

PROJETO - MÓDULO ESTATÍSTICA I

Este projeto busca avaliar os conhecimentos sobre estatística descritiva, probabilidade e estatística inferencial.

Para isso, vocês devem utilizar os seguintes datasets:

- **Squad 1 - Lufa Lufa:** Covid-19: https://brasil.io/dataset/covid19/caso_full/
- **Squad 2 - Corvinal:** NBA Players: <https://www.kaggle.com/datasets/justinas/nba-players-data>
- **Squad 3 - Sonserina:** BTC Dataset: <https://www.cryptodatadownload.com/data/poloniex/>
- **Squad 4 - Grifinória:** Ticket de Cinema: https://www.kaggle.com/datasets/arashnic/cinema-ticket?select=cinemaTicket_Ref.csv

v

ESTATÍSTICA DESCRITIVA

Cada grupo receberá um conjunto de dados e deverá fazer uma análise do tipo de cada variável e apresentá-las graficamente e usando as medidas de resumo para explicar a concentração dos dados, quais são os valores com maiores frequências, sua distribuição e como cada variável correlaciona-se com a outra.

Além disso, o aluno deve fazer uma análise sobre dados anômalos (os *outliers*) e dizer, em cada contexto, qual seria a melhor opção a ser tomada (manter os dados ou descartá-los).

PROBABILIDADE E INFERÊNCIA

Você está trabalhando em um projeto de Ciência de Dados e, para isso, está coletando dados de sensores para a construção de uma base de dados. Esses dados estão armazenados em um arquivo no formato csv. A princípio, não temos informações sobre as suas distribuições; mas é interessante conhecer qual distribuição melhor representa a nossa variável. Para isso, iremos testar se os nossos dados pertencem a uma distribuição exponencial, normal, lognormal ou gamma.

a) Para cada uma das distribuições, pesquise um estimador para o parâmetro θ . Em seguida, escreva um código para a estimação desses parâmetros.

b) Faça o teste do qui-quadrado para verificar qual distribuição teórica mais se assemelha à distribuição dos dados. Não esqueça de montar o teste de hipótese.

Séries temporais geralmente contém informações muito importantes para a análise humana. Considere a série temporal selecionada para o grupo. Com base nas aulas, faça uma análise completa sobre o comportamento da série temporal.

Questionamentos e sugestões para motivação da análise:

Mostre graficamente o comportamento da série temporal. Realce os valores extremos (mínimo e máximo). Existem registros atípicos?

Quais são os componentes da série temporal? Analise tais componentes e plote os gráficos.

Os resíduos da série seguem uma distribuição normal?