

## PROJETO 1

### ANÁLISE EXPLORATÓRIA DE DADOS

## PERFIL DOS DOMICÍLIOS BRASILEIROS AO LONGO DE ALGUNS ANOS

### **PNAD:**

<http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2015/default.shtm>

A Pesquisa Nacional por Amostra de Domicílios - PNAD investiga anualmente, de forma permanente, características gerais da população, de educação, trabalho, rendimento e habitação e outras, com periodicidade variável, de acordo com as necessidades de informação para o País, como as características sobre migração, fecundidade, nupcialidade, saúde, segurança alimentar, entre outros temas. O levantamento dessas estatísticas constitui, ao longo dos 48 anos de realização da pesquisa, um importante instrumento para formulação, validação e avaliação de políticas orientadas para o desenvolvimento socioeconômico e a melhoria das condições de vida no Brasil.

A base de dados resultante dessa pesquisa é oferecida no site do IBGE em dois formatos: arquivo de Pessoas – em que cada linha descreve as características das pessoas residentes em um domicílio; e o arquivo de Domicílios – em que cada linha descreve as características particulares de um domicílio.

Nesse projeto, iremos trabalhar com o arquivo de domicílios. A PNAD considera que uma **unidade domiciliar** é um domicílio particular ou uma unidade de habitação (quarto, apartamento etc.) em domicílio coletivo, podendo esse estar localizado em área urbana ou rural.

É importante que se dedique a entender o problema e traçar um objetivo claro para que possa ser capaz de selecionar e trabalhar com as variáveis mais adequadas. A utilização do dicionário será extremamente importante para localizar o nome e significado de cada variável considerada em cada ano de pesquisa.

## PERFIL DOS DOMICÍLIOS BRASILEIROS AO LONGO DE ALGUNS ANOS

### PROBLEMA:

Esse projeto tem como proposta levantar possíveis *semelhanças, diferença e/ou melhorias* em duas possíveis vertentes:

- Descrever os domicílios brasileiros quanto ao acesso à Internet e à Televisão (PNAD 2013, PNAD 2014 e PNAD 2015); ou
- Descrever os domicílios brasileiros quanto à segurança alimentar (PNAD 2009 e PNAD 2013).

Assim, você deverá:

1. Escolher **APENAS UMA DAS OPÇÕES** acima para ser o seu projeto;
2. Descrever um **OBJETIVO AO SEU PROBLEMA** independente de qual seja a opção escolhida;
3. **ESCOLHER E/OU CRIAR VARIÁVEIS** que julgue importante de acordo com seu objetivo;
4. Fazer um levantamento para todos os anos a fim de que possa **IDENTIFICAR UMA EVOLUÇÃO OU NÃO AO LONGO DO TEMPO**.

Obrigatoriamente, sua análise deverá considerar um misto de variáveis quantitativas e qualitativas e usar as diversas ferramentas estatísticas que serão vistas na disciplina.

**Importante!** Se preocupe não só em fazer um filtro das variáveis que julgar serem relevantes para diferenciar os domicílios (colunas do *dataframe*), mas também avaliar se há a necessidade ou não de ser feito um filtro dos domicílios que deseja analisar (linhas do *dataframe*). Por exemplo, se deseja comparar domicílios apenas de áreas urbanas, deve antes excluir todas as linhas referentes aos domicílios localizados em área não urbana, e assim por diante. A importância desse “filtro de linhas” é avaliar, para o seu particular problema, se a grande heterogeneidade em muitas características domiciliares pode dificultar na análise dos resultados que serão obtidos.

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 setor, gráfico de barras, histograma e boxplot, além de medidas resumo. É válido ressaltar que a escolha dessas ferramentas deve ser norteadada pela adequação aos tipos de variáveis e ao questionamento do problema. Cada recurso estatístico utilizado no seu projeto deve deixar sempre claro o que se deseja investigar e qual a interpretação que se faz dos resultados obtidos.

