

# Atividade Avaliativa

## Tópicos de Amostragem

Ana Maria Alves da Silva

2025-01-11

Observação: a explicação do que foi feito em cada questão estará descrito no código.

Item 1 - Considere o plano amostral de Amostragem Aleatória Simples com reposição e a variável IDADEMAE. Com o objetivo de estimar a média das idades de uma população, determine o tamanho necessário da amostra considerando um erro máximo de 1 ano, um nível de confiança de 95% e o conhecimento da variância populacional ( $\sigma^2 = 44,8$ ). Após dimensionar a amostra, selecione-a aleatoriamente, calcule a média amostral e construa um intervalo de confiança para essa média. Por fim, interprete os resultados obtidos.

O dimensionamento da amostra deverá ser feito utilizando-se a fórmula  $n = \frac{\sigma^2}{(\frac{B}{z_\alpha})^2}$  e o intervalo de confiança deverá ser construído utilizando-se a fórmula  $\left(\bar{x} - z_\alpha \sqrt{\frac{\sigma^2}{n}}; \bar{x} + z_\alpha \sqrt{\frac{\sigma^2}{n}}\right)$ .

**Solução:** Inicialmente, vamos carregar o conjunto de dados que iremos utilizar ao longo da atividade, além disso, não fizemos um novo download da base de dados que iremos usar. Iremos usar a base de dados disponibilizada no curso.

```
# Configurando o ambiente, lendo o dataframe, verificando o dimensionamento
setwd <- "/Users/anamaria/especializacao/modulo_9/atividade"
df <- read.csv("Dataframe_AulaFabiano.csv", sep = ";")
print(dim(df))
```

```
## [1] 779927      6
```

```
# Checando se é um dataframe
print(is.data.frame(df))
```

```
## [1] TRUE
```

Agora, façamos o que se pede na questão.

```
# Variável de interesse da questão: IDADEMAE
variancia <- 44.8 # Variância populacional fornecida na questão
erro_maximo <- 1 # Erro máximo permitido, pelo enunciado
z_alpha <- 1.96 # Valor de Z para 95% de confiança solicitado

# Dimensionamento da amostra
```

```
n <- variancia / (erro_maximo / z_alpha)^2
n <- ceiling(n) # Arredondar para o inteiro mais próximo
cat("Tamanho necessário da amostra:", n, "\n")
```

```
## Tamanho necessário da amostra: 173
```

```
# Seleção da amostra (aleatória com reposição, replace = TRUE)
set.seed(123) # Para reprodutibilidade
amostra <- sample(df$IDADEMAE, size = n, replace = TRUE)

# Cálculo da média amostral
media_amostral <- mean(amostra)
cat("Média amostral:", media_amostral, "\n")
```

```
## Média amostral: 27.31214
```

```
# Construção do intervalo de confiança solicitado
ic_inferior <- media_amostral - z_alpha * sqrt(variancia / n)
ic_superior <- media_amostral + z_alpha * sqrt(variancia / n)
cat("Intervalo de confiança: [", ic_inferior, ", ", ic_superior, "]\n")
```

```
## Intervalo de confiança: [ 26.31473 , 28.30954 ]
```

```
# Interpretação dos resultados
cat("Com base na amostra selecionada, o intervalo de confiança para a média das  
idades das mães é de",  
    round(ic_inferior, 2), "a", round(ic_superior, 2),  
    "anos com 95% de confiança.\n")
```

```
## Com base na amostra selecionada, o intervalo de confiança para a média das  
## idades das mães é de 26.31 a 28.31 anos com 95% de confiança.
```

Note que para garantir o erro máximo de 1 ano na estimativa da média das idades das mães, com um nível de confiança de 95%, foi necessário selecionar uma amostra de tamanho 173. Além disso, a média das idades das mães, calculada com base na amostra selecionada, é 27,31 anos, representando a estimativa pontual da média da população. Com o intervalo de confiança solicitado, a média verdadeira das idades das mães na população está entre 26,31 anos e 28,31 anos. Com isso, podemos concluir que a amostra foi suficiente para estimar a média das idades das mães com alta precisão, respeitando as condições do problema.

