

# Lista 6 - Introdução a Análise de Dados

## Raspagem de dados

### Gabarito

Guilherme Masuko

May 2023

Vamos estudar como três variáveis macroeconômicas brasileiras se relacionam.

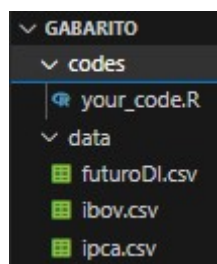
- Inflação
- Taxa de Juros
- Ibovespa

Para cada um desses indicadores, utilizaremos uma base de dados na forma de série temporal.

#### Questão 1

Faça a importação dos dados. Certifique-se de manter seu diretório organizado. Exemplo:

Figure 1: Diretório de Exemplo



---

```
setwd('D:/Economia/PUC_RIO/Monitoria/INF1514 Introdução à Análise  
de Dados/Lista6/Gabarito')
```

```
getwd()
```

---

- a) Faça a importação do arquivo 'futuroDI.csv' em um dataframe. Utilizaremos o futuro DI<sup>1</sup> como proxy da taxa de juros brasileira (Selic). Essa base de dados é composta de 22 colunas contendo dados de Junho de 2012 à Outubro de 2022. Abaixo descrevemos os dados.

- X: Data
- BRPRE\_\_\_.BMF: é a taxa do futuro DI onde os \_\_ são preenchidos pelo horizonte à frente referente à expectativa de taxa de juros, ex: 1M representa o futuro DI um mês à frente. Fonte: Reuters.

### Solução

---

```
library("descr")

file.head("Lista6/Gabarito/data/futuroDI.CSV")

futuroDI <- read.csv("Lista6/Gabarito/data/futuroDI.CSV",
                    dec=".",
                    fileEncoding = "UTF-8-BOM")

View(futuroDI)
```

---

- b) Faça a importação do arquivo 'ipca.csv' em um dataframe. Essa base de dados é composta de 2 colunas contendo dados de Junho de 2012 à Abril de 2023. Abaixo descrevemos os dados.

- X: Data
- valor: Variação da inflação do mês anterior em relação ao mês atual.

### Solução

---

```
file.head("Lista6/Gabarito/data/ipca.CSV")

ipca <- read.csv("Lista6/Gabarito/data/ipca.CSV",
                dec=".",
                fileEncoding = "UTF-8-BOM")

View(ipca)
```

---

---

<sup>1</sup><<https://www.bradesco corretora.com.br/SiteBradescoCorretora/Produtos/Mercados-Futuros/Produtos-Oferecidos/DI-Futuro>>

- c) Faça a importação do arquivo 'ibov.csv' em um dataframe. Essa base de dados é composta de 2 colunas contendo dados de Junho de 2012 à Abril de 2023. Abaixo descrevemos os dados.

- X: Data
- IBOV: Pontos da bolsa brasileira Ibovespa.

### Solução

---

```
file.head("Lista6/Gabarito/data/ibov.CSV")

ibov <- read.csv("Lista6/Gabarito/data/ibov.CSV",
                dec=".",
                fileEncoding = "UTF-8-BOM")

View(ibov)
```

---

### Questão 2

Para cada um dos dataframes, transforme a coluna X em índice (nome das linhas).

### Solução

---

```
# indexando pela data

rownames(futuroDI) <- futuroDI$X
rownames(ipca) <- ipca$X
rownames(ibov) <- ibov$X
```

---

### Questão 3

Para cada dataframe, faça as alterações abaixo:

- a) Para o dataframe futuroDI, mantenha somente a coluna referente ao futuro DI para um mês.

### Solução

---

```
futuroDI <- futuroDI['BRPRE1M.BMF']
```

---

- b) Para o dataframe ipca, mantenha somente a valor.

### Solução

---

```
ipca$X <- NULL
```

---

c) Para o dataframe `ibov`, mantenha somente a IBOV.

### Solução

---

```
ibov$X <- NULL
```

---

### Questão 4

Renomeie as colunas remanescentes para os dataframes `futuroDI`, `ipca` e `ibov`, de `BRPRE1M.BMF`, valor e IBOV, para `Futuro_DI`, `IPCA` e `Ibovespa`.

### Solução

---

```
# renomeando algumas colunas

colnames(futuroDI)[1] <- 'Futuro_DI'
colnames(ipca)[1] <- 'IPCA'
colnames(ibov)[1] <- 'Ibovespa'
```

---

### Questão 5

Precisamos fazer manipulações no dataframe `ipca` para que cada linha tome o valor acumulado da inflação dos últimos 12 meses (assim como o Banco Central mede em <<https://www.bcb.gov.br/>>). Obtemos essa medida calculando a seguinte fórmula.

$$\begin{aligned}\pi_{12t} &= \prod_{j=0}^{11} (1 + \pi_{t-j}) - 1 \\ &= (1 + \pi_t) \cdot (1 + \pi_{t-1}) \cdot \dots \cdot (1 + \pi_{t-11}) - 1\end{aligned}$$

onde  $\pi$  é a inflação.

Para isso, siga os passos:

- Crie colunas com os valores defasados. Seu dataframe deve ficar da seguinte maneira.

