



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO SOCIOECONÔMICO
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

PROGRAMA DE ENSINO

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, Phyton, Julia e SPSS). 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.

1. IDENTIFICAÇÃO DA DISCIPLINA:

Código: CNM 8012

Nome: Ciência de Dados para Economistas

Créditos: 04

Carga Horária: 72h

2. PRÉ-REQUISITOS:

Sem pré-requisitos

3. IDENTIFICAÇÃO DA OFERTA:

Curso de Graduação em Ciências Econômicas.

4. OBJETIVOS DA DISCIPLINA:

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

5. CONTEÚDO PROGRAMÁTICO:

1. Introdução à Análise de Microdados

- O que são microdados e suas principais características.
- Fontes de microdados no Brasil: IBGE, Ministério do Trabalho, Inep, DataSUS.
- Introdução ao formato de arquivos (CSV, DTA, SAV, TXT) e ferramentas de leitura.

2. Manipulação de Grandes Bases

- Introdução ao Pandas (Python) e data.table/dplyr (R).
- Limpeza, transformação e junção de datasets.
- Estruturas de dados: cross-section, séries temporais e painéis.
- Métodos para otimização de processamento de grandes bases.

3. Estatística Descritiva e Análise Exploratória

- Construção de indicadores econômicos e sociais.
- Medidas de desigualdade: Gini, Theil e decomposições.
- Exploração espacial de dados com mapas e georreferenciamento (geopandas, ggmap).

4. Modelagem Estatística e Inferência

- Aplicação de modelos lineares e logit/probit para análise de microdados.
- Análise causal com Diferenças em Diferenças e Regressão Descontinua.
- Construção de perfis econômicos e demográficos.

5. Estudos de Caso com Microdados

- PNAD Contínua (IBGE): Caracterização do mercado de trabalho e desigualdades regionais.
- Censo Demográfico: Análise de pobreza, habitação e acesso a serviços.
- RAIS e CAGED: Dinâmica do emprego formal e mobilidade laboral.
- Dados do Inep (ENEM, Censo Escolar): Relação entre educação e indicadores econômicos.
- PNS: Saúde da população e impactos econômicos.

6. BIBLIOGRAFIA:

Bibliografia básica:

Basuchoudhary, Atin; Bang, James T.; Sen, Tinni. "Machine-learning Techniques in Economics". Springer, 2017.

Cameron, A. C.; Trivedi, P. K. "Microeconometrics: methods and applications". Cambridge University Press, 2005.

Provost, Foster; Fawcett, Tom. "Data Science for Business: What you need to know about data mining and data-analytic thinking". O'Reilly Media, 2013.

Shikida, Claudio; Monasterio, Leonardo; Nery, Pedro. "Guia brasileiro de análise de dados". ENAP, 2021.