



TESTE 2 (INDIVIDUAL OU EM DUPLA) INF-0612 – ANÁLISE DE DADOS

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

1 Distribuidora

A distribuidora “X” é bastante reconhecida no mercado e fornece produtos para supermercados em diversas cidades. No arquivo “teste2a.R” você encontra os vetores “C”, “L” e “V” representam os produtos distribuídos por “X” nas cidades de Campinas, Limeira e Vinhedo, respectivamente.

O distribuidora “X” precisa da sua ajuda para responder algumas perguntas e elaborar um plano estratégico de vendas. Perguntas:

- (0.25 ponto) Quais os produtos que são vendidos em Campinas, mas não são vendidos em Limeira?
- (0.25 ponto) Quais os produtos que são vendidos em Vinhedo, mas não são vendidos em Campinas?
- (0.5 ponto) Quais os produtos que são vendidos em pelo menos uma cidade?
- (0.5 ponto) Quais os produtos que são vendidos em todas as cidades?
- (1.0 ponto) Se a filial de Limeira parar a distribuição de produtos, a filial de Campinas possui todos os itens necessários para atender a demanda de Limeira?

2 Média Final

Foi solicitado a você calcular a média final de cada aluno. No arquivo “teste2b.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 deverá ser calculada utilizando a média quadrática das três maiores notas que o aluno obteve nas provas, ou seja, a menor nota obtida pelo aluno deve ser desconsiderada.

A média quadrática entre n valores x_1, x_2, \dots, x_n é calculada a partir da fórmula:

$$\sqrt[2]{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

Por exemplo, Joãozinho obteve as notas 8.3, 9.0, 5.2 e 6.7. Utilizando as três maiores notas a média de Joãozinho será:

$$\sqrt[2]{\frac{8.3^2 + 9.0^2 + 6.7^2}{3}} = 8.06$$

Pontos obrigatórios (a não observância destes itens acarretará em nota 0 naquele item):

- (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).
- (1.0 ponto) Você deve salvar no vetor “medquad” a média final de cada aluno (média quadrática das provas, desconsiderando a menor nota obtida), com duas casas decimais.
- (0.5 ponto) Você deve salvar nas variáveis “mp1”, “mp2”, “mp3” e “mp4” a média aritmética das notas das provas 1, 2, 3 e 4, respectivamente.

- (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.
- (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.

3 Chuvas

Um estagiário fez um programa que armazena para cada dia a quantidade de chuva acumulada em cada um dos sensores existentes nas cidades de Campinas, Limeira e Vinhedo. Após os 15 dias do mês de abril, seu chefe percebeu que o acumulado de chuva estava muito maior do que deveria, e, ao olhar o arquivo, percebeu que o programa do estagiário salvava o acumulado de chuvas em cada um dos sensores mais de uma vez ao dia. Por exemplo, para o dia 03, no sensor de Campinas, existiam três informações: a primeira era de que o acumulado de chuva era de 3mm, a segunda que o acumulado era de 4.3mm e a terceira que o acumulado era de 5.2mm. Claramente, o acumulado para o dia 3 em Campinas deveria ser apenas a informação com o maior valor, neste caso de 5.2mm. Por sorte, para um mesmo dia e mesma cidade, a última observação no arquivo era sempre a maior (e, portanto, deveria ser a única existente). Você foi consultado então para corrigir o arquivo, e fornecer algumas estatísticas sobre as chuvas na região.

Você encontrará no arquivo “teste2c.R” três vetores de mesmo tamanho: o vetor “dia”, contendo o dia da informação, o vetor “cidade” contendo as cidades de onde é aquela informação, podendo ser Campinas, Limeira ou Vinhedo, e o vetor “chuva” com o acumulado de chuva daquela observação. Por exemplo, se o arquivo contém as seguintes linhas:

Dia	Cidade	Chuva
01	Campinas	1.3
01	Limeira	0.4
02	Campinas	0.2
02	Vinhedo	0.1
02	Vinhedo	1.1
02	Campinas	0.8
03	Limeira	0.1
03	Campinas	0.2

Devem ser removidas as linhas 3, pois a linha 6 contém o mesmo dia e mesma cidade, e a linha 4, pois a linha 5 contém o mesmo dia e mesma cidade.

Pontos obrigatórios (a não observância destes itens acarretará em nota 0 naquele item):

- (0.5 ponto) Você deve criar um data frame utilizando os vetores fornecidos e, sempre que utilizar algum dado destes vetores, referir-se apenas a este data frame (ou seja, você só pode utilizar os vetores fornecidos para criar este data frame).
- (1.25 ponto) Você deve remover do dataframe todas as linhas i tais que exista uma linha j com $j > i$ e que os campos contendo dia e cidade sejam o mesmo em i e j .
- (0.75 ponto) Salve nas variáveis “acumCamp”, “acumLim” e “acumVin” o total de chuvas observados nos 30 dias nas cidades de Campinas, Limeira e Vinhedo, respectivamente.
- (0.75 ponto) Você deve salvar nas variáveis “dmaxCamp”, “dmaxLim” e “dmaxVin” o dia do mês que mais choveu nas cidades de Campinas, Limeira e Vinhedo, respectivamente. Se existir mais de um dia com o valor máximo, você pode escolher qualquer um dos dias.
- (0.75 ponto) Você deve salvar nas variáveis “dminCamp”, “dminLim” e “dminVin” o dia do mês que menos choveu nas cidades de Campinas, Limeira e Vinhedo, respectivamente. Se existir mais de um dia com o valor mínimo, você pode escolher qualquer um dos dias.

Considerações Finais

- A utilização de qualquer função não vista em sala de aula, ou seja, que não encontram-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.
- 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 2” 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: 12 de março de 2018 (segunda-feira) às 23:55hs.

Forma de entrega: via sistema Moodle:

- <https://moodle.lab.ic.unicamp.br/moodle/course/view.php?id=243>

Pontuação: Este teste será pontuado de 0 a 10, e corresponderá 15% da nota final.