Figure 2: Dataframe do IPCA

	IPCA	IPCA_1	IPCA_2	IPCA_3	IPCA_4	IPCA_5	IPCA_6	IPCA_7	IPCA_8	IPCA_9	IPCA_10	IPCA_11
2012-06-30	0.0008	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2012-07-31	0.0043	0.0008	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2012-08-31	0.0041	0.0043	0.0008	NA	NA	NA	NA	NA	NA	NA	NA	NA
2012-09-30	0.0057	0.0041	0.0043	0.0008	NA	NA	NA	NA	NA	NA	NA	NA
2012-10-31	0.0059	0.0057	0.0041	0.0043	0.0008	NA	NA	NA	NA	NA	NA	NA
2012-11-30	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008	NA	NA	NA	NA	NA	NA
2012-12-31	0.0079	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008	NA	NA	NA	NA	NA
2013-01-31	0.0086	0.0079	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008	NA	NA	NA	NA
2013-02-28	0.0060	0.0086	0.0079	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008	NA	NA	NA
2013-03-31	0.0047	0.0060	0.0086	0.0079	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008	NA	NA
2013-04-30	0.0055	0.0047	0.0060	0.0086	0.0079	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008	NA
2013-05-31	0.0037	0.0055	0.0047	0.0060	0.0086	0.0079	0.0060	0.0059	0.0057	0.0041	0.0043	0.0008

## Solução

```

ipca[2:nrow(ipca), 'IPCA_1'] <- ipca[1:(nrow(ipca) - 1), 'IPCA']
ipca[3:nrow(ipca), 'IPCA_2'] <- ipca[1:(nrow(ipca) - 2), 'IPCA']
ipca[4:nrow(ipca), 'IPCA_3'] <- ipca[1:(nrow(ipca) - 3), 'IPCA']
ipca[5:nrow(ipca), 'IPCA_4'] <- ipca[1:(nrow(ipca) - 4), 'IPCA']
ipca[6:nrow(ipca), 'IPCA_5'] <- ipca[1:(nrow(ipca) - 5), 'IPCA']
ipca[7:nrow(ipca), 'IPCA_6'] <- ipca[1:(nrow(ipca) - 6), 'IPCA']
ipca[8:nrow(ipca), 'IPCA_7'] <- ipca[1:(nrow(ipca) - 7), 'IPCA']
ipca[9:nrow(ipca), 'IPCA_8'] <- ipca[1:(nrow(ipca) - 8), 'IPCA']
ipca[10:nrow(ipca), 'IPCA_9'] <- ipca[1:(nrow(ipca) - 9),
  'IPCA']
ipca[11:nrow(ipca), 'IPCA_10'] <- ipca[1:(nrow(ipca) - 10),
  'IPCA']
ipca[12:nrow(ipca), 'IPCA_11'] <- ipca[1:(nrow(ipca) - 11),
  'IPCA']

```

- Compute  $\pi_{12t}$ . Chame essa coluna de IPCA\_Acumulado.

## Solução

```

ipca['IPCA_Acumulado'] <- (1 + ipca['IPCA']) * (1 +
  ipca['IPCA_1']) * (1 + ipca['IPCA_2']) * (1 +
  ipca['IPCA_3']) * (1 + ipca['IPCA_4']) * (1 +
  ipca['IPCA_5']) * (1 + ipca['IPCA_6']) * (1 +
  ipca['IPCA_7']) * (1 + ipca['IPCA_8']) * (1 +
  ipca['IPCA_9']) * (1 + ipca['IPCA_10']) * (1 +
  ipca['IPCA_11']) - 1

```

- Mantenha apenas a coluna IPCA\_Acumulado no dataframe ipca.

### Solução

---

```
ipca <- ipca['IPCA_Acumulado']
```

---

### Questão 6

- a) Una os três dataframes: futuroDI, ipca e ibov.

### Solução

---

```
df = merge(x = ipca, y = ibov, by = 'row.names')

rownames(df) <- df$Row.names

df <- df[, c('IPCA_Acumulado', 'Ibovespa')]

df <- merge(x = df, y = futuroDI, by = 'row.names')

rownames(df) <- df$Row.names

df <- df[, c('IPCA_Acumulado', 'Ibovespa', 'Futuro_DI')]
View(df)
```

---

- b) Drobe todas linhas que tenham NA em alguma das colunas.

### Solução

---

```
library(tidyr)

df <- df %>%
  drop_na()
View(df)
```

---

O dataframe final deve parecer como:

Figure 3: Dataframe Final

	IPCA_Acumulado	Ibovespa	Futuro_DI
2013-05-31	0.06503960	53506	0.07719
2013-06-30	0.06695514	47457	0.08104
2013-07-31	0.06270559	48234	0.08265
2013-08-31	0.06090637	50008	0.08728
2013-09-30	0.05858560	52338	0.09118
2013-10-31	0.05837513	54256	0.09366
2013-11-30	0.05774389	52482	0.09778
2013-12-31	0.05910818	51507	0.09957
2014-01-31	0.05585294	47639	0.10339
2014-02-28	0.05679754	47094	0.10567
2014-03-31	0.06153088	50415	0.10743

### Questão 7

Calcule e interprete as correlações entre as variáveis:

- IPCA\_Acumulado e Futuro\_DI.

#### Solução

---

```
cor(df$IPCA_Acumulado, df$Futuro_DI)
```

```
cor.test(df$IPCA_Acumulado, df$Futuro_DI)
```

---

- Ibovespa e Futuro\_DI.

#### Solução

---

```
cor(df$Ibovespa, df$Futuro_DI)
```

```
cor.test(df$Ibovespa, df$Futuro_DI)
```

---