

# PROJETO 1

ANÁLISE DESCRITIVA

## PERFIL DOS APOSENTADOS

### Problema:

Esse projeto deve conter uma resposta para a seguinte pergunta:

**“Em quais características certo *grupo de aposentados* está melhor (ou pior ou igual) do que algum outro *grupo de comparação*?”**

Você deve escolher dois grupos de brasileiros selecionados da PNAD, desde que ***pelo menos um*** seja só de aposentados. Dependendo do seu objetivo, ambos os grupos podem ser de aposentados, desde que sejam diferenciados por alguma característica, como, por exemplo, Sexo.

Invente a metodologia baseada nos dados considerando ***pelo menos uma*** das PNADs disponibilizadas desde que inclua a PNAD2014.

Sua análise deve olhar para um misto de variáveis quantitativas e qualitativas e usar todos os recursos vistos até o momento na disciplina.

Importante! Se preocupe não só em fazer um filtro das variáveis que julga ser relevantes para diferenciar os grupos (colunas do *dataframe*), mas também um filtro do público que deseja analisar (linhas do *dataframe*).

Procure, sempre que possível, comparar entre aposentados e demais pessoas em condições razoavelmente iguais. Por exemplo, entre os indivíduos chefes de família, com 1 família por domicílio, todos homens, existe diferença entre os aposentados e não aposentados? Desta forma, ganhamos mais força para dizer que as diferenças se devem ao fato de um grupo ser aposentado e outro não, eliminando outras possíveis causas. Nem sempre vai ser possível fazer esta separação.

Por fim, escolha as ferramentas estatísticas mais adequadas para explorar nas suas análises. Seu projeto deve conter um *mix* entre tabela de frequências, gráfico de pizza (ou setor), gráfico de barras, histograma, boxplot, gráfico de dispersão, medidas resumo e ajuste de retas. É válido ressaltar que a escolha dessas ferramentas deve ser norteadas pelas que julgar mais adequadas aos tipos de variáveis e ao questionamento do problema. Cada recurso estatístico usado deve deixar sempre claro o que se deseja investigar e qual a interpretação que se faz dos resultados.

**Base de dados (links para download):**

**PNAD 2014**

VIA DROPBOX:

<https://dl.dropboxusercontent.com/u/65763596/Projeto1/PNAD%20e%20Dicionario%202014.zip>

VIA GITHUB:

<https://github.com/mirwox/CDEng/blob/master/Projeto1/PES2014.zip>

**PNAD 2013**

<https://dl.dropboxusercontent.com/u/65763596/Projeto1/PNAD%20e%20Dicionario%202013.zip>

**PNAD 2011**

<https://dl.dropboxusercontent.com/u/65763596/Projeto1/PNAD%20e%20Dicionario%202011.zip>

**PNAD 2001**

<https://dl.dropboxusercontent.com/u/65763596/Projeto1/PNAD%20e%20Dicionario%202001.zip>

**PNAD 1992**

<https://dl.dropboxusercontent.com/u/65763596/Projeto1/PNAD%20e%20Dicionario%201992.zip>

**Regras:**

1. O projeto 1 é estritamente INDIVIDUAL.
2. Elaborar um relatório que contenha as análises que contribuíram para as conclusões. Use o arquivo layout disponibilizado em:  
<https://github.com/mirwox/CDEng/blob/master/Projeto1/Projeto1%20Layout.ipynb>
3. O relatório deve ser feito em IPython Notebook (com células de texto formatadas em Markdown) e deve conter, pelo menos, as seguintes seções:
  - a. Introdução,
  - b. Análises e explicação dos resultados,
  - c. Conclusão.

A estrutura do documento deve ser clara e de fácil compreensão da linha de raciocínio. Nesse caso, não deve haver excesso de impressões não discutidas de variáveis e *dataframe*. Aconselhamos fazer a análise geral e, após finalizada, salve com outro nome, limpe seu IPython Notebook apenas com os resultados relevantes e melhore seu texto.
4. Deve ser enviado no link que ainda será disponibilizado no Blackboard.
5. O arquivo DEVE ser com extensão .ipynb

**Cronograma:**

**Turmas A, B e C**

DATA	EM SALA – Aula Studio
26/08	Entendendo problema e base de dados para traçar possíveis análises Escolha dos dois (ou mais) grupos de comparação Tenha o projeto CD criado em seu Github e dentro uma pasta Projeto1
30/08	Explorando base de dados e/ou Construção do relatório
<b>30/08</b>	Enviar arquivo com extensão .ipynb via Blackboard – <b>link irá expirar às 23:59.</b>  Fazer git push em seu Github até 23:59.  Um script dos professores vai pegar sua cópia neste horário. Assegure-se de  Deixe o arquivo pes2014.txt em seu .gitignore para não estourar o armazenamento do Github, ou trabalhe com o PES2014 zipado, como exemplo dos professores.

RUBRICS DE AVALIAÇÃO DO OBJETIVO DE APRENDIZADO

Objetivo de aprendizado	Insatisfatório (I)	Em desenvolvimento (D)	Essencial (C)	Proficiente (B)	Avançado (A)
<b>Elaborar análises exploratórias de dados (univariadas e multivariadas), utilizando ferramentas estatísticas e computacionais adequadas.</b>	Não consegue trabalhar com bases de dados de forma proficiente. Apresenta problemas com arquivos, formatos de arquivos ou não tem habilidades básicas de filtragem de dados.	<p>Consegue abrir e filtrar dados de forma básica, mas não consegue conectar os dados à prática de análise.</p> <p>O notebook é praticamente uma coleção de comandos do panda e gráficos, com pouca análise e interpretação.</p> <p>Tem dificuldade para lidar com valores inválidos e não consegue eliminá-los, corrompendo a análise</p>	<p>Identifica em uma base de dados as variáveis pertinentes à resposta de uma pergunta e é capaz de filtrá-las, cruzá-las e processá-las para embasar sua análise.</p> <p>Utiliza pelo menos cinco variáveis mistas entre qualitativas e quantitativas e emprega os recursos adequados a aquelas variáveis de forma pertinente.</p>	<p>Apresenta os comportamentos da rubrica C para 7 variáveis mais diversificação excelente de recursos de acordo com o tipo de variável (qualitativa ou quantitativa).</p> <p>Analisa ocorrências conjuntas com recursos adequados ao tipo de variável (crosstab, gráficos de barras, pizza, histograma e boxplot e gráfico de dispersão e ajuste de retas) respeitando o tipo de variável, e interpreta corretamente o que as ocorrências conjuntas significam em função dos objetivos da pesquisa.</p> <p>Para pelo menos uma comparação quali e uma comparação quant procurou grupos iguais que diferem apenas em relação a serem aposentados.</p> <p>Comunica os resultados com bastante clareza, deixando claro que pergunta está respondendo.</p>	<p>Executa os comportamentos da rubrica B de maneira excepcional e incorpora aspecto histórico, analisando a PNAD 2014 e mais uma.</p>