

Análise Prova Brasil 2011

Lendo os dados:

```
setwd("..") # acessar os dados
arquivo <- "dados/Amostra_g01_Adrielly_Amanda_Erick_Raquel.xlsx"
arquivo_amostra <- "dados/Amostra_50_linhas.xlsx"

# variáveis globais
dados <- read_excel(arquivo) # caso queria testar para a amostra
amostra <- read_excel(arquivo_amostra)
n <- nrow(dados) # caso a amostra seja alterada
ic <- 1.96 # defini com 95%
```

1. Descrever as características das escolas e o desempenho de seus estudantes na Prova de Brasil em 2011.

```
#NULL
```

2. Estimar a proporção de escolas que menos de 75% de seus estudantes participaram da Prova Brasil em 2011.

```
dados$BAIXA_PARTICIPACAO <- dados$PARTICIPACAO < 75
proporcao <- mean(dados$BAIXA_PARTICIPACAO)
print(proporcao)

## [1] 0.105

erro_padrao <- sqrt(proporcao * (1-proporcao)/n)
ic_inf <- proporcao-ic*erro_padrao
ic_sup <- proporcao+ic*erro_padrao

# coloquei 95% mas poderia ser outro
cat(sprintf("IC[%.4f ± %.4f]", proporcao, ic * erro_padrao), "\n")
```

```
## IC[0.1050 ± 0.0425]
cat(sprintf("IC[%.4f; %.4f]", ic_inf, ic_sup))
```

```
## IC[0.0625; 0.1475]
```

```
# obs: escolher o mais usual
```

#3. Estimar a proficiência média em Língua Portuguesa e em Matemática das escolas na Prova Brasil em 2011.

```
# Língua portuguesa
media_pt <- mean(dados$NOTA_LP)
erro_pt <- sd(dados$NOTA_LP)/sqrt(n)
```

```

ic_lp_inf <- media_pt - ic * erro_pt
ic_lp_sup <- media_pt + ic * erro_pt

cat(sprintf("IC língua portuguesa: [%.2f ± %.2f]\n", media_pt, ic * erro_pt))

## IC língua portuguesa: [182.77 ± 3.30]

cat(sprintf("IC língua portuguesa: [%.2f; %.2f]\n", ic_lp_inf, ic_lp_sup))

## IC língua portuguesa: [179.46; 186.07]

# Matemática
media_mt <- mean(dados$NOTA_MT)
erro_mt <- sd(dados$NOTA_MT)/sqrt(n)
ic_mt_inf <- media_mt - ic * erro_mt
ic_mt_sup <- media_mt + ic * erro_mt

cat(sprintf("IC matemática: [%.2f ± %.2f]\n", media_mt, ic * erro_mt))

## IC matemática: [201.36 ± 3.80]

cat(sprintf("IC matemática: [%.2f; %.2f]\n", ic_mt_inf, ic_mt_sup))

## IC matemática: [197.56; 205.16]

