

Lista 01 - Artur Damião*

Artur Damião

15 de setembro de 2025

Índice

1	Exercício 1	1
2	Analisando o Estado de São Paulo	2
3	Estatística descritiva dos dados	3
4	Trabalhando com gráficos	3
5	Gráfico de densidade de pequenos municípios	4
6	Trabalhando com escala logarítmica	6
7	Médias da população por estado	7

Ambientação

Instalando e carregando os pacotes necessários:

```
devtools::install_github("tbrugz/ribge")
pacman::p_load(
  e1071,
  tidyverse,
  janitor,
  devtools,
  ribge,
  scales,
  patchwork,
  knitr,
  kableExtra
)

# Removendo notação científica
options(scipen = 999)
```

*Número USP: 10701251. E-mail: arturcardoso@usp.br

1 Exercício 1

Carregando a base de dados:

```
pop2020 <- ribge::populacao_municipios(ano = 2020)
```

1. Qual é a unidade de análise desse banco de dados? Quantas observações estão listadas?

Resposta: A unidade de análise do banco de dados é o município. Ao todo, são 5570 observações.

2. Quantas variáveis estão contidas nesse banco de dados?

Resposta: Estão contidas 8 variáveis.

3. Comente sobre outros aspectos da base de dados que considerar interessante, curioso ou que tenha gerado algum tipo de dúvida.

Resposta: A base de dados possui as mesmas variáveis tanto no formato número quanto no formato textual. Por exemplo, a variável `populacao_str` armazena as informações de população no formato de string, enquanto a variável `populacao` armazena as informações no formato numeric. Além disso, não reconheço a variável `codigo_munic` e precisaria consultar o dicionário de variáveis para um melhor entendimento.

4. Utilize a função `select()` do pacote `dplyr` para estas três variáveis: “uf”, “nome_munic”, “populacao”.

```
pop2020 <- pop2020 |>
  dplyr::select(uf, nome_munic, populacao)
```

5. Renomeie a variável “nome_munic” para “municipio”. Utilize a função `rename()`.

```
pop2020 <- pop2020 |>
  dplyr::rename(
    municipio = nome_munic
  )
```

6. Por fim, transforme todos os caracteres dos nomes de município na coluna “municipio” em minúsculo. A função `mutate()` permite criar novas variáveis, e a função `tolower()`¹ transforma todos as letras em minúsculas. Combine as duas para resolver o exercício.

```
pop2020 <- pop2020 |>
  dplyr::mutate(
    municipio_lower = stringr::str_to_lower(municipio)
  )
```

2 Analisando o Estado de São Paulo

Filtrando os municípios do Estado de São Paulo:

¹Optei por usar a função `str_to_lower` do pacote `stringr` para manter a consistência do Tidyverse.

```
pop_sp_2020 <- pop2020 |>
  dplyr::filter(uf == "SP")
```

1. Quantos municípios há no Estado de São Paulo?

Resposta: O Estado de São Paulo possui 645 municípios.

2. Qual o menor município do Estado para o ano de 2020? Quantos habitantes ele tinha?

Resposta: O menor município do Estado em 2020 foi Borá, com 838 habitantes.

3. Quantos municípios paulistas tem uma população maior que um milhão de habitantes? Liste-os, mencionando também a população de cada um.

```
pop_sp_milhao <- pop_sp_2020 |>
  filter(
    populacao > 1000000
  )
```

Resposta: Ao todo, 3 possuem mais de um milhão de habitantes. São eles: Campinas (1.213.792), Guarulhos (1.392.121) e São Paulo (12.325.232).

3 Estatística descritiva dos dados

Estatísticas descritivas usando a função `summarize()`

```
descritivas <- pop_sp_2020 |>
  summarize(
    media = mean(populacao),
    mediana = median(populacao),
    desvio_padrao = sd(populacao),
    variancia = var(populacao)
  )
```

1. Média

Resposta: A média da população do Estado é 71.766,41.

2. Mediana

Resposta: A mediana da população do Estado é 14.141.

3. Desvio-padrão

Resposta: O desvio padrão da população do Estado é 498.489,9.

