

Módulo 2. Estatística para análise de dados

Semana 5. Estatística - Projeto do módulo

Bem-vindo(a) à quinta semana deste módulo.



[https://www.freepik.com/free-vector/people-giving-high-five-theme-illustration_6600419.htm#query=hello&position=23&from_view=search]

Nesta semana, vocês devem realizar um desafio em que poderão usar todas as habilidades aprendidas no módulo. O desafio pode ser realizado em Python ou Excel e vocês podem agregá-lo ao portfólio de vocês!

Instruções



[https://www.freepik.com/free-vector/choose-concept-illustration_5784488.htm#query=probability&position=5&from_view=search]

- 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:

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

<https://drive.google.com/file/d/1pXHL4aUGgQshj6DeglQQcmXLYxkQkPln/view?usp=sharing>

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>

Através dos indicadores registrados na tabela, o PHQ9 é calculado como:

$$\text{PHQ9} = \text{DPQ010} + \text{DPQ020} + \text{DPQ030} + \text{DPQ040} + \dots + \text{DPQ080} + \text{DPQ090}$$

Dessa forma, os sintomas de saúde mental são classificados da seguinte forma:

PHQ9	Classificação
0 - 4	Sem sintomas
1 - 9	Sintomas leves
10 - 14	Sintomas moderados
15 - 19	Sintomas moderadamente severos
20 ou mais	Sintomas severos

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

Sugestão

- **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;
- Monte o score que resume a informação do questionário PHQ-9.
- **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, reagrpe 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?