**Item 2 - Considere o plano amostral de Amostragem Aleatória Simples sem reposição e a variável IDADEMAE. Com o objetivo de estimar a média das idades de uma população, determine o tamanho necessário da amostra considerando um erro máximo de 1 ano, um nível de confiança de 95% e o conhecimento da variância populacional ( $\sigma^2 = 44,8$ ). Após dimensionar a amostra, selecione-a aleatoriamente, calcule a média amostral e construa um intervalo de confiança para essa média. Por fim, interprete os resultados obtidos.**

O dimensionamento da amostra deverá ser feito utilizando-se a fórmula  $n = \frac{1}{\frac{B^2}{z_\alpha^2 \sigma^2} + \frac{1}{N}}$ , onde  $N$  é o tamanho da população e o intervalo de confiança deverá ser construído utilizando-se a fórmula  $\left( \bar{x} - z_\alpha \sqrt{\left(1 - \frac{n}{N}\right) \frac{\sigma^2}{n}}; \bar{x} + z_\alpha \sqrt{\left(1 - \frac{n}{N}\right) \frac{\sigma^2}{n}} \right)$ .

**Solução:** Note que o dataframe utilizado para essa questão será o mesmo da questão anterior.

```
# Variável de interesse: IDADEMAE
variancia <- 44.8 # Variância populacional fornecida na questão
erro_maximo <- 1 # Erro máximo permitido no enunciado
z_alpha <- 1.96 # Valor de Z para 95% de confiança solicitado
N <- nrow(df) # Tamanho da população (número de registros no dataframe)

# Cálculo do dimensionamento da amostra
n_amostra <- (variancia / (erro_maximo / z_alpha)^2) / (1 + (variancia / (erro_maximo / z_alpha)^2) / N)
n_amostra <- ceiling(n_amostra) # Arredondar para o inteiro mais próximo
n_amostra <- min(n_amostra, N) # Garantir que n <= N
cat("Tamanho necessário da amostra:", n_amostra, "\n")
```

## Tamanho necessário da amostra: 173

```
# Seleção da amostra (aleatória sem reposição, replace = FALSE)
set.seed(123) # Garantir reprodutibilidade
amostra <- sample(df$IDADEMAE, size = n_amostra, replace = FALSE)

# Cálculo da média amostral
media_amostral <- mean(amostra)
cat("Média amostral:", media_amostral, "\n")
```

## Média amostral: 27.31214

```
# Construção do intervalo de confiança
erro_padrao <- sqrt((1 - n_amostra / N) * variancia / n_amostra)
ic_inferior <- media_amostral - z_alpha * erro_padrao
ic_superior <- media_amostral + z_alpha * erro_padrao
cat("Intervalo de confiança: [", ic_inferior, ", ", ic_superior, "]\n")
```

## Intervalo de confiança: [ 26.31484 , 28.30943 ]

```
# Interpretação dos resultados obtidos
cat("Com base na amostra selecionada, o intervalo de confiança para a média das  
idades das mães é de", round(ic_inferior, 2), "a", round(ic_superior, 2),  
"anos com 95% de confiança.\n")
```

```
## Com base na amostra selecionada, o intervalo de confiança para a média das  
## idades das mães é de 26.31 a 28.31 anos com 95% de confiança.
```

Note que os resultados obtidos foram similares ao da questão 1.

**Item 3 - Existem diversas maneiras de classificar as pessoas, cada uma com um propósito diferente. Uma das classificações úteis, por exemplo, para questões de marketing, é a classificação em classes sociais. Considerando os critérios de classificação empregados atualmente no Brasil, as categorias podem ser generalizadas da seguinte forma:**

**Classe A: Famílias com renda mensal igual ou superior a R\$ 14.400,00.**

Classe B: Famílias com renda mensal entre R\$ 7.100,00 e R\$ 14.399,00.

