

MBA EXECUTIVO EM BUSINESS ANALYTICS E BIG DATA

TRABALHO DE MODELAGEM ESTATÍSTICA AVANÇADA

Professor: Paulo Maranhão

Prazo: 08/05/2022

Questão 1: Seguem abaixo cinco bases de dados que se encontram do próprio R onde cada variável resposta encontra-se entre parêntesis. Cada grupo ou cada aluno deve escolher três base de dados. Para cada base de dados deve ser feito uma análise estatística conforme a teoria da disciplina de Modelagem Estatística Avançada. Cada análise deve responder os seguintes itens:

- Usando o R **determine o melhor modelo linear possível** para a estimar o valor médio de cada variável resposta. Proponha quaisquer transformações que achar necessário para melhorar o ajuste do modelo.
- Os pressupostos dos modelos foram atendidos? **Apresente os resultados graficamente e analiticamente** (testes formais) comprovando se as suposições foram satisfeitas.
- Utilize uma parte dos dados para gerar previsões e intervalos de confiança. É possível dizer que o modelo foi bem ajustado?

Bases de Dados:

- **mtcars (hp)**
- **LifeCycleSavings (sr)**
- **quakes (mag)**
- **swiss (fertility)**
- **trees (volume)**

Questão 2: Imagine que uma empresa de determinado país quer fazer um estudo sobre a produção bruta do país. Para isso resolveu utilizar a base de dados da Organização para a Cooperação e Desenvolvimento Econômico (OECD) cujos dados podem ser encontrados no seguinte sítio:

https://stats.oecd.org/Index.aspx?DataSetCode=IO_GHG_2021#

A empresa decidiu utilizar os dados encontrados na **popular queries**, especificamente os dados contidos nas planilhas disponíveis na **Trade in Value Added (TiVA) 2021 ed: Principal Indicators**.

Dessa forma, foi solicitado ao grupo de estudos do **MBA Executivo em Business Analytics e Big Data da FGV** um **relatório** sobre os dados constantes na base de dados, onde a **variável resposta** a ser estudada é a variável **Produção Bruta**. Dessa forma, cada grupo ou cada aluno (trabalho individual) deve escolher uma país ou um grupo de países (ex: União Europeia, G20, países não membros da OECD, etc) e responder as seguintes questões:

- Usando o R **determine o melhor modelo linear possível** para a estimar o valor médio da variável resposta. Proponha quaisquer transformações que achar necessário para melhorar o ajuste do modelo.
- Quais fatores afetam a variável resposta?

MBA EXECUTIVO EM BUSINESS ANALYTICS E BIG DATA

- b) O modelo é adequado para estimar um valor futuro da variável resposta?
- c) Os pressupostos dos modelos foram atendidos? **Apresente os resultados graficamente e analiticamente** (testes formais) comprovando se as suposições foram satisfeitas.
- d) Utilize uma parte dos dados para gerar previsões e intervalos de confiança. É possível dizer que o modelo foi bem ajustado?

Obs1: A análise deve ser feita em forma de relatório e deve ser enviada obrigatoriamente para a pasta “**Trabalho de Modelagem Estatística**” que se encontra dentro da aba **Entrega de Atividades do E-Class**. Quaisquer dúvidas sobre o trabalho podem ser encaminhadas para o e-mail maranhao@ime.eb.br.

Obs2: Um vídeo contendo uma explicação mais detalhada sobre a **questão 2** pode ser encontrado no E-Class.

Obs3: A data de entrega do trabalho é 08 de maio de 2022.