi ajustado para todos os conjuntos de dados: o original (completo) e os imputados. Os resultados foram comparados e verificou-se que o método de imputação múltipla foi o que forneceu resultados mais próximos dos resultados com os dados originais (completos).

Para a terceira aplicação, com dados longitudinais sobre parâmetros hematológicos e bioquímicos em serpentes, ocorreu somente a imputação múltipla sendo aplicada com o algoritmo MICE. Os dados originais apresentavam vários valores ausentes tornando a análise de perfis inviável uma vez que o plano dos pesquisadores seria coletar dados de apenas 20 serpentes, divididas em dois grupos, em 7 ocasiões. Na realidade, vários dados foram perdidos. Percebe-se que há variabilidade grande entre as medidas dos indivíduos, e entre coletas considerando os dados originais. Após a imputação, com cinco conjuntos de dados imputados, vemos que as médias dos valores imputados estão de acordo com os intervalos de valores observados e não há aumento nem diminuição da variabilidade dos dados. Considerando os resultados das três aplicações pode-se concluir que o método de imputação múltipla se mostrou eficaz e com boas propriedades, possibilitando a análise estatística posterior confiável. Esse método é de simples implementação no software R e deve ser considerado quando a ausência de dados não pode ser desprezada.

Referências

- [1] ALLAN, F. E.; WISHART, J. A method of estimating the yield of a missing plot in field experimental work. *The Journal of Agricultural Science*, [S. l.], v. 20, n. 3, p. 399–406, 1930.
- [2] AUSTIN, P. C. et al. Missing data in clinical research: a tutorial on multiple imputation. *Canadian Journal of Cardiology*, [S. l.], v. 37, n. 9, p. 1322–1331, 2021.
- [3] BROWN, M.; KROS, J. F. Data mining and the impact of missing data. *Industrial Management & Data Systems*, [S. l.], v. 103, n. 8, p. 611–621, 2003.
- [4] CAO, Y. et al. Review and evaluation of imputation methods for multivariate longitudinal data with mixed-type incomplete variables. *Statistics in Medicine*, [S. l.], v. 41, n. 30, p. 5844–5876, 2022.
- [5] CASELLA, G.; GEORGE, E. I. Explaining the gibbs sampler. *The American Statistician*, [S. l.], v. 46, n. 3, p. 167–174, 1992.
- [6] United States Bureau of the Census. *Census of Manufactures 1958*. [S.l.]: Washington, D.C., 1961.
- [7] CETESB. *Relatórios de Qualidade das Águas Interiores do Estado de São Paulo*. São Paulo: [s.n.], 2025. Disponível em: <<https://cetesb.sp.gov.br/aguas-interiores/publicacoes-e-relatorios/>>. Acesso em: 16 abr. 2025.
- [8] COSTA, H. C.; GUEDES, T. B.; BERNILS, R. S. Lista de répteis do brasil: padrões e tendências. *Herpetologia Brasileira*, [S. l.], p. 110–279, 2021.
- [9] CUNHA, L. E. R. Soros antiofídicos: história, evolução e futuro. *Journal Health NPEPS*, [S. l.], v. 2, n. 1, p. 1–4, 2017.
- [10] GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. *Markov Chain Monte Carlo in Practice*. London: Chapman-Hall, 1996.
- [11] GRAHAM, J. W.; HOFER, S. M.; PICCININ, A. M. K. Analysis with missing data in prevention research. In: BRYANT, M.; WINDLE, M.; WEST, S. (Ed.). *The science of prevention: methodological advances from alcohol and substance abuse research*. Washington, D.C.: American Psychological Association, 1997. p. 325–366.
- [12] GRAHAM, J. W.; HOFER, S. M. Multiple imputation in multivariate research. In: LITTLE, T. D.; SCHNABEL, K. U.; BAUMERT, J. (Ed.). *Modeling longitudinal and multilevel data*. [S. l.]: Psychology Press, 2000. p. 189–204.

