# **DESAFIO INFERÊNCIA**

## **Por Tamy Tsujimoto**

## 

## **Desafio 1 - inferência estatística**

([**https://www.youtube.com/watch?v=hTi5U2\_5pJ4**](https://www.youtube.com/watch?v=hTi5U2_5pJ4))

## **Depressão e Hábitos Saudáveis**

O transtorno depressivo representa um grande problema de saúde pública e é apontado como uma das principais causas de doenças e debilitações segundo a Organização Mundial da Saúde (OMS). A depressão resulta de uma interação complexa de fatores sociais, psicológicos e biológicos e, embora o acesso a diagnósticos e tratamentos esteja aumentando, muitos ainda não possuem acesso ao controle adequado dos sintomas depressivos.

Para garantir um estilo de vida saudável, a OMS recomenda a prática regular de exercícios, bem como a adoção de uma dieta saudável rica em alimentos in natura e com reduzido teor de gordura saturada, sal e açúcares refinados.

Neste desafio, utilizaremos dados de uma pesquisa realizada anualmente nos Estados Unidos para avaliar (1) qual o perfil de indivíduos (adultos maiores de 18 anos) com sintomas depressivos nos EUA no período de 2005-2006, e (2) se hábitos saudáveis de alimentação e atividade física estão associados a menores índices de depressão nesta população.

## **National Health and Nutrition Examination Survey**

O National Health and Nutrition Examination Survey (NHANES) é uma pesquisa anual conduzida pelo National Center for Health Statistics (NCHS) do Centro de Controle e Prevenção de Doenças (Centers for Disease Control - CDC) para avaliar a saúde e nutrição de adultos e crianças dos Estados Unidos. Dados coletados incluem questões demográficas, socioeconômicas, dietéticas e relacionadas à saúde, com o componente de exame contendo medidas médicas, odontológicas, fisiológicas e exames laboratoriais.

A pesquisa examina uma amostra de cerca de 5.000 pessoas a cada ano, selecionada à partir de amostragem complexa a fim de selecionar uma amostra representativa da população civil não institucionalizada dos EUA. Sendo assim, as análises utilizando este estudo devem ser realizadas utilizando técnicas e ferramentas que levem em conta a amostragem complexa.

Neste desafio, no entanto, iremos assumir que os dados foram obtidos usando uma amostra aleatória da população de interesse e utilizaremos técnicas e ferramentas de análise usuais para amostras aleatórias para fins didáticos.

## **Patient Health Questionnaire-9 (PHQ-9)**

O Patient Health Questionnaire-9 (PHQ-9) é um instrumento utilizado para avaliar o grau de depressão em pacientes. O questionário consiste de 9 itens em que os respondentes indicam a frequência (0 = “nenhuma vez”, 1 = “menos de uma semana”, 2 = “uma semana ou mais” e 3 = “quase todos os dias”) de sintomas de depressão nas duas últimas semanas.

O PHQ-9 inclui os seguintes itens para a pergunta “Nas últimas 2 semanas, com que frequência você ficou incomodado por algum dos problemas a seguir?” (0 = “nenhuma vez”, 1 = “menos de uma semana”, 2 = “uma semana ou mais” e 3 = “quase todos os dias”):

1 - Pouco interesse ou pouco prazer em fazer as coisas

2 - Se sentiu para baixo, deprimido(a) ou sem perspectiva

3 - Dificuldade para pegar no sono ou permanecer dormindo ou dormiu mais do que o costume

4 - Se sentiu cansado(a) ou com pouca energia

5 - Falta de apetite ou comeu demais

6 - Se sentiu mal consigo mesmo(a) ou achou que é um fracasso ou que decepcionou sua família ou a você mesmo(a)

7 - Dificuldade para se concentrar nas coisas (como ler o jornal ou ver televisão)

8 - Teve lentidão para se movimentar ou falar (a ponto de outras pessoas perceberem), ou ao contrário, esteve tão agitado(a) que você ficava andando de um lado para o outro mais do que costume

9 - Pensou em se ferir de alguma maneira ou que seria melhor estar morto(a)

O escore total é calculado à partir da soma dos itens 1-9 e varia de 0 a 27, em que maiores valores do escore indicam maiores frequências de sintomas de depressão. Aqueles com pontuação maior ou igual a 5 para o escore total de PHQ-9 são considerados como tendo sintomas leves (5-9), moderados (10-14), moderadamente severos (15-19) e severos de depressão (>= 20).

# **HEALTHY EATING INDEX - (HEI)**

O Healthy Eating Index (HEI) é uma medida de qualidade da dieta baseado nas orientações dietéticas do governo federal americano (Dietary Guidelines for Americans). O HEI utiliza diferentes grupos alimentares para o cálculo do escore, variando de 0 a 100, em que maiores valores do escore refletem dietas mais próximas das orientações alimentares em vigor.