```

8. Comparar a proporção de escolas que menos de 75% de seus estudantes participaram da Prova Brasil em 2011 segundo:

Local da escola

Análise para amostra de 200

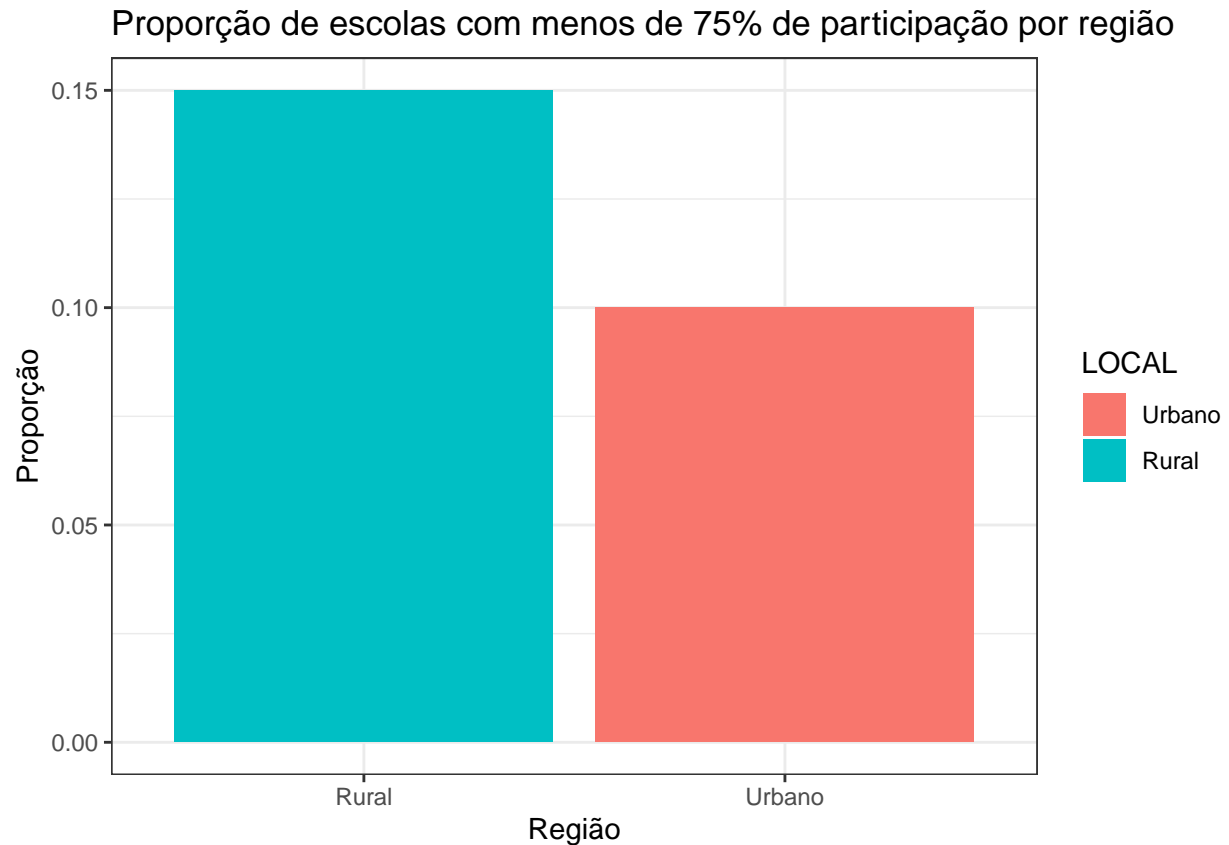
```

menor_75 <- dados %>%
  group_by(LOCAL) %>%
  summarise(
    total = n(),
    menor_que_75 = sum(PARTICIPACAO < 75),
    prop = round(menor_que_75 / total, 2)
  )

