

Exercícios de preparação para o exame

Para resolver os seguintes exercícios, pode usar um ficheiro de dados de uma das fichas de exercícios das aulas práticas, um ficheiro com dados proveniente de uma qualquer outra fonte, ou ainda, usar uma tabela já existente em **R** (e.g., `iris`).

Manipulação de ficheiros

1. Carregar um ficheiro de texto com dados, organizados em várias colunas. Identificar a parametrização necessária para carregar o ficheiro.
2. Guardar o conteúdo de um `data.frame` num ficheiro de texto, de acordo com determinadas especificações (por exemplo, com valores separados por vírgulas).

Aritmética

1. Calcular a média ponderada das colunas, de acordo com uma dada ponderação.
2. Subtrair a cada linha a respectiva média.

Lógica

1. Seleccionar as linhas em que o valor da última coluna está entre o 2º e o 3º quartil, dessa mesma variável.
2. Assinalar as linhas em que existem pelo menos 3 valores abaixo da média da variável, colocando **"OK"** ou **"Fail"** numa nova coluna.
3. Remover 10% das linhas com variância mais alta.

Factores

1. Carregar um novo ficheiro, com duas colunas, em que uma existe em comum com o primeiro ficheiro. A nova coluna tem valores do tipo `factor`, correspondendo a um ID único.
2. Juntar essa nova tabela à anterior, confirmando que existe pelo menos uma coluna em comum entre as duas tabelas.
3. Quantos registos há em cada um dos grupos?
4. Fazer o somatório de cada uma das variáveis, para cada ID.
5. Encontrar o ID com menor variância.
6. Somar 10% do valor de cada variável, às linhas do ID com média mais baixa.

Sequências

1. Seleccionar linhas, começando na 2ª linha, acabando na 42ª a contar do fim, de 7 em 7 linhas.
2. Quais são os ID's existentes entre os elementos seleccionados anteriormente?
3. Criar uma amostra com 20 elementos, retirados aleatoriamente, com igual probabilidade, mas com reposição.

Funções

1. Definir uma função que receba dois vectores e calcule o RMSE (Equação 1) entre eles. Assuma que um vector contém valores estimados de uma variável ($\hat{\theta}$), e que o outro tem valores observados da mesma variável (θ).

$$\text{RMSE}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})} = \sqrt{\text{E}((\hat{\theta} - \theta)^2)}. \quad (1)$$

2. Aplicar essa função em cada linha, em que o 2º vector é a média de todas as linhas. Guarde os valores calculados numa nova coluna.

Manipulação de dados

1. Removendo os registos duplicados, com quantos registos fica a tabela?
2. Modificar a coluna com os valores de RMSE, colocando cada valor elevado ao quadrado.