

R PARA ANÁLISE DE DADOS

FICHA DO CURSO

1. Resumo do curso**Curso:** R para a análise de dados**Modalidade:** Ensino à distância**Carga horária total:** 40 horas**Carga horária síncrona:** 2,5 horas por semana**Início da oferta:** Janeiro/2022**Pré-requisito:** Nenhum**Duração do curso:** 10 semanas**Carga horária semanal:** 4 horas**Carga horária assíncrona:** 1,5 horas por semana**Fim da oferta:** Abril/2022**2. Objetivos**

Habilitar o(a) aluno(a) a utilizar o software R de computação estatística e gráficos para análise de dados. Realizar tarefas de importação, tratamento e organização, manipulação e análise de dados com estatística descritiva e construção de gráficos.

Ao final do curso o participante será capaz de instalar e configurar o ambiente para análise de dados com R, instalar e gerir pacotes, importar, manipular dados, confeccionar, interpretar e exportar gráficos, aplicar e interpretar medidas resumo numéricas e exportar resultados de análise.

3. Programa

Instalação e configuração dos ambientes R e RStudio IDE para análise de dados; Visão geral e convenções da linguagem R; Aritmética e manipulação básica de vetores; Estruturas de programação e funções; Estruturas de dados; Importação de dados; Visualização de dados com ggplot2; Limpeza e organização para o formato tidy; Processamento e transformação de dados; Agregação e conciliação de dados; Manipulação de datas, strings e fatores; Estimacão e testes de hipótese; Análise de regressão linear simples.

4. Procedimentos didáticos

A cada semana (em um total de **10 semanas**) serão desenvolvidas as seguintes atividades pelos cursistas:

1. Acessar o material pré aula para preparar-se para a aula.
2. Assistir as aulas programadas para a semana no horário definido.
3. Trabalhar nos tutoriais guiados disponibilizados pelo professor.
4. Fazer o estudo individual de materiais indicados como: leituras complementares, resolução de exercícios e acesso a vídeos adicionais.
5. Participar do fórum do curso contribuindo com tópicos para a discussão ou respondendo e complementando tópicos em aberto relacionados ao conteúdo apresentado.
6. Realizar as atividades de avaliação semanal de desempenho e avaliação do curso.

Como suporte ao Ensino Remoto será utilizada a infraestrutura de AVA (Ambiente Virtual de Aprendizado) que é acessado pelo endereço <https://treinamento.omegadatascience.com.br> para cursistas cadastrados.

Serão observados os seguintes detalhes sobre os procedimentos didáticos no período de vigência do curso, conforme itens a seguir.

4.1 Sistema de comunicação durante aulas ao vivo

Será usado o sistema de comunicação por mensagens instantâneas do serviço de streaming utilizado para transmissão das aulas ao vivo durante a apresentação do professor. Durante a apresentação do professor, a comunicação por texto é com a equipe de monitores do curso. Durante a aula, em momentos de pausa para dúvidas definidos pelo professor, os alunos poderão interagir por voz e vídeo.

4.2 Sistema de comunicação entre as aulas ao vivo

Será usado o recurso de fórum do AVA, permitindo que alunos abram tópicos e contribuam com a discussão ou resolução de dúvidas em tópicos abertos.

4.3 Sistema de comunicação nos plantões de dúvida

Será usado o sistema de comunicação por mensagens instantâneas, voz e vídeo do serviço de streaming utilizado para transmissão dos plantões de dúvidas.

4.4 Materiais didáticos

1. No Moodle terá o guia rápido do curso, onde constarão detalhes das plataformas de ensino, instruções sobre a comunicação com os instrutores, monitores e coordenação, cronograma e programa da disciplina, orientação das atividades remotas assíncrona e síncrona, seus prazos e pesos na determinação de desempenho e participação.
2. Dentro da conta de cada usuário na Sympla estarão disponíveis os links para a transmissão das aulas e dos plantões de dúvidas.
3. No Moodle estarão os links para acesso às transmissões gravadas das aulas. Elas ficarão hospedadas no canal do Youtube designado para o treinamento.
4. Slides (em formato pdf) usados nas aulas bem como eventuais arquivos como scripts, planilhas e arquivos de dados serão disponibilizados no Moodle.
5. No Moodle será usado o recurso de fórum para cursistas abrirem e contribuírem em tópicos, tirarem dúvidas e contribuírem com materiais.
6. No Moodle poderão ser indicados materiais complementares àqueles produzidos pelos professores do curso, como: vídeos que demonstram conceitos e ideias sobre os conteúdos das unidades didáticas e links para outros sítios eletrônicos contendo materiais relacionados aos temas das aulas.
7. No Moodle serão disponibilizadas atividades avaliativas de desempenho e participação bem como formulários de avaliação da qualidade do curso.

