

Nome: Ana Paula Pudo

Ao entregar este trabalho final no e-class Você estará declarando:

"Estou ciente das razões pelas quais estamos fazendo este trabalho de forma remota. Assim, declaro que utilizei somente formas éticas de resolver as questões e não consultei ninguém, de nenhuma forma, para respondê-las "

Instruções:

INSTRUÇÕES:
Escrevi para que Você leia, por favor!!!
A não observação das instruções será penalizada

- Este trabalho tem **5 páginas**, incluindo esta.
- Cada questão tem vários itens. Todos os itens tem o mesmo valor. A resposta a cada questão só será considerada se estiver completamente correta (não há “meio certo”)
- A prova tem que ser resolvida utilizando o software R.
- Consulta a notas de aula, apontamentos, net, scripts etc. é permitida.
- **A consulta a colegas ou outras pessoas, qualquer que seja o meio, é proibida.**
- Responda as perguntas digitando as respostas neste arquivo, nos locais indicados. Se não for incômodo, coloque as respostas em vermelho ou azul ou verde.
- Ao terminar, **não esqueça de colocar seu nome**. Verifique cuidadosamente se não esqueceu nenhuma resposta nos espaços indicados.
- **Salve como .pdf** e verifique se não prejudicou a formatação. Faça upload do arquivo em .pdf no ECLASS na caixa de **entrega de atividades**, na pasta correspondente. **Trabalhos entregues no formato .doc não serão aceitos**
- Não cole os scripts do R, os outputs do R nem as tabelas de dados originais (a menos que seja solicitado).
- A interpretação das questões é parte da prova. Não responderei dúvidas a respeito das mesmas.
- A data de entrega é **até as 23:00 horas do dia 26/11/2021** (dentro do prazo regulamentar fixado pela FGV).
- Não envie as respostas por e mail ou WhatsApp, por favor (**não receberei**).

Questão 1: Considere as variáveis quantitativas da planilha ECONOMIA (arquivo em EXCEL enviado em conjunto com este Word).

inflacao	Inflação (% em 12 meses)
desemprego	Desemprego (% força de trabalho)
PIB	PIB
reservas	Reservas (US\$ bilhões)
juros	Juros (% ano. taxa básica)

Padronize-as utilizando subtraindo a média e dividindo pelo desvio padrão correspondentes de cada variável. Utilize as variáveis padronizadas como drivers para agrupar as observações (países)

Não elimine observações mesmo que lhe pareçam outliers.

Não transforme as variáveis (exceto a padronização solicitada)

Não elimine variáveis mesmo que correlacionadas com outras

Não divida a base de dados em duas partes.

1.a) Qual a correlação entre *desemprego* e *inflacao*?

Resp: **0,322** (3 casas decimais)

1.b) Utilize a matriz de distâncias calculada com a função **dist**. Utilizando o método hierárquico de Ward determine K=3 clusters. Mostre quantos elementos tem cada cluster que Você determinou. Dê o número de elementos que cada cluster contém.

Cluster 1 elementos=**10** Cluster 2 elementos= **2**

Cluster 3 elementos= **1**

1.c) Quais os países que compõem o menor cluster obtido pelo método de Ward.D2 (em caso de empate, selecione um deles. Diga qual.)? **Japão**

1.d) Quais as médias (**valores não padronizados**) da variável *juros* nos diferentes clusters? .

Cluster 1 =**12,60** Cluster 2 = **51,70** Cluster 3 = **0.21** (2 casas decimais)

Continua na página seguinte

1.e) Com auxílio de boxplots compare os clusters. Qual apresenta maiores valores de PIB, sem considerar eventuais outliers? (não cole a figura)

Maiores valores de PIB → **cluster 1**. Quantos países há neste cluster? **10**

1.f) Utilizando o método k medoids, determine o melhor número de clusters sugerido pela função `pamk` fazendo o `k` variar de 2 a 4 e adotando como critério a medida “asw”. Utilize a matriz de distâncias calculada com a função `dist`. Rode utilizando o `set.seed(123)` e diga quantos clusters encontrou.

Número de clusters com `set.seed(123)` **2**

Tamanhos dos clusters obtidos: **11 e 2**

Continua na página seguinte

Questão 2: Considere a planilha ECONOMIA.

Determine as componentes principais das variáveis quantitativas após padronizá-las (utilizando a opção `scale=T`). Selecione o **menor** número de componentes que expliquem **pelo menos 90%** da variância total.

2.a) Quantas componentes principais Você selecionou? Resp: **3**

2.b) Qual a proporção da variância explicada por essas componentes? Resp: **94,38%** (em porcentagem, com 2 casas decimais)

2.c) Qual a correlação da variável *juros* com a terceira componente principal? Resp: **-0,356** (3 casas decimais)

2.d) Seja R matriz de correlações das variáveis quantitativas. Determine a matriz inversa de R. Qual o primeiro elemento (1ª linha, 1ª coluna) da diagonal principal da matriz inversa?
Resp: **10,0465** (4 casas decimais)

2.e) Qual o maior auto valor da matriz de correlações R?
Resp: **2,758** (3 casas decimais)

2.f) Qual a correlação entre o 1º e o 3º autovetor de R? Resp: **-0,08** (2 casas decimais)

Continua na página seguinte

Questão 3: Considere a seguinte matriz binária de transações (consta do arquivo em Excel da prova como MKT). Para facilitar a leitura, coleí apenas os valores iguais a 1 (No Excel da prova a matriz está completa)

TID	FRALDAS	CERVEJA	LEITE	PÃO	QUEIJO	SALGADOS
1	1	1	1			
2	1			1	1	
3	1	1	1			1
4				1	1	1
5		1				1
6	1		1	1	1	
7			1	1	1	
8		1	1			
9	1		1	1	1	1
10		1				1
11			1	1	1	1
12		1	1			
13	1		1	1	1	1
14		1		1		1

3.a) calcule (utilizando as fórmulas) o suporte, a confiança e o lift para a seguinte regra: **{ leite, fraldas} → {cerveja}** (dê as respostas com 2 casas decimais)

Suport **0,36** confiança=**0,40** lift= **1,14**

3.b) utilize o pacote **arules** para determinar as regras de associação com suporte superior a 10% e confiança superior a 70%. Considere apenas regras com **2 a 4** itens. Não esqueça de eliminar a coluna 1 (TID). Quantas regras foram geradas? Resp: **29**

3.c) verifique no output o suporte, a confiança e o lift para a regra de associação seguinte

{FRALDAS,PÃO,QUEIJO} => {LEITE}

Suporte= **0,214** confiança= **0,750** lift= **1,167**

(3 casas decimais)

3.d) Eliminando as regras redundantes, quantas regras não redundantes sobraram? Resp: **12** Das regras de associação não redundantes, qual a que possui menor lift? Dê a confiança dela. **{QUEIJO} → {LEITE} Conf = 0,714 Lift = 1,111** .

Continua na página seguinte