

Análise de Amostragem: Estudo de Caso sobre Sorologia Positiva para COVID-19

Pedro Lucas Moreira Barbosa Silva

Outubro de 2024

1 Defina a população de estudo, a unidade de observação e as variáveis envolvidas no estudo

- **População de estudo:** Todos os domicílios do bairro, representados por 90 casas, divididos entre Estrato A e Estrato B.
- **Unidade de observação:** Os domicílios do bairro.
- **Variáveis envolvidas no estudo:**
 - **Estrato:** Localidade do bairro (Estrato A ou B).
 - **n_positivo:** Quantidade de pessoas com sorologia positiva para COVID-19 em cada domicílio.
 - **Casa:** Identificação numérica que varia de 1 a 90 em cada quarteirão.

2 Total de pessoas com sorologia positiva para COVID-19 no bairro

O total de pessoas com sorologia positiva no bairro foi obtido através da soma dos valores da variável `n_positivo` em todos os domicílios. O código R a seguir ilustra como calcular o total de pessoas positivas:

```
# Carregando a base de dados
dados <- COVID_bairro

# Calculando o total de pessoas com sorologia positiva
total_positivo <- sum(dados$n_positivo)
total_positivo
```

3 Selecione uma amostra aleatória de 9 domicílios e calcule o total de pessoas com sorologia positiva para COVID-19. Utilize cada uma das seguintes técnicas

3.1 Amostra Aleatória Simples Sem Reposição

Neste método, são selecionados 9 domicílios de forma aleatória, sem reposição.

```
#Amostra aleatoria simples sem reposicao, usando a funcao sample
#que embaralha os elementos do objeto, ou seja, monta um
#vetor de mesmo tamanho com os elementos alocados aleatoriamente.

amostra_simples <- dados[sample(1:nrow(dados), 9, replace = FALSE), ]
total_amostra_simples <- sum(amostra_simples$n_positivo)
total_amostra_simples
```

3.2 Amostra Sistemática

Aqui, selecionamos domicílios a intervalos regulares, com o intervalo j calculado como:

```
# Definindo o intervalo j
j <- floor(nrow(dados) / 9)

# Selecionar amostra sistemática
amostra_sistemática <- dados[seq(1, nrow(dados), by = j), ][1:9, ]
total_amostra_sistemática <- sum(amostra_sistemática$n_positivo)
total_amostra_sistemática
```

3.3 Amostra Estratificada Proporcional

Dividimos a amostra proporcionalmente ao tamanho dos estratos A e B. O código a seguir utiliza o pacote `sampling` para realizar a amostragem estratificada:

```
library(dplyr)
library(sampling)

amostra_estratificada <- strata(
  data = dados,
  stratanames = "Estrato",
  size = rep(9, length(unique(dados$Estrato))),
  method = "srswor"
)

indices_amostra <- amostra_estratificada$ID_unit
amostra_estratificada_final <- dados[indices_amostra, ]
total_amostra_estratificada <- sum(amostra_estratificada_final$n_positivo)
total_amostra_estratificada
```

3.4 Amostragem por Conglomerados em Dois Estágios

Nesta técnica, primeiro selecionamos 3 bairros, e depois 3 domicílios em cada bairro:

```
bd_auxiliar = dados
bd_auxiliar$quarteirao <- rep(1:9, each = 10)

# Selecao de quarteiroes
quarteiroes = sample(1:9, 3)
bd_auxiliar <- bd_auxiliar[bd_auxiliar$quarteirao %in% quarteiroes, ]

# Selecao de domicilios
sorteio4 = strata(bd_auxiliar, c("quarteirao"), size=c(3, 3, 3), method="srswor")
sorteio4 = sorteio4$ID_unit
amostra_conglomerada = bd_auxiliar[sorteio4, ]
casos_amostra_conglomerada = sum(amostra_conglomerada$n_positivo)
```

4 Características das Amostragens

- **Amostra Aleatória Simples (sem reposição):** É simples de aplicar, mas pode não ser representativa no caso de haver muita variação nos dados.
- **Amostra Sistemática:** Seleciona domicílios a intervalos regulares, útil quando os dados já estão aleatoriamente distribuídos.
- **Amostra Estratificada Proporcional:** Garante representatividade de todos os estratos, sendo mais precisa em casos de variação significativa entre eles.
- **Amostra por Conglomerados:** É eficiente em grandes áreas, mas a precisão depende da representatividade dos conglomerados selecionados.

5 Conclusão

A técnica de **Amostragem Estratificada Proporcional** foi a que melhor estimou a quantidade de pessoas com sorologia positiva para COVID-19, pois garantiu que todos os estratos do bairro fossem representados proporcionalmente, minimizando o viés e aumentando a precisão.