O índice é composto por 13 componentes baseados nos grupos alimentares descritos nas recomendações dietéticas. Detalhes dos valores máximos e interpretações estão descritos no quadro abaixo:

([**https://storage.googleapis.com/tera-originals/desafio-inferencia/download.png**](https://storage.googleapis.com/tera-originals/desafio-inferencia/download.png))

## **Physical Activity Guidelines for Americans (PAGA)**

O Physical Activity Guidelines for Americans (PAG) é emitido pelo Departamento de Saúde e Serviços Humanos (U.S. Department of Health and Human Services (HHS)) e possui recomendações de atividades físicas. Este documento é utilizado em conjunto com as orientações dietéticas para americanos (Dietary Guidelines for Americans) para promover a importância de ser fisicamente ativo e seguir uma dieta saudável.

O PAGA recomenda que adultos se engajem em pelo menos 150 minutos de atividades aeróbicas de intensidade moderada ou 75 minutos de atividades aeróbicas de intensidade vigorosa semanalmente. A partir dos dados coletados do NHANES, é possível calcular o número de minutos de atividades físicas, definido como a total minutos semannais de atividades físicas moderadas + 2\*(total minutos de atividades aeróbicas vigorosas).

## **Roteiro para o Desafio**

## **1 - Bases de dados**

Segue os seguintes bancos de dados em formato csv:

[**https://drive.google.com/drive/folders/1N4sCFSZDXLYzccP-9t0JXGkF06wiAQS-?usp=sharing**](https://drive.google.com/drive/folders/1N4sCFSZDXLYzccP-9t0JXGkF06wiAQS-?usp=sharing)

**DEMO\_PHQ.csv**: banco de dados contendo 5334 observações de adultos pesquisados no NHANES 2005-2006, segue PDF com detalhamento das variáveis:

[**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, segue PDF com detalhamento das variáveis:

[**https://drive.google.com/file/d/1Qc63Z0t5v7zHKUszWOQlHVCwBWG1Ps7i/view?usp=sharing**](https://drive.google.com/file/d/1Qc63Z0t5v7zHKUszWOQlHVCwBWG1Ps7i/view?usp=sharing)

## **2 - Tratamento de dados**

1. Faça a leitura dos bancos de dados e faça o tratamento que julgar necessário para as variáveis (exemplo: tratamento das categorias 7 = “Se recusou a responder” e 9 = “Não sabe”).

2. Em seguida, combine os dois bancos de dados fornecidos, utilizando a variável SEQN como chave única. O banco de dados final deve conter 5334 observações dos adultos respondentes do NHANES.

3. Crie a variável phq9 - escore Patient Health Questionnaire-9 como a soma das variáveis DPQ010 a DPQ090

4. Crie a variável phq\_grp - grupo de sintomas de depressão que assume 0 (“sem sintomas”) se phq9 < 5, 1 (“sintomas leves”) se 5 <= phq9 < 10, 2 (“sintomas moderados”) se 10 <= phq9 < 15, 3 (“sintomas moderadamente severos”) se 15 <= phq9 < 19 e 4 (“sintomas severos”) se phq9 >= 20

## **3 - EDA: Análise Univariada**

1. Faça uma análise univariada de todas as variáveis da análise.

2. Julgue como proceder com os casos faltantes nas variáveis

3. Re-agrupe ou re-codifique as variáveis que julgar necessário

## **4 - EDA: Análise Bivariada**

1. Faça uma análise bivariada de sintomas de depressão (phq9 ou phq\_grp) com as características demográficas. Descreva o perfil com maiores incidências de sintomas de depressão.

2. Faça uma análise bivariada de características demográficas x hábitos saudáveis. Qual perfil possui hábitos mais saudáveis?

3. Faça uma análise bivariada de sintomas de depressão (phq9 ou phq\_grp) com os hábitos saudáveis (variáveis de atividade física e healthy eating index). O que podemos observar?

## **5 - Testes de hipóteses**

1. Faça o(s) teste(s) de hipóteses adequado(s) para avaliar a significância estatística das diferenças nas características demográficas apontadas na análise bivariada. Quais são as características apresentam diferenças estatisticamente significativas para a frequência de sintomas de depressão?

2. Faça o(s) teste(s) de hipóteses adequado(s) para avaliar se existe associação entre hábitos saudáveis e sintomas de depressão.

## **6 - Aspectos da análise**

1. Qual o tipo de estudo está sendo empregado pelo NHANES? experimental ou observacional?

2. Discuta as possíveis fontes de viés presentes na análise

3. A partir da análise realizada, podemos afirmar que hábitos saudáveis possuem um efeito causal na prevenção de depressão?

4. Quais são as fraquezas das análises realizadas? O que pode ser feito para melhorar?

5. Quais outras variáveis/informações poderiam ter sido coletadas para esta análise?