

PLANO DE AULA - MINICURSO

Estatística Básica e Análise de Dados no R

Versão Intensiva - Fundamentos

Nível:	Básico
Carga Horária:	20 horas
Modalidade:	Presencial ou Online
Pré-requisitos:	Conhecimentos básicos de informática
Data de Elaboração:	21/12/2025

1. APRESENTAÇÃO DO MINICURSO

Este minicurso intensivo de 20 horas oferece uma introdução prática e aplicada à análise de dados utilizando a linguagem R, abrangendo desde os fundamentos da programação até estatística inferencial e criação de visualizações profissionais. Focado em capacitar os participantes rapidamente para realizar análises de dados completas, o curso combina exposições teóricas objetivas com exercícios práticos intensivos, preparando os alunos para aplicar imediatamente os conhecimentos adquiridos em suas atividades profissionais ou acadêmicas.

2. OBJETIVOS DO MINICURSO

Objetivo Geral:

Capacitar os participantes a realizar análises estatísticas básicas e intermediárias completas utilizando R, incluindo manipulação de dados, testes estatísticos e visualizações profissionais.

Objetivos Específicos:

- Dominar a sintaxe básica do R e operações fundamentais
- Utilizar tidyverse para manipulação eficiente de dados
- Importar, limpar e processar dados de diferentes fontes
- Aplicar estatística descritiva e testes inferenciais
- Criar gráficos profissionais e tabelas formatadas
- Desenvolver fluxo completo de análise de dados

3. PÚBLICO-ALVO

Estudantes, pesquisadores e profissionais de diversas áreas que necessitem realizar análises estatísticas e desejam uma formação intensiva e prática em R. Ideal para quem busca rapidamente adquirir habilidades essenciais em análise de dados.

4. METODOLOGIA

O minicurso será ministrado de forma intensiva com foco na prática. Cada conceito será seguido imediatamente de exercícios aplicados. Os participantes trabalharão com dados reais desde o primeiro dia, desenvolvendo um projeto cumulativo ao longo do curso.

MÓDULO 1: FUNDAMENTOS DO R E TIDYVERSE

Carga Horária: 6 horas

Conteúdo Programático:

1.1 Introdução ao R e RStudio

- Instalação e ambiente de trabalho
- Interface do RStudio
- Pacotes essenciais (install.packages, library)

1.2 Sintaxe Básica e Estruturas de Dados

- Operadores e atribuições
- Vetores, data frames e listas
- Indexação e subsetting
- Funções básicas (sum, mean, median, sd, etc.)

1.3 Introdução ao Tidyverse

- Filosofia tidy data
- Pipe operator (%>%)

1.4 dplyr: Manipulação de Dados

- **select()**: Seleção de colunas
- **filter()**: Filtragem de linhas
- **mutate()**: Criação/transformação de variáveis
- **group_by() + summarise()**: Agregação de dados
- **arrange()**: Ordenação

1.5 tidyr: Reorganização de Dados

- pivot_longer() e pivot_wider()
- Tratamento de valores NA

Prática Intensiva:

- Exercícios progressivos de manipulação de dados
- Pipelines de transformação

MÓDULO 2: IMPORTAÇÃO, LIMPEZA E PROCESSAMENTO DE DADOS

Carga Horária: 4 horas

Conteúdo Programático:

2.1 Importação de Dados

- Arquivos de texto: `read_csv()`, `read_delim()`, `read_tsv()`
- Arquivos Excel: `read_excel()` (`readxl`)
- Parâmetros importantes: `col_types`, `na`, `skip`

2.2 Inspeção e Diagnóstico

- `glimpse()`, `head()`, `str()`, `summary()`
- Identificação de problemas de qualidade
- Valores missing e duplicados

2.3 Limpeza de Dados

- Padronização de nomes (`janitor::clean_names`)
- Conversão de tipos de dados
- Tratamento de NA (`drop_na`, `replace_na`, `fill`)
- Remoção de duplicatas (`distinct`)

2.4 Processamento Avançado

- Joins (`left_join`, `inner_join`, etc.)
- Combinação de datasets
- Criação de variáveis derivadas

2.5 Exportação

- `write_csv()`, `writexl::write_xlsx()`

Prática Intensiva:

- Pipeline completo: importação → limpeza → processamento → exportação
- Trabalho com múltiplos datasets reais

MÓDULO 3: ESTATÍSTICA DESCRITIVA, UNIVARIADA E INFERENCIAL

Carga Horária: 6 horas

Conteúdo Programático:

3.1 Estatística Descritiva

- Medidas de tendência central (média, mediana, moda)
- Medidas de dispersão (variância, desvio padrão, IQR)
- Medidas de posição (quartis, percentis)
- Resumos estatísticos (summary, skimr::skim)

3.2 Análise Univariada

- Distribuição de variáveis contínuas
- Análise de frequências para categóricas
- Identificação de outliers
- Testes de normalidade (Shapiro-Wilk)

