



# CURSO DE CIÊNCIA DE DADOS APLICADA AO PODER JUDICIÁRIO



**R PARA ANÁLISE DE DADOS**

## FICHA DO CURSO

**Curso:** R para a análise de dados**Modalidade:** EAD - autoinstrucional**Carga horária total:** 40 horas**Carga horária semanal:** 4 horas**Início da oferta:****Fim da oferta:****Pré-requisito:** Não há pré-requisitos recomendados para o cursista**Conteudistas:** Walmes Marques Zeviani**1. Objetivos**

Habilitar o(a) aluno(a) a utilizar o software R de computação estatística e gráficos para análise de dados. Isso corresponde a realizar tarefas de importação, tratamento e organização de dados, 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, exportar resultados de análise.

**2. Programa**

Instalação e configuração da ambientação ao 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; Estimação e testes de hipótese; Análise de regressão linear simples.

### 3. Procedimentos didáticos

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

1. Acessar o material pré aula para preparar-se para a aula.
2. Assistir as vídeos-aulas programadas para o módulo
3. Fazer o estudo individual de materiais indicados como: leituras complementares, resolução de exercícios e acesso a vídeos adicionais.
4. Realizar as atividades de avaliação do módulo no formato de quiz.

Como suporte ao Ensino Remoto será utilizada a infraestrutura da plataforma moodle do Centro de Formação e Aperfeiçoamento de Servidores do Poder Judiciário (CEAJUD).

### 4. Procedimentos metodológicos

Conjunto de vídeo-aulas separadas por módulos de conhecimento acompanhado de avaliações para reforçar o aprendizado.

### 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%
Assistir vídeo-aulas	02:00	50%
Estudo individual pós aula	00:30	12,5%
Atividades avaliativas	01:00	25%

Total	04:00	100%
-------	-------	------

## 6. Conteúdo programado

O curso está dividido em 10 módulos sendo que cada módulo configura uma semana de atividades. O conteúdo programado é apresentado na Tabela 2.

Tabela 2 - Cronograma detalhado do conteúdo das unidades didáticas.

Módulo	Conteúdo programado	Duração
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 Arquivos da linguagem R (.R, .Rprofile, .Rhistory, .RData, etc) Instalação e inspeção de pacotes Acesso à documentação	4 horas
2	Aritmética básica Operações matemáticas (+, -, *, /, ^) 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)	4 horas

- 
- 3 Estruturas de programação e funções  
 Estruturas de controles (if-else, switch)  
 Estruturas de repetição (for, while, repeat, next/break)  
 Componentes de uma função (nome, argumentos, corpo, retorno, namespace) 4 horas  
 Argumentos, valores padrão, dots e escopo  
 Definição de funções  
 Tratamento de exceções (stop, warning, message, try)
- 
- 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) 4 horas  
 Argumentos para encoding, comentários, aspas, datas, valores ausentes, etc  
 Inspeção das dimensões, atributos e estrutura  
 Conversões de tipo de valor, transformações e medidas resumo
- 
- 5 Visualização de dados  
 O ecossistema tidyverse  
 Importação com {readr}  
 A gramática dos gráficos do {ggplot2}  
 Gráficos para exibir a distribuição (barras, setores, histograma, ecdf, densidade) 4 horas  
 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()) 4 horas  
 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())

---

## 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()) 4 horas  
 Agregação (summarise(), \_if(), \_at(), \_all())  
 Operações de transformação por estrato (group\_by(),  
 group\_split(), etc)  
 Aplicações: estratificação por ano, mês, semana, etc

---

## 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 4 horas  
 Uso de defasagens em datas (lag(), lead())  
 Detecção de padrões em strings (grep(), str\_detect())  
 Extração e substituição de padrões (gsub(), str\_replace())  
 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.) 4 horas  
 Teste de hipótese para a variância  
 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) Medidas corretivas (transformação Box-Cox, uso de polinômios) Intervalo de confiança e de predição	4 horas
----	---	---------

---

Encerramento do curso

---

## 7. Desempenho no curso

O desempenho no curso será determinado pela nota em atividade avaliativa aplicada aos cursistas. Requer-se, no mínimo, 70% de aproveitamento para obtenção de certificado.

## 8. 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, 2015. Print.
7. Dalgaard, Peter. **Introductory statistics with R**. New York: Springer, 2008. Print.

## 9. Informações sobre os conteudistas

### Walmes Marques Zeviani

Engenheiro Agrônomo pela Universidade Federal da Grande Dourados (2003-2007) e Mestre (2008-09) e Doutor (2009-13) em Estatística e Experimentação Agropecuária pela Universidade Federal de Lavras. Professor Adjunto III do Departamento de Estatística da Universidade Federal do Paraná (2010-hoje). Atua principalmente com ensino no Curso de Bacharel em Estatística e na Especialização em Data Science & Big Data (UFPR). Tem experiência com cursos de capacitação do software R para planejamento e análise de experimentos, modelos de regressão linear e não linear, manipulação e visualização de dados e construção de interfaces gráficas e pacotes R.

## 10. Requisitos Técnicos

Computador com acesso à internet. Permissão para instalar programas.



CURSO DE CIÊNCIA  
DE DADOS APLICADA  
AO PODER JUDICIÁRIO

