

# Exercício Programa (EP) 01

Ana C V de Melo

MAC113 - FEA diurno – 2025

## 1 EP: Informação sobre Empresas e Empregos no Brasil

Este EP utiliza um conjunto de informações disponibilizadas pelo IBGE (Instituto Brasileiro de Geografia e Estatística) sobre **Demografia das Empresas e Estatísticas de Empreendedorismo**, referente ao ano de 2020 (<https://www.ibge.gov.br/estatisticas/economicas/comercio/22649-demografia-das-empresas-e-estatisticas-de-empreendedorismo.html>).

No IBGE (<https://www.ibge.gov.br/>), vários dos dados utilizados aqui para o exercício são disponibilizados. Contudo, podemos ainda extrair novas informações a partir desses dados.

## 2 Preparação

1. Abra o RStudio.
2. Na janela de *Files* (canto direito inferior) crie um diretório com seu nome e/ou identificação do EP.
3. Copie do *moodle* todos os arquivos disponíveis para o EP atual.
4. Atualize o diretório de execução do RStudio para a sessão atual: clique em *Session/Set Working Directory/Choose Directory...* e escolha o diretório onde colocou os seus arquivos.
5. Abra o script disponível e pode iniciar a solução do EP. **Não mude o nome do Script!! você vai entregar o EP com esse nome.**

## 3 Sobre este EP

Existem atualmente várias fontes de dados para os mais diversos temas (economia, saúde, esportes, finanças, etc.). Muitas vezes queremos informações que não são exibidas diretamente nos sites que armazenam os dados.

### 3.1 Os dados utilizados:

Como já mencionado, utilizaremos dados fornecidos pelo IBGE. As informações fornecidas aqui foram extraídas de planilhas disponível no site do IBGE (Tabelas 1.5 e 1.6), a partir das quais foram geradas as seguintes planilhas:

1. Planilha *NumeroEntradaSaidaEmpresasNacionalSetor.csv*, na qual temos as seguintes colunas de informações:
  - **Local:** nome do país, no caso o Brasil
  - **Atividade:** atividade econômica a qual pertence a empresa (código) - setor da economia.
  - **Absoluto\_Entrada:** número de empresas abertas num setor da economia.
  - **Taxas\_IN:** percentual de empresas abertas em um período de um ano num setor.
  - **Absoluto\_Saida:** número de empresas fechadas num setor da economia.
  - **Taxas\_OUT:** percentual de empresas fechadas em um período de um ano num setor.

2. Planilha *NumeroEntradaSaidaEmpresasRegionalSetor.csv* , na qual temos as seguintes colunas de informações:

- **Local:** nome da região
- **Atividade:** atividade econômica a qual pertence a empresa (código) - setor da economia.
- **Absoluto\_Entrada:** número de empresas abertas em um setor da economia.
- **Taxas\_IN:** percentual de empresas abertas em um período de um ano num setor.
- **Absoluto\_Saida:** número de empresas fechadas num setor da economia.
- **Taxas\_OUT:** percentual de empresas fechadas em um período de um ano num setor.

3. Planilha *PessoalOcupadoRegional.csv* , na qual temos as seguintes colunas de informações:

- **Local:** nome da região
- **Atividade:** atividade econômica da ocupação (código) - setor da economia.
- **Absoluto\_Entrada:** número de trabalhadores admitidos no setor da economia.
- **Taxas\_IN:** percentual de trabalhadores admitidos no setor no período de um ano.
- **Absoluto\_Saida:** número de trabalhadores demitidos no setor da economia.
- **Taxas\_OUT:** percentual de trabalhadores demitidos no setor no período de um ano.