Classe C: Famílias com renda mensal entre R\$ 2.600,00 e R\$ 7.099,00.

Classe D: Famílias com renda mensal igual ou inferior a R\$ 2.599,00.

Suponha que uma determinada população em estudo esteja distribuída nesses estratos da seguinte forma:

Classe A: 60 famílias

Classe B: 90 famílias

Classe C: 120 famílias

Classe D: 480 famílias

Pretende-se selecionar 100 famílias para analisar o comportamento de consumo dessa população. Responda:

1) Quantas unidades amostrais devem ser retiradas de cada classe, se adotarmos o plano de Amostragem Aleatória Estratificada Uniforme?

Lembre-se de que o número de elementos do estrato  $h$  que comporá a amostra é dado por  $n_h = \frac{n}{H}$ , onde  $n$  é o tamanho da amostra e  $H$  é a quantidade de estratos.

```
# Definir os parâmetros do problema
n_total <- 100 # Tamanho total da amostra fornecido na questão (100 famílias)
H <- 4        # Quantidade de estratos

# Calcular o tamanho da amostra para cada estrato (amostragem uniforme)
n_h <- n_total / H # Divisão uniforme
n_h <- ceiling(n_h) # Arredondar para cima para garantir números inteiros

# Classes e distribuição da população
classes <- c("Classe A", "Classe B", "Classe C", "Classe D")
populacao <- c(60, 90, 120, 480) # Número de famílias em cada classe

# Criar uma tabela com os resultados obtidos
amostragem_uniforme <- data.frame(
  Classe = classes,
  Populacao = populacao,
  Amostra_Aleatoria_Uniforme = rep(n_h, H)
)
```

```
# Exibir os resultados obtidos
print(amostragem_uniforme)
```

Solução:

```
##      Classe Populacao Amostra_Aleatoria_Uniforme
## 1 Classe A          60                      25
## 2 Classe B          90                      25
## 3 Classe C         120                      25
## 4 Classe D         480                      25
```

```
cat("Cada classe terá aproximadamente", n_h, "famílias na amostra.\n")
```

```
## Cada classe terá aproximadamente 25 famílias na amostra.
```

2) Quantas unidades amostrais devem ser retiradas de cada classe, se adotarmos o plano de Amostragem Aleatória Estratificada Proporcional?

```
# Parâmetros do problema
n_total <- 100 # Tamanho total da amostra
populacao <- c(60, 90, 120, 480) # Tamanho da população em cada classe
classes <- c("Classe A", "Classe B", "Classe C", "Classe D")
N <- sum(populacao) # Tamanho total da população

# Cálculo do tamanho proporcional da amostra para cada estrato
n_h <- n_total * populacao / N # Fórmula:  $n_h = n * (N_h / N)$ 
n_h <- round(n_h) # Arredondar para inteiros

# Criar tabela com os resultados
amostragem_proporcional <- data.frame(
  Classe = classes,
  Populacao = populacao,
  Amostra_Proporcional = n_h
)

# Exibir os resultados
print(amostragem_proporcional)
```

Lembre-se de que o número de elementos do estrato  $h$  que comporá a amostra é dado por  $n_h = n \cdot \frac{N_h}{N}$ , onde  $n$  é o tamanho da amostra,  $N_h$  é o tamanho de estrato  $h$  e  $N$  é o tamanho da população.

```
##      Classe Populacao Amostra_Proporcional
## 1 Classe A          60                      8
## 2 Classe B          90                     12
## 3 Classe C         120                     16
## 4 Classe D         480                     64
```

```
cat("Tamanho total da população (N):", N, "\n")
```

```
## Tamanho total da população (N): 750
```

```
cat("Tamanho total da amostra (n):", n_total, "\n")
```

```
## Tamanho total da amostra (n): 100
```