4.5 Distribuição da carga horária do cursista

A Tabela 1 a seguir apresenta distribuição da carga horária semanal total do cursista, de 4 horas por semana, nas atividades semanais previstas na vigência do curso.

Tabela 1 - Carga horária semanal do(a) cursista.

| Atividade semanal | Carga horária (horas) | Fração (%) |
|--|-----------------------|------------|
| Estudo individual de conteúdo pré-aula | 00:30 | 12,5 |
| Participação na aula ao vivo | 01:30 | 37,5 |
| Estudo individual pós aula | 00:30 | 12,5 |
| Participação no plantão de dúvidas | 01:00 | 25,0 |
| Participação no fórum | 00:30 | 12,5 |

| | | |
|-------|-------|-----|
| Total | 04:00 | 100 |
|-------|-------|-----|

4.6 Modelo de tutoria a distância

A dinâmica deste curso é disponibilizar para o cursista um serviço de autoaprendizagem que consta de vídeo-aulas síncronas com acesso posterior às gravações, disponibilização de materiais didáticos usados nos vídeos e complementares, atividades para a prática do conteúdo.

Além das aulas e respectivos materiais, o cursista terá um encontro remoto síncrono de plantão de dúvidas para interagir com o professor e pares.

No AVA haverá durante todo período de vigência do curso, fórum separados por encontro no qual os alunos poderão abrir e contribuir em tópicos diversos, como por exemplo, dúvidas em relação ao conteúdo, solicitar e/ou recomendar materiais complementares, contribuir com tópicos abertos pelos pares, apresentar resultados de atividades desenvolvidas, entre muitas outras possibilidades. Será gerado um FAQ (frequently asked questions) com as questões mais relevantes ou populares do fórum.

Em cada unidade didática, o cursista poderá avaliar seu aprendizado por meio de quizzes sobre os tópicos cobertos no encontro imediatamente anterior.

E por fim, toda unidade didática terá uma enquete anônima com questões voltadas à avaliação do curso e espaço de texto livre para expressar sua opinião.

5 Conteúdo programado

O curso está dividido em unidades didáticas (UD), sendo uma UD para cada semana de atividades. Na sequência, estão listadas as 10 UD's, com respectivo conteúdo planejado, na Tabela 2.

Tabela 2 - Cronograma detalhado do conteúdo das unidades didáticas.

| UD | Conteúdo previsto | Período |
|----|---|---------------|
| 1 | Ambientação ao R Instalação e configuração do ambiente R Instalação e tour pela RStudio IDE Outras IDEs para trabalhar com R Instruções e comentários e uso REPL Atribuição, espaço de trabalho e diretório de trabalho | 31/01 - 04/02 |

Arquivos da linguagem R (.R, .Rprofile, .Rhistory, .RData, etc)
 Instalação e inspeção de pacotes
 Acesso à documentação

2 Aritmética básica

Operações matemáticas (+, -, *, /, ^), de
 Operações de comparação (<, <=, ==, !=, >=, >)
 Operadores lógicos (&, &&, |, ||, !)
 Tipos e classes para vetores (int, dbl, fct, chr, lgl)
 Predicados lógicos (is.*) e conversões (as.*)
 Criação de vetores e manipulação de vetores
 Seleção por indexação e condicional
 Aritmética entre vetores e regra da reciclagem
 Funções matemáticas (log, exp, sin, ..., round, floor, ceiling, mod)
 Funções estatísticas (mean, median, sd, min, max, etc)

07/02 - 11/02

3 Estruturas de programação e funções

Estruturas de controle (if-else, switch)
 Estruturas de repetição (for, while, repeat, next/break)
 Componentes de uma função (nome, argumentos, corpo, retorno, namespace)
 Argumentos, valores padrão, dots e escopo
 Definição de funções
 Tratamento de exceções (stop, warning, message, try)

14/02 - 18/02

4 Importação de dados

Estruturas de dados (vetor, matriz, lista e data.frame)
 Operações com listas
 Importação pela entrada padrão (scan()) e clipboard.
 Leitura de arquivos de campos delimitados (tsv, csv) e comprimento fixo (fwf)
 Argumentos para encoding, comentários, aspas, datas, valores ausentes, etc
 Inspeção das dimensões, atributos e estrutura
 Conversões de tipo de valor, transformações e medidas resumo