Obs: As atividades econômicas possuem um código como apresentado abaixo. Esse código corresponde ao dado da coluna **Atividade** das planilhas fornecidas:

Código	Atividade Econômica
A	Agricultura pecuária produção florestal pesca e aquicultura
B	Indústrias extrativas
C	Indústrias de transformação
D	Eletricidade e gás
E	Água esgoto atividades de gestão de resíduos e descontaminação
F	Construção
G	Comércio; reparação de veículos automotores e motocicletas
H	Transporte armazenagem e correio
I	Alojamento e alimentação
J	Informação e comunicação
K	Atividades financeiras de seguros e serviços relacionados
L	Atividades imobiliárias
M	Atividades profissionais científicas e técnicas
N	Atividades administrativas e serviços complementares
O	Administração pública defesa e seguridade social
P	Educação
Q	Saúde humana e serviços sociais
R	Artes cultura esporte e recreação
S	Outras atividades de serviços
U	Organismos internacionais e outras instituições extraterritoriais

### 3.2 Objetivo do EP

Extrair novos dados a partir de informações dessas planilhas. Queremos, com este EP, fazer funções que possam extrair e processar informações das planilhas para responder a novos questionamentos. As informações necessárias e as tarefas a serem realizadas estão descritas nos próximos itens.

## 4 A sua Tarefa será:

### 4.1 Exercício 1

Ler as planilhas dadas e criar os respectivos data frames: código disponível no script fornecido. Você só precisa executar essa parte do script.

```
df_empresasNac <- read.table(  
  file = "NumeroEntradaSaidaEmpresasNacionalSetor.csv",  
  header = TRUE, sep = ";", dec = ",",  
  fill = TRUE  
) # lê a planilha  
  
df_empresasReg <- read.table(  
  file = "NumeroEntradaSaidaEmpresasRegionalSetor.csv",  
  header = TRUE, sep = ";", dec = ",",  
  fill = TRUE  
) # lê a planilha  
  
df_ocupadosReg <- read.table(  
  file = "PessoalOcupadoRegional.csv",  
  header = TRUE, sep = ";", dec = ",",  
  fill = TRUE  
) # lê a planilha
```

### 4.2 Exercício 2

Quando trabalhamos com dados, principalmente em grande quantidade, precisamos fazer alguns ajuste nestes dados para que não atrapalhem o processamento (para aplicar operações aritméticas, por exemplo). Algumas planilhas do IBGE têm células vazias, as quais são preenchidas com *NA* quando colocadas no data frame. Sendo assim, faça uma função *subs\_NA\_por\_zero* que: dada uma coluna de um data frame, substitui todos os objetos *NA* ou *X* pelo valor 0. Iremos substituir, por exemplo, todos os objetos *NA* ou *X* da coluna *Absoluto\_Entrada* do data frame *df\_empresasNac* pelo valor 0 (veja que alguns valores iniciais estavam vazios).

```
tail(df_empresasNac)
```

##	Local	Atividade	Absoluto_Entrada	Taxas_IN	Absoluto_Saida	Taxas_OUT
## 15	Brasil	O	12	5.2	7	3.2
## 16	Brasil	P	10048	13.2	5567	8.1
## 17	Brasil	Q	24651	15.5	7468	5.9
## 18	Brasil	R	5749	20.9	2672	11.9
## 19	Brasil	S	11296	19.3	6125	11.9
## 20	Brasil	U	NA	NA	NA	NA

```
df_empresasNac$Absoluto_Entrada <- subs_NA_por_zero(df_empresasNac$Absoluto_Entrada)  
tail(df_empresasNac)
```

##	Local	Atividade	Absoluto_Entrada	Taxas_IN	Absoluto_Saida	Taxas_OUT
## 15	Brasil	O	12	5.2	7	3.2
## 16	Brasil	P	10048	13.2	5567	8.1
## 17	Brasil	Q	24651	15.5	7468	5.9
## 18	Brasil	R	5749	20.9	2672	11.9
## 19	Brasil	S	11296	19.3	6125	11.9
## 20	Brasil	U	0	NA	NA	NA

A função *tail()* imprime na tela as últimas linhas de um data frame... Você pode imprimir o data frame inteiro para conferir como ficam os valores quando já tiver definido sua função.

Observe que na linha 20 o NA da coluna *Absoluto\_Entrada* foi substituído por zero! Claro que podemos também substituir os NAs das outras colunas aplicando a função *subs\_NA\_por\_zero*. No *script* fornecido com o EP, você encontrará a aplicação dessa função a todas as outras colunas que possuem valores numéricos nas planilhas dadas. Você precisará executar todos os comandos depois de definir a sua própria função.