3.3 Fundamentos de Inferência

- Conceitos: população, amostra, hipóteses
- Valor-p e nível de significância
- Intervalos de confiança

3.4 Testes Paramétricos

- Teste t de Student (uma e duas amostras)
- Teste t pareado
- ANOVA one-way
- Testes post-hoc (Tukey HSD)

3.5 Testes Não-Paramétricos

- Mann-Whitney U (wilcox.test)
- Kruskal-Wallis
- Quando usar cada teste

3.6 Testes de Associação

- Qui-quadrado (chisq.test)
- Correlação (Pearson, Spearman)

3.7 Pacotes Auxiliares

- rstatix: Pipeline tidy para testes
- Interpretação e apresentação de resultados

Prática Intensiva:

- Análise exploratória completa
- Aplicação de múltiplos testes estatísticos
- Interpretação de resultados

MÓDULO 4: VISUALIZAÇÃO DE DADOS E CRIAÇÃO DE TABELAS

Carga Horária: 4 horas

Conteúdo Programático:

4.1 ggplot2: Gramática de Gráficos

- Conceito e estrutura em camadas
- Estética (aes) e geometrias (geom)

4.2 Gráficos Essenciais

- Dispersão (geom_point)
- Barras (geom_bar, geom_col)
- Boxplot (geom_boxplot)
- Histogramas e densidade
- Linhas (geom_line)

4.3 Personalização

- Títulos e labels (labs)
- Cores e temas (scale_*, theme_*)
- Facetas (facet_wrap)

4.4 ggpubr: Gráficos Estatísticos

- Gráficos com p-valores automáticos
- stat_compare_means()
- ggboxplot, ggviolin com comparações

4.5 Combinação de Gráficos

- ggarrange() para painéis
- Exportação (ggsave)

4.6 Tabelas Profissionais com gt

- Estrutura básica de tabelas gt
- Formatação de colunas (fmt_number, fmt_percent)
- Estilização:
 - * Cabeçalhos e títulos (tab_header)
 - * Cores e destaque (tab_style, cell_fill)
 - * Bordas e alinhamento
- Resumos e totais (summary_rows)
- Agrupamento de linhas (tab_row_group)
- Notas de rodapé (tab_footnote, tab_source_note)
- Exportação de tabelas (gtsave)
- Integração com dplyr para dados sumarizados

4.7 gtsummary: Tabelas Estatísticas

- tbl_summary(): Estatísticas descritivas por grupo
- tbl_regression(): Tabelas de regressão formatadas

- `add_p()`: Adição de p-valores
- Customização e exportação

Prática Intensiva:

- Criação de dashboard visual com múltiplos gráficos
- Desenvolvimento de tabelas publicáveis
- Projeto: Relatório completo com gráficos e tabelas profissionais

5. AVALIAÇÃO

Avaliação contínua baseada em:

- Participação ativa nos exercícios práticos (30%)
- Mini-projeto ao final de cada módulo (30%)
- Projeto final integrador: Análise completa de dataset real (40%)

6. PROJETO FINAL

O projeto final será desenvolvido nas últimas 2 horas e deverá incluir:

- Importação e limpeza de dataset fornecido
- Análise exploratória com estatísticas descritivas
- Aplicação de testes estatísticos apropriados
- Criação de pelo menos 4 visualizações profissionais
- Elaboração de 2 tabelas formatadas com gt
- Interpretação e conclusões dos resultados

7. RECURSOS NECESSÁRIOS

- Computador com R (≥ 4.0) e RStudio instalados
- Pacotes: tidyverse, readxl, janitor, skimr, rstatix, ggpubr, gt, gtsummary
- Datasets de exemplo fornecidos
- Material didático digital

8. CRONOGRAMA SUGERIDO

Dia	Horário	Módulo	Tópico
1	4h	Módulo 1	Fundamentos R + Tidyverse (Parte 1)
2	4h	Módulo 1	Tidyverse (Parte 2) + Exercícios
3	4h	Módulo 2	Importação e Processamento de Dados
4	4h	Módulo 3	Estatística Descritiva e Inferencial (Parte 1)
5	4h	Módulo 3 + 4	Inferencial (Parte 2) + Visualização e Tabelas + Projeto Final

9. BIBLIOGRAFIA ESSENCIAL

- Wickham, H., & Grolemund, G. (2017). *R for Data Science*. <https://r4ds.had.co.nz/>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage.
- RStudio Cheatsheets: <https://www.rstudio.com/resources/cheatsheets/>
- gt package documentation: <https://gt.rstudio.com/>
- STHDA: <http://www.sthda.com/english/>

10. OBSERVAÇÕES FINAIS

Este minicurso intensivo foi projetado para maximizar o aprendizado em 20 horas. Recomenda-se dedicação integral durante o período e prática adicional após a conclusão. Os participantes receberão todos os scripts, datasets e material de apoio. Certificados serão emitidos para participantes com frequência mínima de 75% e conclusão do projeto final.