

TESTE 1 (INDIVIDUAL OU EM DUPLA)

INF-0612 – ANÁLISE DE DADOS

O objetivo deste teste é exercitar os conceitos vistos até agora no curso. São dois problemas a serem resolvidos.

1 Média Final

Foi solicitado a você calcular a média final de cada aluno. No arquivo “**teste1a.R**” você encontra 4 vetores: o vetor “**ids**” representa o registro acadêmico dos n alunos matriculados no curso e os vetores “**p1**”, “**p2**”, “**p3**” e “**p4**” representam o vetor das notas das provas 1, 2, 3 e 4 de cada aluno, respectivamente. A média final é uma média ponderada das notas, onde o peso p_i de cada nota x_i é definida como

$$p_i = \begin{cases} 1, & \text{se } x_i \geq 5 \\ 1.5, & \text{se } 2.5 \leq x_i < 5 \\ 2, & \text{se } x_i < 2.5. \end{cases}$$

1. (0.5 ponto) Você deve criar um data frame utilizando os vetores fornecidos e sempre que utilizar algum dado já existente no arquivo referir-se a este data frame (ou seja, você só pode utilizar os vetores fornecidos para criar este data frame).
2. (0.5 ponto) Você deve salvar nas variáveis “**dp1**”, “**dp2**”, “**dp3**” e “**dp4**” o desvio padrão das notas das provas 1, 2, 3 e 4, respectivamente.
3. (1.0 ponto) Você deve salvar no vetor “**medpond**” a média final de cada aluno, com duas casas decimais.
4. (1.0 ponto) Você deve salvar na variável “**notasrank**” os registros acadêmicos dos alunos tal que na primeira posição conste o registro acadêmico do aluno com a maior média final, na segunda posição conste o registro acadêmico do aluno com a segunda maior média final, e assim por diante.

Observação: para esse problema, as suas respostas devem ser feitas no arquivo “**teste1a.R**”.

2 Taxa de Cancelamento

Uma empresa de telecomunicações contratou você para fazer uma análise dos registros dos seus clientes. O objetivo principal dessa análise é aprender quais as características dos clientes que cancelaram a sua conta e dos clientes que permaneceram na empresa. No arquivo “**teste1b.R**” você encontra o data frame **customer_churn**. Esse data frame possui 7500 registros com as seguintes colunas:

- “customerID”: Identificador único do cliente;
- “gender”: Sexo do cliente;
- “Partner”: Se o cliente possui um parceiro ou não;
- “Dependents”: Se o cliente possui dependentes ou não;
- “tenure”: Número de faturas que o cliente já pagou;
- “Contract”: Tipo de contrato (“Month-to-month”, “One year” e “Two year”);
- “MonthlyCharges”: Valor cobrado mensalmente ao cliente;
- “Churn”: Se o cliente cancelou o seu contrato ou não.

Você deve realizar as seguintes tarefas nesse conjunto de dados:

1. (0.5 ponto) Remover linhas duplicadas. Note que você só precisa identificar linhas com customerID repetidos;
2. (0.5 ponto) Transformar todas as colunas com valores “Yes” e “No” para os valores booleanos “TRUE” e “FALSE”, respectivamente;

Agora, você deve responder algumas perguntas (Obs: A resposta de cada pergunta deve ser armazenada na variável indicada no arquivo “`teste1b.R`”). As respostas que contenham o identificador dos clientes devem ser ordenadas pela variável “customerID”.

3. (0.5 ponto) Qual cliente pagou mais faturas?
4. (1 ponto) Dentre os clientes que pagam mensalmente um valor maior ou igual a 50, qual cliente pagou mais faturas?
5. (1 ponto) Dentre os clientes com contrato “Month-to-month”, qual cliente pagou menos faturas?
6. Dentre as pessoas que não cancelaram o contrato:
 - (a) (1 ponto) Qual o total pago por mês para cada um dos três tipos de contrato (“Month-to-month”, “One year” e “Two year”)?
 - (b) (0.5 ponto) Quantos são clientes da empresa há mais de um ano?
7. Dentre as pessoas que cancelaram o contrato:
 - (a) (0.5 ponto) Qual o número de clientes casados e que possuem dependentes?
 - (b) (0.5 ponto) Qual o número de clientes para cada tipo de contrato (“Month-to-month”, “One year” e “Two year”)?
 - (c) (0.5 ponto) Qual o número de clientes que já haviam pago pelo menos dois anos de faturas?
 - (d) (0.5 ponto) Qual seria o desconto acumulado em um ano para cada um desses clientes se fosse oferecido um desconto de 5% na fatura mensal?

Observação: para esse problema, as suas respostas devem ser feitas no arquivo “`teste1b.R`”.

Considerações Finais

- A utilização de qualquer função não vista em sala de aula, ou seja, que não encontra-se nos slides fornecidos no Moodle, acarretará em um desconto de 50% da nota daquele item.
- Você não deve remover qualquer linha já existente nos arquivos.
- Respostas com informações não requisitadas pelo enunciado serão penalizadas.
- Para testes realizados em dupla, apenas um membro da dupla deve enviar a solução. Os nomes dos membros devem constar no cabeçalho de cada arquivo “.R”.
- Salve os arquivos utilizando o mesmo nome, e os envie no sistema Moodle, clicando no link “**Teste 1**” da Seção “**Avaliações**”. Clique em “**Adicionar tarefa**”, anexe os arquivos e, por fim, clique em “**Salvar mudanças**”. Você voltará para a tela da atividade e deverá constar o status “**Enviado para avaliação**”. A qualquer momento, antes do prazo final de submissão, você pode alterar sua submissão clicando em “**Editar envio**”.

Prazo de entrega: 06 de março de 2019 (Quarta-Feira), até às 23h55.

Forma de entrega: via sistema Moodle:

- <https://moodle.lab.ic.unicamp.br/moodle/course/view.php?id=336>

Pontuação: Este teste será pontuado de 0 a 10, e corresponderá 30% da nota final.