### 4.3 Exercício 3

Faça uma função *demitidos\_regiao\_faixa* que recebe um data frame com os dados contidos em *df\_ocupadosReg*, uma região, dois valores reais *min* e *max*, e devolve um novo data frame que é o *subset* contendo os setores cujo número de demitidos está entre os valores *min* e *max* (os demitidos estão na coluna *Absoluto\_Saida* do data frame *df\_ocupadosReg*)

Em seguida, faça uma nova função *soma\_demitidos\_faixa* que recebe esse novo dataframe e devolve o número de funcionários demitidos nessa faixa.

```
demitidos_regiao_faixa <- function(df, regiao, min, max) {  
  df_demitidos <-  
    ... complete ...  
  
  )  
  return(df_demitidos)  
}
```

```
soma_demitidos_faixa <- function(df_faixa) {  
  ...  
  
  return( ... )  
}
```

Por exemplo, veja abaixo o resultado obtido quando essas funções são executadas para a região “Norte” com os valores 100 e 200 ( *min* e *max*) e verifique os valores correspondentes nas planilhas fornecidas.

```
demitidos_Norte_100a200 <- demitidos_regiao_faixa(df_ocupadosReg, "Norte", 100, 200)  
print(demitidos_Norte_100a200)
```

```
##      Local Atividade Absoluto_Entrada Taxas_IN Absoluto_Saida Taxas_OUT  
## 1  Norte          A             386      21.0             146      10.0  
## 10 Norte          J             492      18.8             176       8.7  
## 11 Norte          K             270      10.6             119       5.2
```

```
soma_demitidos_Norte_100a200 <- soma_demitidos_faixa(demitidos_Norte_100a200)  
cat("Soma dos demitidos na região Norte (100 a 200): ", soma_demitidos_Norte_100a200)
```

```
## Soma dos demitidos na região Norte (100 a 200): 441
```

### 4.4 Exercício 4

Um dado importante é saber quais setores da economia demitiram mais que contrataram trabalhadores em uma determinada região. Para isso, faça uma função ( *deficit\_empregos\_regiao*) que dados como entrada o data frame *df\_ocupadosReg* e uma região, devolve como resultado um novo data frame com apenas os setores que demitiram mais do que contrataram trabalhadores na dada região.

```
deficit_empregos_regiao <- function(df, regiao) {  
  df_deficit <-
```

```

    ... complete ...

)
return(df_deficit)
}

```

Podemos, por exemplo, aplicar essa função para a região Sul e ainda imprimir quais os setores onde houve deficit de empregos.

```

deficit_Sul <- deficit_empregos_regiao(df_ocupadosReg, "Sul")
print(deficit_Sul)

```

```

##      Local Atividade Absoluto_Entrada Taxas_IN Absoluto_Saida Taxas_OUT
## 24    Sul          D                48      6.9              73      10.3

```

```

setores_deficit_Sul <- deficit_Sul$Atividade
cat("Setores com deficit de empregos na região Sul: ", setores_deficit_Sul)

```

```

## Setores com deficit de empregos na região Sul:  D

```

Veja que V. pode aplicar essa mesma função a outras regiões.

## 4.5 Exercício 5

Semelhante ao que fizemos no exercício anterior, é importante saber quais setores da economia admitiram mais do que demitiram trabalhadores em uma determinada região. Para isso, faça uma função ( *superavit\_empregos\_regiao*) que dados como entrada o data frame *df\_ocupadosReg* e uma região, devolve como resultado um novo data frame com apenas os setores que admitiram mais do que demitiram trabalhadores na dada região.

```

superavit_empregos_regiao <- function(df, regioao) {
  df_superativ <-
    ... complete ...

  )
  return(df_superativ)
}

```

Podemos, por exemplo, aplicar essa função para a região Sul e ainda imprimir quais os setores onde houve superavit de empregos.

```

superavit_Sul <- superavit_empregos_regiao(df_ocupadosReg, "Sul")
print(superavit_Sul)