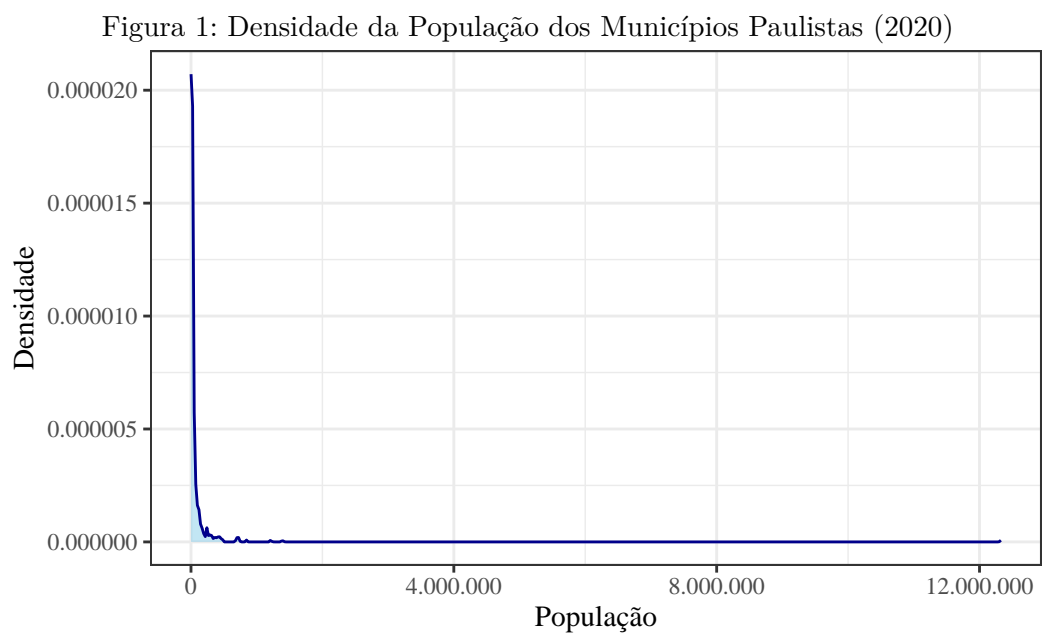
4. Variância

Resposta: A variância da população do Estado é 248.492.160.624.

4 Trabalhando com gráficos

1. Construa um gráfico de densidade (`geom_density()`) da variável população dos municípios paulistas. Extra: tente mudar a cor (argumento `colour =`), e preenchimento da curva (argumento `fill =`), e a transparência da curva (argumento `alpha =`) de seu gráfico.

```
pop_sp_2020 |>
  ggplot(aes(x = populacao)) +
  geom_density(
    colour = "darkblue",    # cor da linha
    fill = "skyblue",      # cor do preenchimento
    alpha = 0.5            # transparência do preenchimento
  ) +
  scale_x_continuous(labels = scales::comma_format(
    big.mark = ".", decimal.mark = ",")) +
  labs(
    x = "População",
    y = "Densidade",
    caption = "Elaborado por Artur Damião\n com dados do pacote ribge"
  ) +
  theme_bw(base_family = "serif")
```



Elaborado por Artur Damião
com dados do pacote ribge

(a) Escala linear no eixo X

2. Comente o gráfico gerado. O que você observa?²

Resposta: É um gráfico bastante assimétrico à direita, em virtude de poucos municípios com muitos milhões de habitantes. Entretanto, a maior parte dos municípios paulistas possuem uma população menor que 100.000 habitantes. Mais especificamente, a vasta maioria dos municípios

²Eu havia feito o gráfico inicialmente com escala logarítmica porque a visualização estava difícil, seguindo os comentários de revisão da Lista. Mas, só depois, vi que o exercício 06 se trata dessa transformação.

possuem uma população ao redor de 10.000 habitantes. Isso pode ser confirmado pelo valor da mediana (14.141, conforme exercício 3.2).

3. A partir da observação do gráfico, qual parece ser a medida mais adequada de tendência central da população: a média ou a mediana? Justifique sua resposta.

