

INDICAÇÕES PRÁTICAS

1. O relatório (.pdf), o *script* em R (.Rdata) e a base de dados (*.csv, *.xls) deverão ser submetidos no moodle até ao **dia 18 de março de 2024, 23h59**, na secção do moodle “Trabalho Época de Recurso” numa pasta comprimida (*.zip, *.rar, etc.)
2. O ficheiro (*.pdf) com a resolução do trabalho deverá conter:
 - a. Capa com identificação dos estudantes.
 - b. Identificação das perguntas, seguida das respostas. Cada resposta deverá ser justificada enunciando os métodos usados, não deixando de referir os pressupostos necessários para a utilização desses métodos; a resposta às questões deverá ser acompanhada de tabelas e/ou gráficos sempre que se justifique; análise e conclusão.
3. A resolução do trabalho deverá ser realizada com recurso ao *software* R.
4. O ficheiro (*.Rdata) com o *script* deverá estar devidamente comentado de modo que cada questão seja identificável.
5. A base de dados deverá estar no formato (*.csv,*.xls) com o título das variáveis na primeira linha.
6. O nome dos ficheiros e da pasta deverá conter a indicação do 1º e último nomes dos estudantes que compõe o grupo e o número do grupo (ex: Grupo1_nomeApelido1_nomeApelido2_nomeApelido3.Rdata).
7. O trabalho tem uma ponderação final de 100% na nota final da unidade curricular.
8. Sempre que necessário utilize um nível de significância de 5%.
9. **Selecione aleatoriamente 20% dos indivíduos** da BD BD_Epoca_Recurso_DIOS_2024.xls e guarde a nova DB num ficheiro *.xls. Esta será a nova BD que irão usar para a resolução do trabalho e que deverá ser submetida juntamente com o projeto e *script* do R.

CRITÉRIOS DE AVALIAÇÃO

Critérios de Avaliação	Descrição
Processamento de Dados no RStudio	Manipulação dos dados, conversão de tipos de dados e processamento.
Exploração de Dados	Utilização eficaz de visualizações para explorar e apresentar distribuições e relações entre variáveis. Análise correta e interpretação de padrões identificados nos dados.
Análise Estatística Descritiva	Aplicação correta de técnicas de estatística descritiva para resumir e caracterizar os dados. Uso adequado de medidas de tendência central, dispersão e outras estatísticas descritivas relevantes.
Estimação e Testes de Hipóteses	Correta aplicação de técnicas de estimação e testes de hipóteses. Interpretação apropriada dos resultados, incluindo conclusões sobre significância estatística.
Análise de Correlação e Associação	Uso adequado de métodos estatísticos para análise de correlação. Interpretação correta das associações identificadas entre variáveis.
Implementação no RStudio	Código limpo, bem comentado e estruturado no ambiente RStudio. Utilização eficaz de bibliotecas e funções do R para realizar análises estatísticas.
Comunicação dos Resultados	Relatórios claros e bem estruturados que incluam interpretações significativas dos resultados. Uso eficiente de gráficos e visualizações para comunicar.
Inovação e Criatividade	Introdução de abordagens inovadoras ou criativas na análise, destacando originalidade na exploração dos dados.

ENUNCIADO

A vinculação entre os dados de prescrição e dispensação de medicamentos pode ser feita por meio do sistema de informação de gestão regional de saúde - SIARS, que é um banco de dados administrativo, desenvolvido para facilitar a análise e a monitorização da atividade das unidades de cuidados primários. Esse sistema automatizado inclui informações sobre diagnósticos e registos nas unidades de cuidados de saúde primários, bem como dados demográficos e administrativos dos pacientes.

Nesta base de dados encontra-se informação relativa à adesão à terapia de hipertensão arterial (AHT) em pacientes hipertensos recém-tratados em unidades de cuidados de saúde primários.

Foram identificados todos os pacientes que foram diagnosticados com hipertensão e receberam uma primeira prescrição (prescrição de índice) para pelo menos um medicamento AHT entre março e abril de 2019.

VARIÁVEIS DA BASE DE DADOS:

1. Idade
2. Sexo: masculino / feminino
3. inicio: Doente inicia tratamento (sim ou não)
4. ICPC: código da hipertensão: K86 – hipertensão sem complicações / K87 – hipertensão com complicações
5. tempo_inicio: Tempo até à primeira aquisição do medicamento
6. grupo_terap: Grupo terapêutico. As classes de drogas AHT avaliadas, com os códigos do sistema de classificação química terapêutica anatómica (QTA) correspondentes foram:
1 = C02: Anti-hipertensivos
2 = C03: Diuréticos
3 = C07: agentes bloqueadores beta
4 = C08: bloqueadores do canal de cálcio
5 = C09: Agentes que atuam no sistema renina-angiotensina
6 = 2 ou mais
7. nr_receita: número de medicamentos na 1ª receita (1-um; 2-dois ou mais)
8. marca_generico: Comprou a marca ou o genérico (0-genérico; 1-marca)
9. Poder_compra: poder de compra (dados segundo a PORDATA)
10. Custo_inicial: custo da 1ª receita
11. Descontinua: O doente descontinua prematuramente: 0=não, 1=sim
12. Persist_6m: persistência ao fim de 6 meses: 0=não, 1=sim
13. Persits_24m: persistência ao fim de 24 meses: 0=não, 1=sim
14. Descontinua2: doente descontinua ao fim de 2 anos: 0=não, 1=sim
15. Custo_24: custo da receita passada 24 meses depois

Observação: o valor 999 na BD corresponde a um valor omissos. A análise deverá ter em consideração a remoção destes valores, pois caso contrário irão ser considerados nos diferentes cálculos. No R funções como `na.rm = TRUE`, `na.omit()`, servem para lidar com os valores omissos, porém no R os valores omissos são identificáveis através de NA.

Questões

1. Classifique as variáveis da BD (tipo e escala de medição).
2. Caracterize as variáveis **Idade**, **Sexo**, **grupo_terap**, **nr_receita** recorrendo a gráficos e/ou estatísticas de acordo com as características das variáveis e analise os resultados.
3. Recorrendo a gráficos, compare o tempo até à primeira aquisição (**tempo_inicio**) por grupo terapêutico (**grupo_terap**).
4. Obtenha uma representação gráfica que permita comparar a compra de medicamentos genéricos ou marca (**marca_generico**) por sexo e interprete-o.
5. Construa um diagrama de dispersão entre o poder de compra (**Poder-compra**) e o custo inicial (**custo_inicial**) e conclua. Complemente com uma estatística que considere adequada.
6. Verifique se existe associação entre a variável “início” e o código ICPC.
7. Analise as variáveis **descontinua**, **persist_6m**, **persist_24m**, **descontinua2**, recorrendo a estatísticas e/ou gráficos e conclua sobre a adesão à terapêutica dos pacientes.
8. Obtenha uma estimativa pontual para o valor médio do poder de compra (**Poder_compra**) e uma estimativa intervalar com uma confiança de 99% apenas para os indivíduos que aviam dois ou mais medicamentos.
9. Compare a estimativa por intervalo de confiança a 95% para o mesmo parâmetro da alínea anterior e conclua.
10. Considerando apenas os indivíduos com idade inferior ou igual a 45 anos, pode afirmar-se que o valor médio dos valores obtidos para o custo inicial (**Custo_inicial**) é significativamente superior a 3?
11. Verifique se existem diferenças significativas entre o valor médio do poder de compra (**Poder_compra**) entre homens e mulheres.
12. Utilizando a base de dados de que dispõe proceda a uma análise estatística suplementar elaborando 3 questões e tire conclusões.

Carina Silva