```

```

##      Local Atividade Absoluto_Entrada Taxas_IN Absoluto_Saida Taxas_OUT
## 21    Sul          A                1010    15.7              463      8.6
## 22    Sul          B                 102     7.3               63      4.8
## 23    Sul          C               9795    10.5             5385      6.4
## 25    Sul          E                 264    10.2              112      4.9
## 26    Sul          F               6868    20.7             3496     12.0
## 27    Sul          G            32955    12.5            21829      8.8
## 28    Sul          H               5102    12.4             4015     10.3
## 29    Sul          I               8118    18.0             5108     12.5
## 30    Sul          J               1428    12.7               910      8.8
## 31    Sul          K                910     7.1               678      5.6
## 32    Sul          L               1228    15.9               549      8.6
## 33    Sul          M              4839    15.5             2144      8.1
## 34    Sul          N               5281    16.6             2535      9.4

```

```
## 36 Sul P 1642 13.3 977 8.9
## 37 Sul Q 4099 14.4 1384 5.9
## 38 Sul R 1163 20.3 582 12.6
## 39 Sul S 2083 18.6 1135 11.7
```

```
setores_superavit_Sul <- superavit_Sul$Atividade
cat("Setores com superavit de empregos na região Sul: ", setores_superavit_Sul)
```

```
## Setores com superavit de empregos na região Sul: A B C E F G H I J K L M N P Q R S
```

Veja que V. pode aplicar essa mesma função a outras regiões.

## 4.6 Exercício 6

Utilizando as funções dos exercícios anteriores, faça uma nova função ( *setores\_superavit*) que dados como entrada o data frame *df\_ocupadosReg*, devolve como resultado uma lista com os setores que admitiram mais do que demitiram trabalhadores em todas as regiões. Veja que nesse caso, V. precisa saber os setores com superativ para cada região e calcular os setores que aparecem em todas as regiões.

```
setores_superavit <- function(df) {
  regioes <- c("Norte", "Nordeste", "CentroOeste", "Sudeste", "Sul")
  setores_br <-
    ... complete ...

  return(setores_br)
}
```

Podemos calcular os setores que tiveram superavit em todas as regiões com o uso da função definida acima.

```
setores_superavit_brasil <- setores_superavit(df_ocupadosReg)
cat("Setores com superavit de empregos no Brasil: ", setores_superavit_brasil)
```

```
## Setores com superavit de empregos no Brasil: A B C E F G H I J K L M N P Q R S
```

## 5 Sobre a entrega e avaliação do EP

### 5.1 Dicas importantes:

- Não modifique o que está escrito no script fornecido. Ele foi feito para ajudá-los a resolver o problema.
- Resolva o problema na ordem em que é sugerida. Alguns precisam de funções dos exercícios anteriores.

### 5.2 Exemplo de Execução e como o seu EP será avaliado:

- Exemplos de execução mostrados em cada exercício.
- a correção será feita mediante comparação dos resultados gerados pelo seu EP com os resultados gerados pelos nossos scripts para os dados atualizados. Vamos basicamente fazer novas perguntas usando as funções que foram pedidas em cada exercício e comparar com os nossos resultados.

### 5.3 O que está sendo fornecido no *moodle*

- as planilhas:
  - a. NumeroEntradaSaidaEmpresasNacionalSetor.csv
  - b. NumeroEntradaSaidaEmpresasRegionalSetor.csv
  - c. PessoalOcupadosRegional.csv
- o script da solução parcial do EP

a. *EP01.script.R*

#### 5.4 O que V. deve entregar:

- V. deve gravar o seu script solução no arquivo *EP01.script.R* (parte da solução já está no script fornecido). **ESTE É O ÚNICO ARQUIVO A SER ENVIADO NO MOODLE - NÃO ENVIE ARQUIVOS .ZIP.**
- Envie o seu arquivo solução para o *moodle* (**precisa ser esse arquivo mesmo e só ele**). Tenha certeza de que você gravou todas as modificações que fez no arquivo antes de entregar.
- Guarde uma cópia para você de todas as listas e EPs que fizer durante a disciplina.
- Observe a **data de entrega**. Só serão recebidos os EPs (pelo próprio sistema) até aquela data.