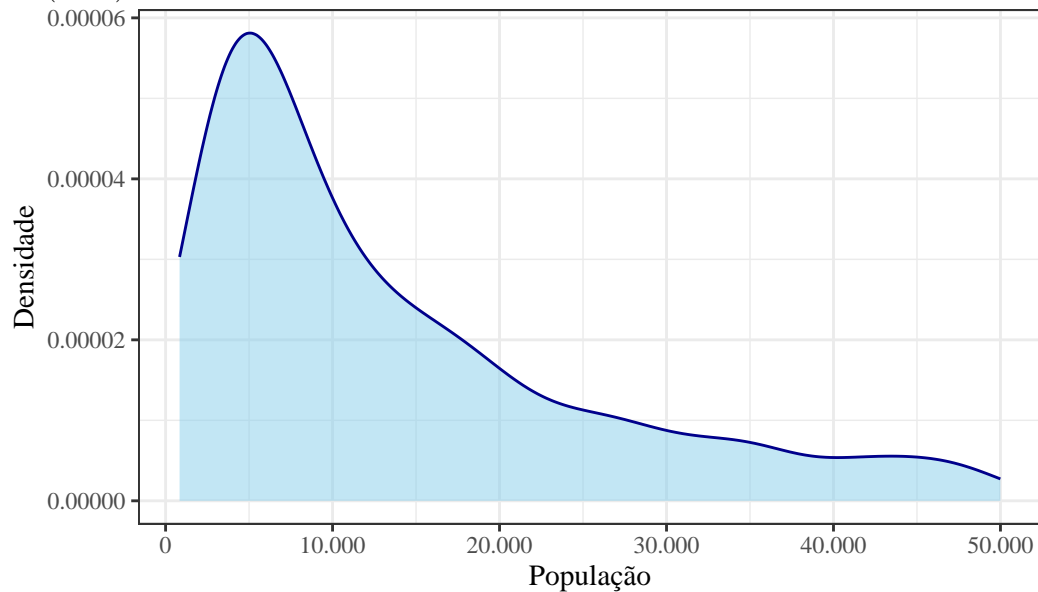
Resposta: A mediana. Isso porque a grande maioria dos municípios, conforme apresentado na Figura 2a, possui população mais próxima de 10.000 habitantes. A média, por sua vez, foi de 71.766,41, bem diferente da realidade. Como possuímos valores discrepantes, a média dá a entender que a maioria dos municípios são maiores do que são de fato.

5 Gráfico de densidade de pequenos municípios

1. Crie um gráfico de densidade apenas para os municípios com menos de 50.000 habitantes.

```
pop_sp_2020 |>
  filter(populacao < 50000) |>
  ggplot(aes(x = populacao)) +
  geom_density(
    colour = "darkblue",    # cor da linha
    fill = "skyblue",       # cor do preenchimento
    alpha = 0.5             # transparência do preenchimento
  ) +
  scale_x_continuous(labels = scales::comma_format(
    big.mark = ".", decimal.mark = ",")) +
  labs(
    x = "População",
    y = "Densidade",
    caption = "Elaborado por Artur Damião\n com dados do pacote ribge"
  ) +
  theme_bw(base_family = "serif")
```

Figura 2: Densidade da População dos Municípios Paulistas Menores que 50 mil Habitantes (2020)



Elaborado por Artur Damião
com dados do pacote ribge

(a) Escala linear no eixo X

2. Quantos são os municípios paulistas com menos de 50.000 habitantes? Qual a porcentagem dessas cidades em relação ao conjunto de municípios do estado? Tente fazer a conta pelo R.

```
# Usando o R base e sem criar novos objetos.  
  
## Número de municípios menores que 50 mil  
sum(pop_sp_2020$populacao < 50000)  
  
## Porcentagem em relação ao total  
sum(pop_sp_2020$populacao < 50000) / nrow(pop_sp_2020) * 100
```

Resposta: Ao todo, o Estado de São Paulo possui 504 com população menor que 50 mil habitantes. Essa porcentagem representa 78,14% do total de municípios.