menor_75$LOCAL <- menor_75$LOCAL %>%
  factor(levels = c("1", "2"), labels = c("Urbano", "Rural"))

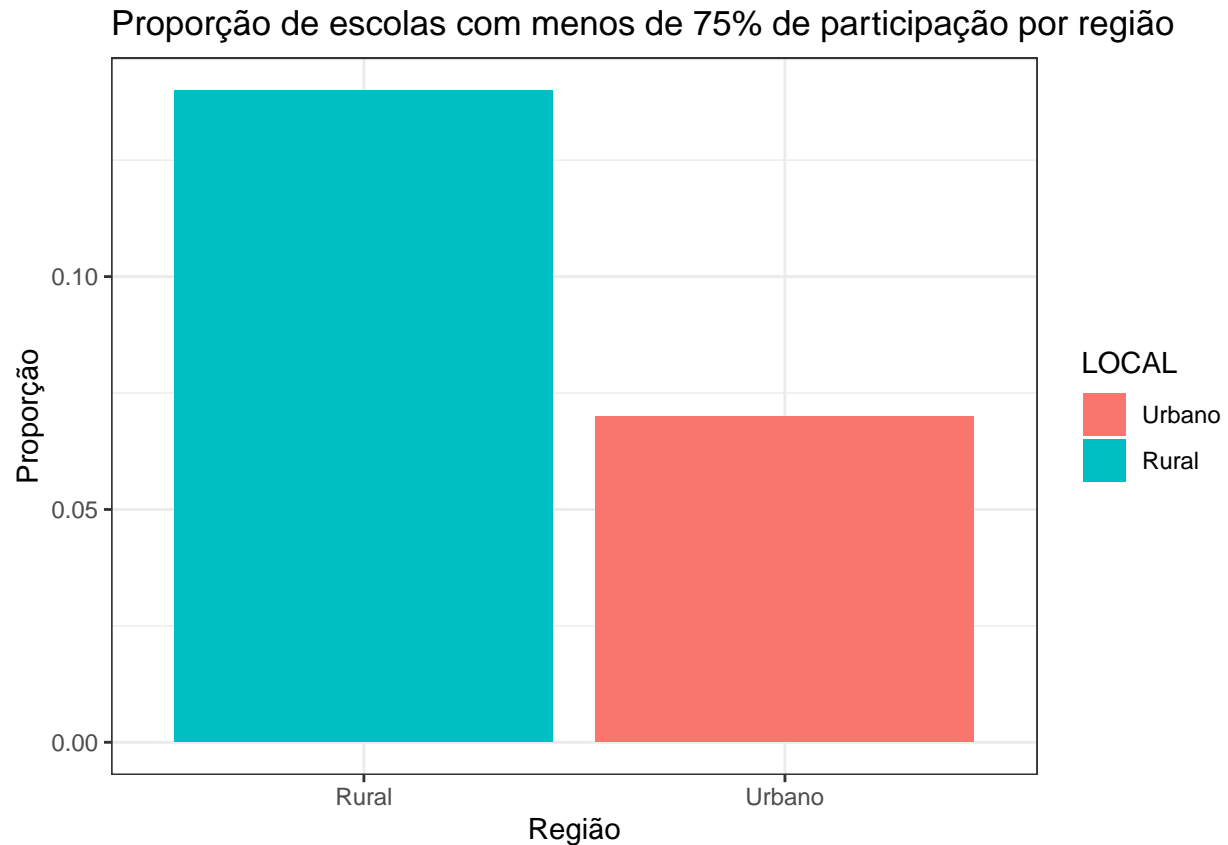
menor_75 %>%
  ggplot(aes(x = reorder(LOCAL, -prop), y = prop, fill = LOCAL)) +
  geom_bar(stat = "identity") +
  labs(title = "Proporção de escolas com menos de 75% de participação por região",
       x = "Região", y = "Proporção") +
  theme_bw()

