

Questionário

1. Qual a diferença entre *missing values* e valores vazios?

- a) *Missing values* são valores omissos, valores em falta. Valores vazios são valores que não existem no contexto da variável.
- b) São a mesma coisa, valores em falta nos dados.
- c) Os *missing values* são valores que aparecem nos *dataframes* como NaN e os valores vazios aparecem como "?".
- d) *Missing values* são valores que temos de recuperar a todo o custo, valores vazios podemos substituir por 0.

2. O que significa que um *missing value* ser do tipo MNAR: *Missing Not At Random*?

- a) São valores em falta devido a erros nos dados.
- b) São valores em falta que têm uma distribuição Normal, em vez de uma distribuição aleatória.
- c) São valores em falta facilmente explicados pela restante informação disponível.
- d) São os *missing values* mais críticos visto que o motivo da sua ausência está relacionada com a variável de interesse do nosso estudo.

3. Verdadeiro ou Falso: "Sempre que a percentagem de *missing values* é inferior a 0,2% podemos eliminar as observações."

- a) Verdadeiro.
- b) Falso.

4. Quando devemos recuperar a informação em falta?

- a) Sempre, visto que é a solução com menor impacto para a nossa análise.
- b) Só em casos extremos, quando a percentagem de *missing* é muito reduzida, por saí muito caro.
- c) Sempre que os valores em causa sejam críticos para o estudo.
- d) Caso haja dinheiro e tempo, é a única solução correta.

5. Qual o principal objetivo da normalização de escalas?

- a) Garantir que todas as variáveis têm o mesmo peso no estudo.
- b) Tornar os modelos mais eficientes de executar.
- c) Por todas as variáveis entre [0:1].
- d) Garantir que todas as variáveis têm uma distribuição Normal.

6. Num estudo sobre o aproveitamento escolar dos alunos de um agrupamento de escolas, temos de *missing values* na variável “nível de escolaridade do encarregado de educação”.

Desenvolve uma abordagem para lidar com esta situação explicitando os pressupostos que consideras (ex.: relevância da variável para o estudo, % *missing values*, tipologia de *missing values*, etc.).

Pontuação: A pergunta 6 = 50%, restantes perguntas = 10%