-
- [13] HUQUE, M. H. et al. Multiple imputation methods for handling incomplete longitudinal and clustered data where the target analysis is a linear mixed effects model. *Biometrical Journal*, [S. l.], v. 62, n. 2, p. 444–466, 2020.
- [14] LACHIN, J. M. Fallacies of last observation carried forward analyses. *Clinical Trials*, London, v. 13, n. 2, p. 161–168, 2016.
- [15] LITTLE, R. J. A.; RUBIN, D. B. *Statistical analysis with missing data*. New York: John Wiley & Sons, 1987.
- [16] LITTLE, R. J. A. Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, [S. l.], v. 6, n. 3, p. 287–296, 1988.
- [17] LITTLE, R. J. A. Regression with missing x's: A review. *Journal of the American Statistical Association*, [S. l.], v. 87, n. 420, p. 1227–1237, 1992.
- [18] LITTLE, R. J. A.; SCHENKER, N. Missing data. In: ARMINGER, G.; CLOGG, C. C.; SOBEL, D. B. (Ed.). *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Boston, MA: Springer US, 1995. p. 39–75.
- [19] LITTLE, R. J. A.; RUBIN, D. B. *Statistical analysis with missing data*. 2. ed. New York: John Wiley & Sons, 2002.
- [20] LOBATO, F. M. F. *Estratégias evolucionárias para otimização no tratamento de dados ausentes por imputação múltipla de dados*. Tese (Tese (Doutorado)) — Universidade Federal do Pará, Belém, 2016.
- [21] MACK, C.; SU, Z.; WEITRICH, D. *Managing missing data in patient registries: addendum to registries for evaluating patient outcomes: a user's guide*. [S. l.]: [s. n.], 2018.
- [22] PEPINSKY, T. B. A note on listwise deletion versus multiple imputation. *Political Analysis*, [S. l.], v. 26, n. 4, p. 480–488, 2018.
- [23] ROBERTO, I. J. *Répteis (Testudines, Squamata, Crocodylia) da Reserva Biológica de Pedra Talhada*. [S. l.]: Melgarejo, 2015.
- [24] ROSATO, R. et al. Missing data in longitudinal studies: comparison of multiple imputation methods in a real clinical setting. *Journal of Evaluation in Clinical Practice*, [S. l.], v. 27, n. 1, p. 34–41, 2021.
- [25] R Core Team. *The R Project for Statistical Computing*. [S. l.], 2026. Disponível em: <<https://www.r-project.org/>>. Acesso em: 1 mar. 2026.
- [26] RUBIN, D. B. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *Journal of the American Statistical Association*, [S. l.], v. 72, n. 359, p. 538–543, 1977.
- [27] RUBIN, D. B. Adjustments in large using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics*, [S. l.], v. 4, n. 1, p. 87–94, 1986.
- [28] RUBIN, D. B. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, 1987.

-
- [29] RUBIN, D. B. The design of a general and flexible system for handling nonresponse in sample surveys. *The American Statistician*, [S. l.], v. 58, n. 4, p. 298–302, 2004.
- [30] SCHAFER, J. L. *Analysis of incomplete multivariate data*. London: Chapman-Hall, 1997.
- [31] SCHAFER, J. L.; GRAHAM, J. W. Missing data: Our view of the state of the art. *Psychological Methods*, [S. l.], v. 7, n. 2, p. 147–177, 2002.
- [32] SCHEUREN, F. *Projeto articulado com o Social Security Administration e o US Census Bureau*. [S. l.]: [s. n.], 1977.
- [33] SCHOBER, P.; VETTER, T. R. Repeated measures designs and analysis of longitudinal data: If at first you do not succeed try, try again. *Anesthesia and Analgesia*, [S. l.], v. 127, n. 2, 2018.
- [34] SILVA, M. J. C. *Imputação múltipla comparação e eficiência em experimentos multi-ambientais*. Tese (Tese (Doutorado)) — Universidade de São Paulo, Piracicaba, 2012.
- [35] SINGER, J. M.; NOBRE, J. S.; ROCHA, F. M. M. *Análise de dados longitudinais*. [S. l.]: [s. n.], 2018.
- [36] BUUREN, S. V.; GROOTHUIS-OUDSHOORN, K. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, [S. l.], v. 45, n. 3, 2011.
- [37] BUUREN, S. V. *Flexible Imputation of Missing Data*. [S. l.]: Chapman & Hall/CRC, 2012.
- [38] HIPPEL, P. T. V. *How to impute interactions, squares and other transformed variables*. London: Chapman-Hall, 2009.
- [39] WHITE, I. R.; ROYSTON, P.; WOOD, A. M. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, [S. l.], v. 40, n. 4, p. 377–399, 2011.
- [40] YATES, F. The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture*, [S. l.], v. 1, n. 2, p. 129–142, 1933.