

Teste de seleção para desenvolvedor estatística R

Regras:

- Prazo de sete dias corridos a partir do envio do teste. Após o término do prazo, pode-se submeter/enviar o código **mesmo que incompleto**.
- O sistema deve ser feito na linguagem R e com base de dados.
- Para a base de dados pode ser utilizado conexão com MySQL ou uso de arquivos em CSV.
- O código deve conter instruções.

Diferenciais e análise:

- Estrutura do projeto, código organizado, comentado e limpo.
- Pode ser utilizado o Rmarkdown.
- Documentação/comentários do código.
- Uso de funções.
- O código pode ser disponibilizado via E-mail, GitHub, Gitlab ou Bitbucket.

Objetivo 1:

Desenvolver scripts com o intuito de analisar o banco de dados disponibilizado no Drive (https://drive.google.com/drive/u/0/folders/11ksjaLWk_N-iczjWmOZpkH3Tvst_gQxw), o **dicionário das variáveis** está no mesmo local do banco de teste. As linhas do banco de dados possuem informações de estudantes que participaram do Enem de 2017, os dados foram retirados do microdados do Inep de 2017. O banco contém uma amostra com 2000 estudantes e 18 variáveis. Qualquer análise pode ser realizada assim como qualquer técnica, entretanto, algumas informações são de importantes ser apresentadas, são elas:

- Rotular os dados de acordo com o dicionário disponibilizado (NA = NR);
- Tabela com média e desvio padrão das notas de Ciências Humanas, Ciências da Natureza, Literatura e Matemática;
- Tabela com número e porcentagem de estudantes com idade maior que 17 e menor que 23;
- Tabela com número e porcentagem de estudantes por Estado;
- Tabela com idade média, mínima e máxima por tipo de escola;
- Tabela com média e desvio padrão da nota de Literatura divididos em Inglês e Espanhol;
- Tabela com número, porcentagem e média da nota de Matemática dos alunos com idade dividido em dois grupos, menores ou com 21 anos de idade e maiores de 21 anos de idade e Região (Centro-Oeste, Sudeste, Sul, Norte, Nordeste);
- Taxa de acerto de cada questão por área (Ciências da Natureza, Ciências Humana, Literatura e Matemática);
- Encontrar um modelo de regressão com objetivo de estimar a nota de Ciências Humanas com o vetor resposta das 45 questões das perguntas aplicadas na prova de 2017 do Enem.

Objetivo 2: Qual outra técnica você indicaria para estimar a nota do Enem de uma pessoa?