```



Análise para amostra de 50

```
menor_75 <- amostra %>%  
  group_by(LOCAL) %>%  
  summarise(  
    total = n(),  
    menor_que_75 = sum(PARTICIPACAO < 75),  
    prop = round(menor_que_75 / total, 2)  
  )  
  
menor_75$LOCAL <- menor_75$LOCAL %>%  
  factor(levels = c("1", "2"), labels = c("Urbano", "Rural"))  
  
menor_75 %>%  
  ggplot(aes(x = reorder(LOCAL, -prop), y = prop, fill = LOCAL)) +  
  geom_bar(stat = "identity") +  
  labs(title = "Proporção de escolas com menos de 75% de participação por região",  
        x = "Região", y = "Proporção") +  
  theme_bw()
```

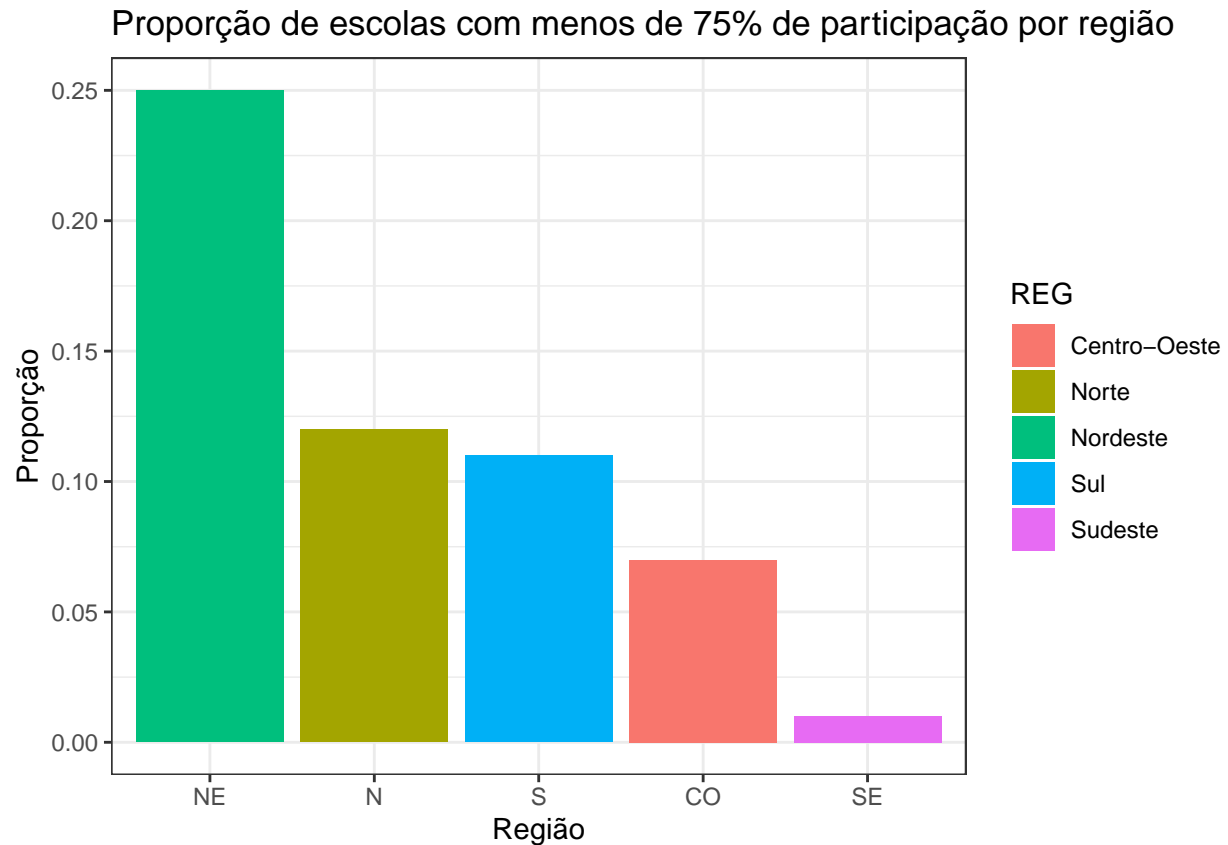


Região de localização da escola

Análise para amostra de 200

```
menor_75 <- dados %>%
  group_by(REG) %>%
  summarise(
    total = n(),
    menor_que_75 = sum(PARTICIPACAO < 75),
    prop = round(menor_que_75 / total, 2)
  )

menor_75 %>%
  ggplot(aes(x = reorder(REG, -prop), y = prop, fill = REG)) +
  geom_bar(stat = "identity") +
  labs(title = "Proporção de escolas com menos de 75% de participação por região",
       x = "Região", y = "Proporção") +
  scale_fill_discrete(
    labels = c("N" = "Norte", "S" = "Sul", "CO" = "Centro-Oeste", "NE" = "Nordeste", "SE" = "Sudeste")
  ) +
  theme_bw()
```



Análise para amostra de 50

```
menor_75 <- amostra %>%
  group_by(REG) %>%
  summarise(
    total = n(),
    menor_que_75 = sum(PARTICIPACAO < 75),
    prop = round(menor_que_75 / total, 2)
  )

menor_75 %>%
  ggplot(aes(x = reorder(REG, -prop), y = prop, fill = REG)) +
  geom_bar(stat = "identity") +
  labs(title = "Proporção de escolas com menos de 75% de participação por região",
       x = "Região", y = "Proporção") +
  scale_fill_discrete(
    labels = c("N" = "Norte", "S" = "Sul", "CO" = "Centro-Oeste", "NE" = "Nordeste", "SE" = "Sudeste")
  ) +
  theme_bw()
```

Proporção de escolas com menos de 75% de participação por região

