

Curso MADEVR100 - Analytics | Introdução ao R

21,00 Horas

Introdução

Este curso aborda o tema da programação em R e a sua utilização para uma análise de dados eficaz. Serão abordadas as formas de instalar e configurar o software necessário para um ambiente de programação estatística. Esta ação ajuda a conhecer como os conceitos da linguagem de programação genérica são implementados numa linguagem estatística de alto nível. O curso cobre questões práticas em computação estatística que inclui programação em R, leitura de dados em R, acesso a pacotes R, escrita de funções R, depuração, criação de perfil de código R, organização e comentários de código R. O curso seguirá uma abordagem fortemente prática com vários exemplos e exercícios práticos.

Público-alvo

Este curso tem como destinatários quem pretenda desenvolver conhecimentos da linguagem R para análise de dados.

Quando completar o curso

Este curso permite obter o conhecimento e experiência para:

- Instalar o R e um interface gráfico R (“R IDE”)
- Dominar os fundamentos da análise de dados manipulando estruturas de dados comuns, como vetores, matrizes e quadros de dados
- Usar o R para limpar, analisar e visualizar dados.
- Conhecer todo o pipeline de ciência de dados, desde a aquisição de dados até à publicação.
- Realizar análises de regressão, mínimos quadrados e inferência usando modelos de regressão.

Pré-requisitos

Os participantes devem ter conhecimentos básicas de análise de dados, manipulando estruturas de dados comuns, como registos, tabelas.

Exames

(não existem exames)

Certificado

Este curso confere certificado de frequência a todos os participantes que frequentem no mínimo 80% das horas do curso.

Conteúdo em detalhe

Módulo 1: Primeiros passos

Este módulo cobre o básico para você começar com R, informações sobre a mecânica do curso e algumas demonstrações sobre a instalação de R. Esta seção cobre a história de R, examina os tipos de dados básicos e m R e descreve as funções para leitura e escrever dados.

- Instalar o R
- Escrever código / Configurar seu diretório de trabalho
- Introdução
- Visão geral e história de R
- Obter ajuda
- Entrada e avaliação do console R
- Tipos de dados

- Objetos e atributos R

- Vetores e Listas

- Matrizes

- Fatores

- Valores ausentes

- Quadros de dados

- Atributo de nomes

- Lendo

- Dados tabulares

- Tabelas grandes

- Formatos de dados textuais
- Conexões: interfaces para dados externos
- Subsetting

- Fundamentos

- Listas

- Matrizes

- Correspondência parcial

- Remover valores ausentes

- Operações vetorizadas
- Introdução ao “swirl”

Módulo 2: Programação com R

Neste módulo cobrimos tópicos importantes como estruturas e funções de controle.

- Estruturas de controle

- Introdução

- IF-ELSE

- Ciclos FOR

- Ciclos WHILE

- REPEAT, NEXT, BREAK

- A primeira função R

- Funções

- Regras de escopo

- “Symbol binding”

- Regras de “scoping” em R

- Regras e padrões de programação

- Datas e horas

Módulo 3: Funções de LOOP e depuração

Este módulo cobre funções de LOOP e as ferramentas de depuração em R. Esses aspetos de R são úteis tanto para trabalho interativo quanto para escrever códigos mais longos e portanto, comuns na prática diária.

- Funções de LOOP
 - lapply
 - apply
 - mapply
 - tapply
 - split
- Ferramenta de depuração
 - Diagnosticando o problema
 - Ferramentas simples
 - Usando as ferramentas

Módulo 4: Simulação & Profiling

Este módulo aborda como simular dados em R, que servirá como base para fazer estudos de simulação. Também cobrimos o Profiler em R que permite recolher informações detalhadas sobre como as funções R estão a ser executadas e identificar “bottlenecks” que podem ser resolvidos. O Profiler é uma ferramenta fundamental para ajudar a otimizar programas. Finalmente, cobrimos a função `str`, muito útil em R.

- A função `str`
- Simulação
 - Gerar Números Aleatórios

- Simular modelo linear

- Amostragem Aleatória

- R Profiler