

Introdução à Econometria

Aula 10: Obtenção e atualização de dados



Obtenção de Dados

- Após a definição do que será analisado (variável dependente e variáveis explicativas) parte-se para a fase de coleta de dados e estimação do modelo. Essa aula se concentra na coleta de dados.
- Ao contrário do que se pode imaginar, A obtenção dos dados tem importância similar à própria definição do modelo. Isso ocorre pois a inexistência ou indisponibilidade de um dado pode impossibilitar a estimação do modelo.
- Aqui nos concentraremos em bancos de dados de séries econômicas e de acesso gratuito.

Fontes de Dados

Hoje em dia, com a internet, a tarefa de obtenção de dados ficou mais fácil. Existe uma grande quantidade de bancos de dados disponíveis de forma gratuita.

Dica: sempre que possível tente obter os dados diretamente na fonte.

Principais Bases de Dados do Brasil

Ipea (Ipeadata)

<http://www.ipeadata.gov.br/>

Banco Central (SGS)

<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>

IBGE (Sidra)

<https://sidra.ibge.gov.br/>

Algumas Bases Internacionais

Fed (Fred)

<https://fred.stlouisfed.org/>

Dbnomics (Cepremap)

<https://db.nomics.world/>

Quandl

<https://www.quandl.com/>

Baixando os Dados

Os dados, em grande parte das vezes, podem ser baixados de duas formas:

- Visitando a página do provedor de informações na internet ou;
- Utilizando algum software de forma direta (ou seja, sem ser necessário entrar na página).

Baixando os Dados: Exemplos

- Baixar o indicador e atividade do Banco Central do Brasil (IBC-Br) no SGS;
- Baixar o índice de produção industrial do Brasil no Ipeadata;
- Baixar o índice de produção industrial dos EUA no Fred.

Baixando os dados de forma “profissional” via R

A forma de baixar os dados vai depender da forma como o provedor de informações os disponibiliza. Assim, infelizmente, não existe uma única forma de obtenção dos dados.

A seguir veremos os casos do Ipea, do Banco Central e do Fed.

Dados do Ipeadata via R

A obtenção dos dados do Ipeadata é a mais simples entre as três que serão vistas. Isto ocorre porque o Ipea disponibiliza um pacote do R para fazer isso. Para instalar o pacote e obter uma lista com todas as séries disponíveis basta executar o seguinte código:

```
install.packages("ipeadatar")  
  
library(ipeadatar)  
  
series_ipea <- available_series(language = "br")
```

Dados do Banco Central via R

Para baixar os dados do Banco central é necessário que se consulte sua página para obter os códigos das séries desejadas. Assim, por exemplo, para baixar a série do Indicador de Atividade do Banco Central (IBC-Br), deve-se utilizar o código 24363 no endereço a seguir:

```
http://api.bcb.gov.br/dados/serie/bcdata.sgs.24363/dados?formato=csv
```

Dados do Fed via R

Para os dados do Fed a lógica é idêntica a utilizada para a obtenção dos dados do Banco Central. Assim, por exemplo, para baixar a série da Produção Industrial dos EUA, deve-se utilizar o código INDPRO no endereço a seguir:

```
https://fred.stlouisfed.org/graph/fredgraph.csv?id=INDPRO
```

Automatização da Obtenção de Dados

A rotina de atualização de bancos de dados pode ser muito dispendiosa em termos de tempo e muito suscetível a erros.

Por esses (e por vários outros) motivos é que recomenda-se fortemente a automatização desse processo.

Boas Práticas no Armazenamento de Dados

- Não mesclar células;
- Não incluir células com comentários;
- Não incluir colunas vazias;
- Nomear colunas com algum código iniciado por letras e que não possua espaços ou acentuação.

Exercício

Criar um script para a automatização da obtenção e da atualização de um conjunto de dados. O script deve ter as seguintes características:

- Baixar 6 séries mensais entre séries do BCB, Ipeadata e do Fed (IPCA, Produção Industrial EUA, PIB Mensal, IGP-DI, Taxa Over Selic, Taxa de Câmbio R\$/US\$);
- Exportar os dados para um arquivo xlsx;
- Os dados devem começar no ano 2010;
- O arquivo deve conter uma aba chamada "dados" e;
- Conter uma aba chamada "metadados".