



## CIÊNCIA DOS DADOS PARA ECONOMISTAS

Plano de aulas atualizado em 26 de junho de 2023

Prof. Francis Petterini ([f.petterini@ufsc.br](mailto:f.petterini@ufsc.br))

**EMENTA:** Estruturas de big data e microdados. Principais fontes de microdados socioeconômicos do Brasil: IBGE (Censo, PNAD, PME, POF e INPC/IPCA), INEP (Censo Escolar, Enem, Enad e SAEB), Ministério do Trabalho (RAIS, CAGED e MEI), Ministério da Justiça (Infopen e Sisnarm), Ministério do Desenvolvimento (ComexStat), Ministério da Saúde (DataSUS), TSE (Eleições), Defesa Civil (S2ID), Tesouro Nacional (FINBra), ANP (SLP), BCB (DadosAbertos), Registros Civil e Comercial (nascimentos, casamentos, óbitos, abertura e fechamento de empresas etc.) e outras. Algoritmos de extração em big data. Leitura de dicionários de microdados. Principais ambientes de tabulação (R, Stata e Python). Identificadores transversais e longitudinais e conexão de big data. Identificação e correção de inconsistências. Interpolação de informação. Análise descritiva. Ferramentas de microeconometria e de machine-learning para a interpretação de big data de microdados socioeconômicos.

**Objetivo:** fornecer ao aluno um ferramental de Ciência dos Dados para leitura, tabulação, conexão e interpretação de informações em big data de microdados socioeconômicos.

**Bibliografia:** as referências básicas são *Econometria Básica*, de Damodar N. Gujarati e Dawn C. Porter; e, as notas técnicas do IBGE e de instituições correlatas. Bibliografias complementares serão disponibilizadas ao longo do curso.

**Avaliação:** mínimo de 75% de presença, e dois seminários em grupos de tamanho a serem definidos – temas indicados no decorrer do semestre, dependendo da quantidade e do perfil e interesse dos alunos. As notas dos seminários são S1 e S2. A nota final é  $(S1+S2)/2$ .

- Para a primeira apresentação deve ser entregue:
  - base de dados (em Excel, Stata ou assemelhado);
  - do file de Stata ou similar; e,
  - PDF da apresentação com +/- 20 slides.
    - Na apresentação, de cerca de 30 minutos, os alunos devem ter:
      - uma introdução ao tema analisado, explicando porque se interessou pelo assunto e qual a relevância disso;
      - uma pequena revisão de Google Scholar sobre o tema;
      - explicação de como a base de dados foi montada;
      - análise descritiva dos dados; e,
      - ideia de modelo econométrico a ser aplicado.
- Para a segunda apresentação, no dia deve ser entregue:
  - base de dados atualizada (em Excel, Stata ou assemelhado);
  - do file atualizado de Stata ou assemelhado; e
  - PDF de apresentação acrescentando os resultados econométricos (regressão + efeitos marginais).
    - Na apresentação, de cerca de 30 minutos, você deve apresentar
      - uma atualização do que foi feito;
      - os resultados modelo econométrico com efeitos marginais estimados e outras explicações; e,
      - uma conclusão e indicação de limitações e sugestões de trabalhos futuros.

**Encontros:** quartas e sextas, no segundo horário regular da manhã, em laboratório.