



**Universidade Federal do Ceará**  
**Campus Sobral**  
**Graduação em Ciências Econômicas e Finanças**

# **Microdados no R**

**Prof. Dr. José Weligton Félix Gomes**

- **DEFINIÇÃO DO IBGE:** “Microdados consistem no menor nível de desagregação dos dados de uma pesquisa, retratando, sob a forma de códigos numéricos, o conteúdo dos questionários, preservado o sigilo das informações”.
- **Por Exemplo:**
  - Pesquisa Nacional por Amostragem de Domicílios (PNAD)
  - PNAD Contínua
  - Pesquisa Orçamentária Familiar (POF)
  - Pesquisa Nacional de Saúde (PNS)



# Fontes de Microdados

- <ftp.ibge.gov.br>
- <ftp.inep.gov.br>
- <ftp.datasus.gov.br>
- Rais e Caged  
(<http://pdet.mte.gov.br/microdados-rais-e-caged>)

# ■ ■ ■ Microdados: Arquivos

- Base de Dados com a Extensão .txt (arquivo de texto);
- Dicionário com a Descrição das Variáveis:
  - Dicionário de variáveis de domicílios - PNAD Contínua
  - Dicionário de variáveis de pessoas - PNAD Contínua
- Inputs para a Leitura dos Dados em Programas Estatísticos (SAS, SPSS, R, Stata...);
- Questionário(s).



## 1 – Pacote PNADcIBGE

- Desenvolvido para facilitar o *download*, **importação** e **análise dos dados amostrais** da Pesquisa Nacional por Amostra de Domicílios Contínua realizada pelo Instituto Brasileiro de Geografia e Estatística - IBGE.
- **A PNAD Contínua possui dois tipos de microdados:**
  - **Trimestral**, que contém a parte básica investigada pela pesquisa, contendo variáveis conjunturais de mercado de trabalho referentes a um trimestre civil;
  - **Anual**, que contém temas estruturais específicos investigados na pesquisa para um ano civil.
- Maiores informações sobre a pesquisa e os temas investigados podem ser encontrados no [site oficial do IBGE](#).

## 2 - Pacote survey

- Devido ao plano amostral utilizado na PNAD Contínua, é necessário que sejam utilizadas ferramentas específicas para a análise de dados amostrais complexos.
- O **pacote survey** é um pacote criado especificamente para análise e modelagem de dados provenientes de pesquisas com estes tipos de **planos amostrais**.
- Cálculo de médias, totais, razões, quantis, tabelas de contingência.
- Modelos de regressão, modelos loglineares, curvas de sobrevivência, testes de classificação, para toda a amostra e para domínios (sub - populações).



# Pacote PNADcIBGE, survey e dplyr

**Todas essas funções especiais começam com svy e incluem:**

- `svytable()` – Realiza uma tabulação cruzada ponderada
- `svymean()` – Exibe a média
- `svyratio()` – Proporção entre as variáveis
- `svychisq()` – Teste Qui-Quadrado (Wald)
- `svytotal()` – Exibe o total ponderado e o erro padrão da variável
- `svyhist()` – Histograma
- `svyplot()` – Gráfico
- `svyglm()` – Regressão
- `svyloglin()` – Regressão

## 3 - Pacote dplyr

- O **dplyr** é o pacote mais útil para realizar transformação de dados, aliando simplicidade e eficiência de uma forma elegante. Os scripts em R que fazem uso inteligente dos verbos dplyr e as facilidades do operador *pipe* (`%>%`) tendem a ficar mais legíveis e organizados sem perder velocidade de execução.
- A filosofia **tidy** é a base do **tidyverse**. Os principais pacotes encarregados da tarefa de estruturar os dados são o dplyr e o tidyr.
- **As principais funções do dplyr são:**
  - `filter()` - filtra linhas
  - `select()` - seleciona colunas
  - `mutate()` - cria/modifica colunas
  - `arrange()` - ordena a base
  - `summarise()` - sumariza a base
- **Observação:** Esse pacote funciona nos dados em formato **tibble**. Uma tibble nada mais é do que um data.frame, mas com um método de impressão mais adequado.



## Aplicações no R

- ✓ Importação de dados da PNAD Contínua com a função `get_pnadc` (Dados trimestrais e anuais);
- ✓ Baixar apenas variáveis selecionadas;
- ✓ **Utilizar o pacote `survey` com as seguintes funções:**
  - ✓ `svytotal`
  - ✓ `svymean`
  - ✓ `svyratio`
  - ✓ `svyquantile`
  - ✓ `svyby`
  - ✓ `svyboxplot`

Também conhecido como “*desenho amostral*”, o plano amostral determina **o conjunto de regras e ações utilizadas para seleção da amostra** de participantes em uma pesquisa. Ele é determinado, principalmente, pelas características da população de interesse e deve ser construído por especialistas no assunto, os chamados amostristas.

[Voltar](#)