21/02 - 25/02

5 Visualização de dados

O ecossistema tidyverse
 Importação com {readr}

07/03 - 11/03

A gramática dos gráficos do {ggplot2}
 Gráficos para exibir a distribuição (barras, setores, histograma, ecdf, densidade)
 Gráficos para exibir a relação (dispersão, séries, barras, mosaico, spider, etc)
 Análise exploratória ágil com {visdat}, {DataExplorer}, {skimr}, etc

6 Arrumação e conciliação de dados

O formato tidy
 Separação e concatenação de campos (separate(), unite())
 Valores ausentes (replace_na(), fill_na())
 Pivotagem (pivot_longer() e pivot_wider())
 Renomear variáveis (rename())
 Mudar ordem das colunas (relocate())
 Concatenação (bind_rows(), bind_cols())
 Junção de tabelas (*_join())
 List columns (nest(), unnest())

14/03 - 18/03

7 Transformação e agregação de dados

Os principais verbos do {dplyr}
 Ordenação de registros (arrange())
 Filtro de registros (filter())
 Seleção de variáveis (select(), select_if(), select_at())
 Transformação de variáveis (mutate(), mutate_at(), mutate_if(), mutate_all())
 Agregação (summarise(), _if(), _at(), _all())
 Operações de transformação por estrato (group_by(), group_split(), etc)
 Aplicações: estratificação por ano, mês, semana, etc

21/03 - 25/03

8 Manipulação de datas, strings e fatores

Formatos para variáveis de data (data, data-hora, duração, etc)
 Extrações de componentes: dia, mês, ano, hora, dia da semana, etc
 Uso de defasagens em datas (lag(), lead())
 Detecção de padrões em strings (grep(), str_detect())
 Extração e substituição de padrões (gsub(), str_replace())

28/03 - 01/04

Operações com fatores: reordenação, rotulação, mesclar níveis

9 **Estimação e testes de hipótese**

Intervalo de confiança para a média e proporção

Teste de hipótese para média, mediana e proporção (z, t, t-pareado, não paramet.)

Teste de hipótese para a variância

04/04 - 08/04

Teste de hipótese para tabelas de contingência

Teste de hipótese para a correlação

Análise de variância (ANOVA)

10 **Análise de regressão linear simples**

Especificação, ajuste e interpretação de regressão linear simples

Análise diagnóstica (falta de ajuste, homocedasticidade, normalidade, pontos influentes)

11/04 - 15/04

Medidas corretivas (transformação Box-Cox, uso de polinômios)

Intervalo de confiança e de predição

Encerramento do curso

6 Avaliação de desempenho e participação

6.1 Desempenho no curso

O desempenho no curso será determinado pela nota média das 7 maiores notas dos 10 quizzes semanais aplicadas aos cursistas.

6.2 Participação no curso

A frequência de participação no curso será determinada pela realização dos quizzes semanais bem como pelo preenchimento das enquetes semanais de avaliação do curso ou questões do material pré-aula/pós-aula. Receberão certificados de participação aqueles com 70% ou mais de aproveitamento nestas atividades.

7 Referências bibliográficas

1. Wickham, Hadley, and Garrett Golemund. **R for data science: import, tidy, transform, visualize, and model data**. Sebastopol, CA: O'Reilly Media, 2016. Print.
2. Matloff, Norman S. **The art of R programming: tour of statistical software design**. San Francisco: No Starch Press, 2011. Print.
3. Wickham, Hadley. **Ggplot2: elegant graphics for data analysis**. Dordrecht New York: Springer, 2009. Print.
4. Golemund, Garrett. **Hands-on programming with R**. Sebastopol, California: O'Reilly Media, Safari Books Online, 2014. Print.
5. Wickham, Hadley. **Ggplot2: elegant graphics for data analysis**. Dordrecht New York: Springer, 2009. Print.
6. Verzani, John. **Using R for introductory statistics**. Boca Raton: Chapman & Hall/CRC, 2005. Print.
7. Dalgaard, Peter. **Introductory statistics with R**. New York: Springer, 2008. Print.

8 Informações sobre o professor

Abaixo estão as informações dos membros da equipe envolvidos diretamente com os cursistas.

Tabela 3 – Informações dos membros da equipe.

| Função | Nome | Contato |
|--------------------|------------------------|------------------------|
| Instrutor do curso | Fernando de Pol Mayer | fernando.mayer@ufpr.br |
| Monitor do curso | A definir | A definir |
| Coordenador | Walmes Marques Zeviani | walmes@ufpr.br |