**Links de algumas notícias do IBGE que podem motivar um perfil de domicílios que deseja investigar:**

<http://g1.globo.com/tecnologia/noticia/2015/04/mais-de-50-de-domicilios-brasileiros-tem- apenas-tv-de-tubo-diz-ibge.html>

<http://g1.globo.com/tecnologia/noticia/2016/11/n-de-casas-com-computador-cai-pela-1-vez- no-brasil-diz-ibge.html>

<http://g1.globo.com/tecnologia/noticia/2016/04/internet-chega-pela-1-vez-mais-de-50-das- casas-no-brasil-mostra-ibge.html>

<http://www4.planalto.gov.br/consea/comunicacao/noticias/2014/ibge-divulga-pnad-sobre- seguranca-alimentar-no-brasil>

<http://g1.globo.com/bemestar/noticia/2014/12/72-milhoes-sofrem-com-inseguranca- alimentar-grave-no-pais-diz-pnad-2013.html>

Entre vários outras matérias que pode encontrar no site <http://g1.globo.com/>

## **BASE DE DADOS (LINKS PARA DOWNLOAD):**

### **PNAD DOMICÍLIO 2015**

**- Microdados das Pesquisas Básica e Suplementar de Acesso à Internet e à Televisão**

<https://www.dropbox.com/s/bnekkp7ed9tvav6/PNAD%20DOM2015.zip?dl=0>

### **PNAD DOMICÍLIO 2014**

**- Microdados das Pesquisas Básica e Suplementares de Acesso à Internet e à Televisão e Cadastro Único para Programa Social do Governo Federal**

<https://www.dropbox.com/s/ggakp9eryes5iby/PNAD%20DOM2014.zip?dl=0>

### **PNAD DOMICÍLIO 2013**

**- Microdados das Pesquisas Básica e Suplementares de Acesso à Internet e à Televisão e de Segurança Alimentar**

<https://www.dropbox.com/s/4kstqleje551wwl/PNAD%20DOM2013.zip?dl=0>

### **PNAD DOMICÍLIO 2009**

**- Microdados das Pesquisas Básica e Suplementares de Segurança Alimentar, Vitimização e Justiça**

<https://www.dropbox.com/s/fvyyz2upub3m2oc/PNAD%20DOM2009.zip?dl=0>

## **REGRAS:**

1. O Projeto 1 é estritamente INDIVIDUAL.
2. Elaborar um relatório que contenha as análises que contribuíram para as conclusões de acordo com seu objetivo inicial. Use o **arquivo layout** disponibilizado na pasta Projeto1 do github:
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, o notebook não deve haver excesso de impressões não discutidas de variáveis e de *dataframe*.**

**Aconselhamos fazer uma análise geral e, após finalizada, salve com outro nome, limpe seu IPython Notebook apenas com os resultados relevantes e melhore seu texto.**

4. Seu projeto deve ser adicionado no seu github dentro de uma pasta chamada Projeto1.
5. O arquivo DEVE ser com extensão .ipynb

## **CRONOGRAMA:**

### **Turmas A, B e C**

<b>DATA</b>	<b>Finalização:</b>
17/08	APS1: Check – a definir
24/08	APS2: Check – a definir
<b>29/08</b>	Fazer git push em seu Github até 23:59 do Projeto 1 finalizado.

## 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>	<p>Não consegue trabalhar com bases de dados de forma proficiente.</p> <p>Apresenta problemas com arquivos, formatos de arquivos ou não tem habilidades básicas de filtragem de dados.</p>	<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>Descreve com riqueza de detalhes o objetivo que deseja avaliar nesse projeto de acordo com a vertente escolhida (acesso à internet e à TV ou segurança alimentar).</p> <p>Constrói uma base de dados com variáveis pertinentes ao objetivo definido, sejam essas do tipo qualitativo ou quantitativo.</p> <p>É capaz de filtrá-las, cruzá-las e processá-las para embasar sua análise.</p>	<p>Executa os comportamentos da rubrica C de maneira excepcional.</p> <p>Analisa ocorrências conjuntas com recursos adequados ao tipo de variável (crosstab, gráficos de barras, de setor, histograma e/ou boxplot), respeitando o tipo de variável.</p> <p>Interpreta corretamente o que as ocorrências conjuntas significam em função dos objetivos da pesquisa.</p> <p>Comunica os resultados com bastante clareza.</p>	<p>Executa os comportamentos da rubrica B de maneira excepcional.</p> <p>Incorpora aspecto histórico, analisando todas as PNADs envolvidas na vertente escolhida.</p> <p>Expõe os resultados obtidos de maneira excepcional.</p>