

Case sobre Depressão

Habilidades:

Analistas: análise exploratória, estatística, excel ou python

Cientistas: análise exploratória, estatística, python

[ANALISTA DE DADOS] [CIENTISTAS DE DADOS]

[JÚNIOR][PLENO][SÊNIOR]

Analistas, sugiro Excel, ou, preferencialmente, Python

Cientistas, é mandatório fazer em Python

Resumo

- O desafio proposto trata sobre o transtorno depressivo, um problema multifatorial que pode ter várias possíveis combinações de situações como causa.
- Os dados disponíveis são resultados da NHNES (National Health and Nutrition Examination Survey), realizada anualmente nos EUA para avaliar a **saúde e nutrição** de adultos e crianças;
- Seu desafio é responder as seguintes perguntas:
 - Qual o perfil de indivíduos (adultos maiores de 18 anos) com sintomas depressivos nos EUA no período de 2005-2006?
 - Hábitos saudáveis de alimentação e atividade física estão associados a menores índices de depressão nesta população?

Bases de dados

Abaixo você tem acesso aos bancos de dados apresentados pela:

https://drive.google.com/file/d/1RtMi9PdKOnVASGP_2285sPcUDUnS9cEH/view?usp=s_haring

https://drive.google.com/file/d/1pXHL4-aUGgQshj6DeglQQcmXLYxkQkPln/view?usp=s_haring

Dicionários das bases

DEMO_PHQ.csv: banco de dados contendo 5334 observações de adultos pesquisados no NHANES 2005-2006:

<https://drive.google.com/file/d/1fiqlsCDDrCGFLujKKV5yRnu2pwTG2P-0/view?usp=sharing>

PAG_HEI.csv: banco de dados contendo 9424 observações de crianças e adultos pesquisados no NHANES 2005-2006:

https://drive.google.com/file/d/1B0tXh0EvFW2nWH2XXVEgWt_61MWx1UXo/view?usp=sharing

Perguntas a serem respondidas

- 1) Qual o perfil de indivíduos (adultos maiores de 18 anos) com sintomas depressivos nos EUA no período de 2005-2006?
- 2) Hábitos saudáveis de alimentação e atividade física estão associados a menores índices de depressão nesta população?

Sugestão

Esse case não terá uma resposta, mas segue aqui uma sugestão do que fazer

• TRATAMENTO DOS DADOS

- Faça a leitura do banco e os tratamentos que achar necessários nos dados;
- Combine os dois bancos e lide com possíveis dados faltantes. O esperado é um banco final com 5334 informações;
- Crie a variável `phq9`, correspondente ao score do questionário PHQ-9, somando as variáveis `DPQ010` a `DPQ090`.
- Crie a variável `phq_grp` com a classificação do score `phq9`, assumindo os valores:
 - 0 ("sem sintomas") se `phq9 < 5`
 - 1 ("sintomas leves") se `5 <= phq9 < 10`
 - 2 ("sintomas moderados a severos") se `phq9 >= 10`

• ANÁLISE EXPLORATÓRIA (EDA) UNIVARIADA

- Identificar corretamente quais variáveis são qualitativas e quais são quantitativas;
- Para medidas quantitativas, podemos usar medidas de posição e dispersão e para gráficos como histogramas ou de densidade;
- Para medidas qualitativas, podemos olhar para as frequências absolutas e relativas, ou olhar para um gráfico de barras;

- Avalie dados faltantes, reagrupe ou os re-codifique.
- **ANÁLISE EXPLORATÓRIA (EDA) BIVARIADA**
 - Para avaliar relação entre duas variáveis numéricas use um gráfico de dispersão;
 - Para avaliar relação entre duas variáveis categóricas use um gráfico de barras;
 - Para avaliar relação entre variáveis numéricas e variáveis categóricas use um gráfico boxplot;
 - Avalie a variável dependente versus todas as variáveis independentes graficamente
- **TESTES DE HIPÓTESES**
 - Para avaliar a relação entre duas variáveis numéricas use teste de correlação de Pearson;
 - Para avaliar a associação entre duas variáveis categóricas use Teste qui-quadrado de independência;
 - Para comparar as médias de dois grupos independentes use Teste-t independente;
 - Para comparar as médias de mais de dois grupos independentes use Teste F (ANOVA).
- **ANÁLISE CRÍTICA DOS RESULTADOS**
 - Pense criticamente sobre os resultados que você observou
 - Quais vieses podem existir nos dados e quais as limitações das análises?
 - O que de fato podemos inferir pensando em efeitos causais?
 - Quais são as limitações das análises realizadas?
O que pode ser feito para melhorar?
 - Quais outras variáveis/informações poderiam ter sido coletadas para esta análise?

Se acharem que cabe, usem também regressões!