

# Desafio Inferência

Convidamos a expert [Fernanda Peres](#), Professora e Consultora na área de Estatística, para propor um **Desafio de Inferência**.

Ela vai apresentar um problema, um possível roteiro de solução e compartilhar como ela abordaria esse desafio. Lembre-se: a resolução proposta **não é a única possível** e, portanto, **não existe solução correta**. O problema pode ser interpretado de diferentes maneiras e diferentes conclusões podem ser adquiridas ao longo das análises.

**Use sua criatividade!**

Antes de começar, entenda quais são os objetivos de aprendizagem deste desafio.

## Objetivos de Aprendizagem

- Aplicar conhecimentos de estatística descritiva;
- Aplicar conhecimentos de estatística inferencial;
- Extrair conclusões relevantes a partir de uma análise crítica dos resultados obtidos.

# ENTENDENDO O PROBLEMA

Neste capítulo, Fernanda apresenta as perguntas que queremos responder e que informações temos disponíveis para resolver o problema proposto.

## 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?

## Baixe as bases de dados

Abaixo você tem acesso aos bancos de dados apresentados pela expert:

<https://drive.google.com/drive/folders/1N4sCFSZDXLYzccP-9t0JXGkF06wiAQS-?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/1a4bGPp-ZY2zZlqtKf\\_b\\_E9cPwiFAUznj/view?usp=sharing](https://drive.google.com/file/d/1a4bGPp-ZY2zZlqtKf_b_E9cPwiFAUznj/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/1Qc63Z0t5v7zHKUzszWOQIHVCwBWG1Ps7i/view?usp=sharing>

## SUGESTÃO DE ROTEIRO

Agora Fernanda propõe uma sugestão de roteiro para essa resolução. Lembre-se: essa sugestão **não é o único caminho possível** e pode ser adaptada como você achar mais conveniente para responder às perguntas definidas.

### Resumo



#### ● 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, 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;
- **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?

## **MÃO NA MASSA!**

Agora você tem todos os dados necessários e a sugestão de roteiro indicada pela expert é a sua vez de tentar executar esse desafio. Tente avançar o máximo possível na sua análise antes de seguir para o próximo capítulo!

## **RESOLUÇÃO DO DESAFIO**

Neste capítulo, Fernanda traz uma possível solução para o desafio. Note que as análises são realizadas em Python, mas essa não é uma obrigatoriedade para o desafio. Você pode utilizar outras ferramentas ou linguagem de programação para resolução.

## Resumo

- **LEITURA DO BANCO E TRATAMENTO DOS DADOS**
  - Importe as bibliotecas necessárias para as análises e leia os bancos de dados
  - Observe as variáveis no banco e decida qual **Join** usar.
  - Na sugestão foi atribuído como valores faltantes as respostas respectivas a informações ausentes;
  - Na variável RIDRETH1, os valores “outros” foram agrupados;
  - A variável INDFMINC foi resumida com as médias de suas categorias;
  - Trate os dados faltantes e construa o score de PHQ-9
- **ANÁLISE EXPLORATÓRIA (EDA)**
  - Observe as estatísticas sumárias das variáveis utilizando *.describe()* e se elas fazem sentido;
  - Observe as distribuições das variáveis numéricas usando histogramas;
  - Observe as frequências observadas para as variáveis qualitativas usando gráficos de barra para comparar proporções;
- **ANÁLISE EXPLORATÓRIA (EDA) BIVARIADA**
  - Use box plot para cruzar informações como PAG\_WH e phq\_grp2;
  - Use gráficos de barras agrupadas para comparar a proporção de variáveis como sexo;
- **TESTES DE HIPÓTESES**
  - Para comparar os **três** níveis de depressão em relação a idade ou tempo de exercício físico realizado na semana use um teste F(ANOVA);
  - Caso o teste F indique alguma diferença, use um pos-hoc de tukey para verificar quais deles diferem entre si;
  - Para comparar os níveis de depressão em relação a outro agrupamento como sexo, use um teste qui-quadrado de independência.

## Dicas Finais

Para ajudar a consolidar as informações mais relevantes da aula, a Fernanda preparou uma lista de "takeaways", informações e dicas que você precisa levar com você.

## Lembre-se

- Não há uma única solução;
- Há várias respostas e caminhos possíveis;
- Há associação entre os sintomas depressivos, gênero e realização de exercícios físicos
- Não podemos fazer inferências causais - estudo observacional;
- A ideia é que tenha sido desafiador!;
- O desafio é diferente para cada um (estamos em níveis diferentes, e tudo bem!).

## Você concluiu este desafio

Que bom que você chegou até aqui.

Este é só o início do seu desenvolvimento abordando desafios que envolvem inferência.

Este desafio ficará disponível para você rever e praticar seus conhecimentos quando quiser.

## Checklist

Você deve sair desta aula sabendo:

- Aplicar conhecimentos de estatística descritiva;
- Aplicar conhecimentos de estatística inferencial;
- Extrair conclusões relevantes a partir de uma análise crítica dos resultados obtidos.