3. Em comparação com o gráfico da questão anterior, o que você observa?

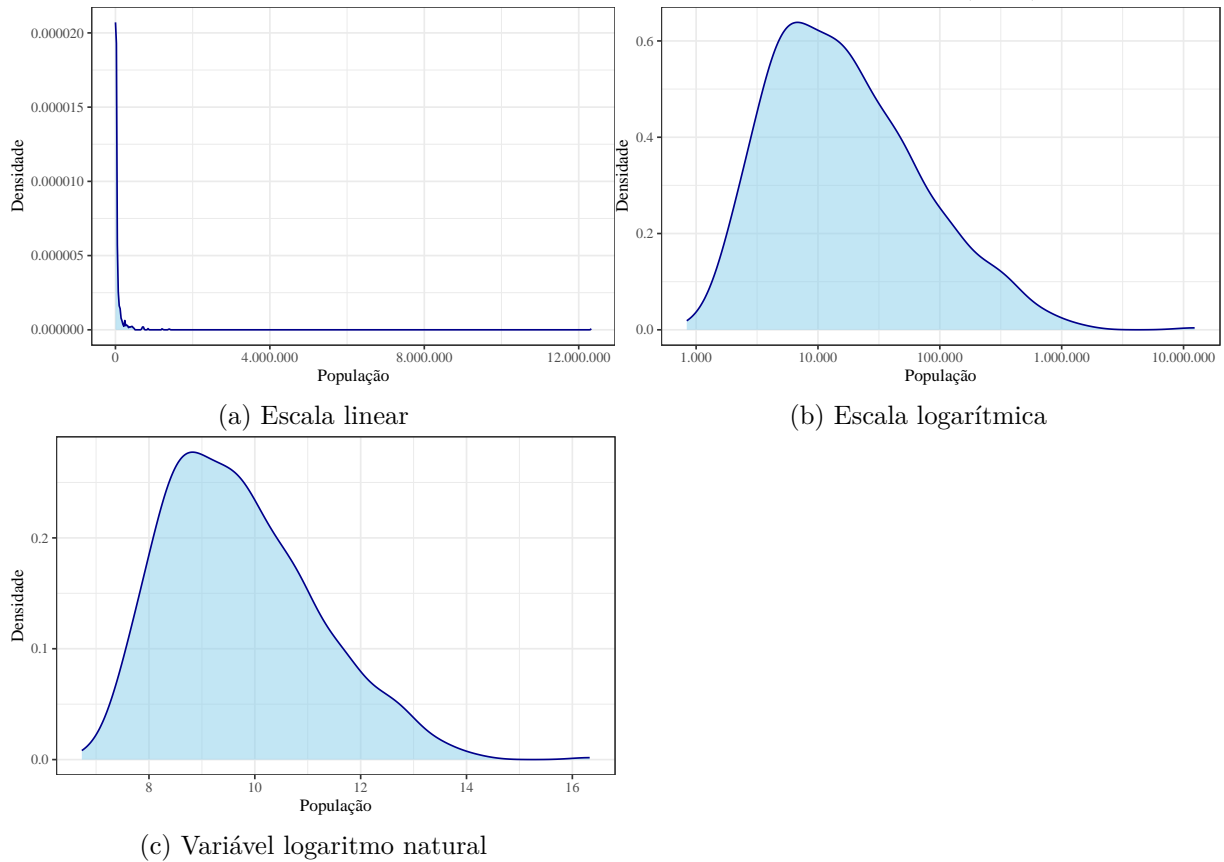
Resposta: É possível ter a dimensão de que a maior parte dos municípios se concentram ao redor dos 10 mil habitantes. A mediana do banco de dados sugere o mesmo.

6 Trabalhando com escala logarítmica

1. Crie novamente um gráfico de densidade para todos os municípios de São Paulo, mas agora utilize a variável de população em logaritmo.

Resposta: Criei 03 gráficos para fins comparativos: o primeiro gráfico, como na Figura 1a, o segundo utilizando a função `scale::scale_x_continuous` e o terceiro criando uma variável de

Figura 3: Densidade da População dos Municípios Paulistas (2020)



logaritmo natural usando a função `log()`. Entre as Figura 3b e Figura 3c, a única diferença é a casa dos milhares para a variável População.

2. Comente o gráfico de densidade, comparando como gráfico do exercício 4.1.

Resposta: A Figura 3b possibilita uma visualização melhor da distribuição da população dos municípios, ao invés da Figura 3a, que, devido ao tamanho de municípios maiores, dificulta a visualização.

7 Médias da população por estado

1. Assim como foi feito para os municípios do estado de São Paulo, calcule a média da população para cada um dos estados brasileiros.

Resposta: A Tabela 1 representa a média da população para os 05 primeiros estados da federação (ordem alfabética pela UF).

```
media_populacao_uf <- pop2020 |>
  group_by(uf) |>
  summarise(
    media = mean(populacao, na.rm = TRUE)
  )

media_populacao_uf |> head(5) |> kable()
```

```
col.names = c("UF", "Média"),
digits = 2,
format.args = list(big.mark = ".", decimal.mark = ",")
)
```

Tabela 1: Média de população por UF

UF	Média
AC	40.657,73
AL	32.858,26
AM	67.866,35
AP	53.860,81
BA	35.804,88

2. Quais dos estados possuem maior e menor população média por município?

Resposta: A Tabela 2a apresenta as 05 UFs com menor média de população por município (Tocantins, Piauí, Paraíba, Rio Grande do Norte e Rio Grande do Sul), enquanto a Tabela 2b apresenta as 05 UFs com maior média de população por município (Distrito Federal, Rio de Janeiro, São Paulo, Amapá e Pará).

```
media_populacao_uf |> arrange(media) |> head(5) |> kable(
  col.names = c("UF", "Média"),
  digits = 2,
  format.args = list(big.mark = ".", decimal.mark = ",")
)
media_populacao_uf |> arrange(desc(media)) |> head(5) |> kable(
  col.names = c("UF", "Média"),
  digits = 2,
  format.args = list(big.mark = ".", decimal.mark = ",")
)
```

Tabela 2: Média de população por município, por UF

(a) Menor média		(b) Maior média	
UF	Média	UF	Média
TO	11.440,63	DF	3.055.149,00
PI	14.649,46	RJ	188.762,92
PB	18.113,35	SP	71.766,41
RN	21.162,66	AM	67.866,35
RS	22.983,85	PA	60.352,40