



Universidade de Brasília
Departamento de Estatística

Intervalo de Confiança

Lucas Fonseca Alves

Projeto apresentado para a matéria Métodos Estatísticos 1 como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2023

Atividade 3.3 - Intervalos de Confiança

1. Gere 1000 (mil) amostras de tamanho 15 de uma variável normalmente distribuída com média μ e desvio padrão σ .
 - (a) A média μ deverá ser gerada aleatoriamente no intervalo entre 100 e 200.
 - (b) Para o cálculo do desvio padrão, considere um CV igual a 10%.
2. Para cada amostra, calcule a média amostral, o desvio padrão amostral e o IC de 95% para μ (variância desconhecida).
3. Faça uma análise descritiva das médias amostrais (medidas de posição e histograma). Comente os resultados.
4. Faça uma análise dos IC95%. Quantos deles contêm o valor de μ ? Selecione aleatoriamente 100 IC95%, e apresente-os graficamente.

1 Relatório

Os códigos para o desenvolvimento da atividade foram escritos em linguagem R e estão disponíveis no Apêndice A. A variável aleatória normalmente distribuída foi gerada com os seguintes parâmetros:

Tabela 1: Medidas resumo do Parâmetros

Parâmetro	Valor
Média	193
Desvio Padrão	19.3

Em seguida, foram geradas e armazenadas em uma matriz 1000 amostras, cada uma com tamanho 15. Posteriormente, calculamos, para cada amostra, a média amostral, o desvio padrão amostral e o IC de 95% para μ , armazenando os resultados em um data frame

Figura 1: Histograma das Médias Amostrais

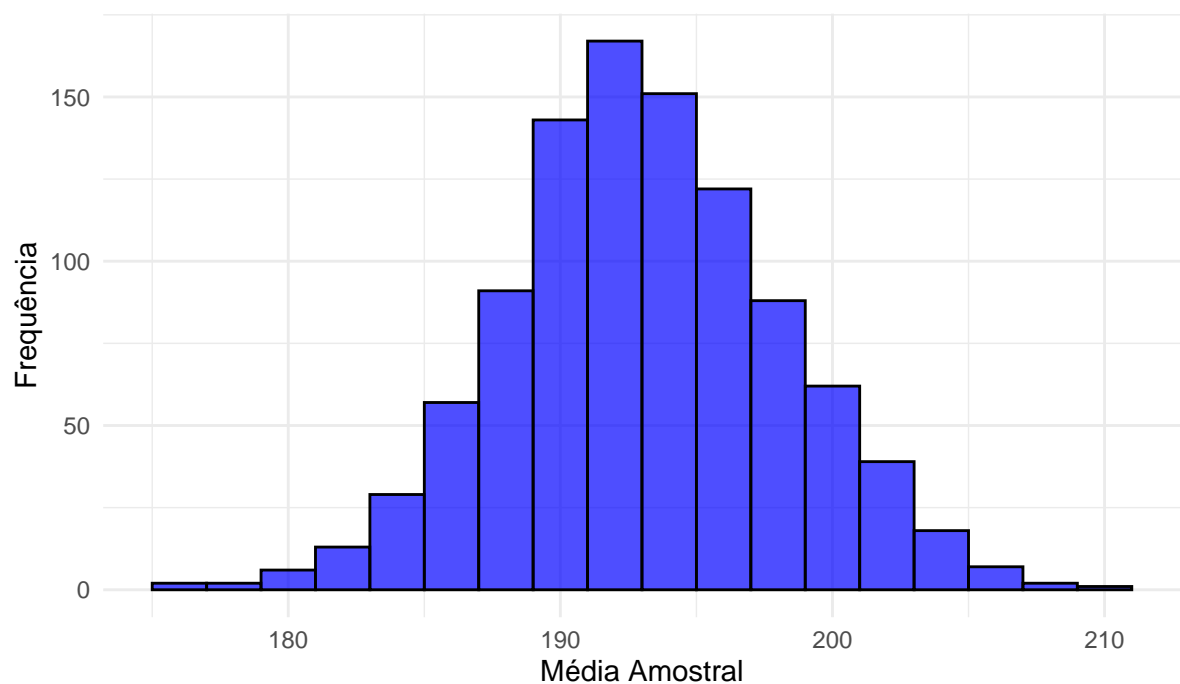


Tabela 2: Medidas de Posição

Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo
176.2	189.8	192.9	193.1	196.4	209.0

As amostras foram geradas a partir de uma distribuição normal com uma média populacional (μ) de 193. Observa-se que todas as médias amostrais também se aproximam de 193, resultando em um histograma que exibe um formato semelhante ao de uma distribuição normal, caracterizado por uma curva sinoide, com a maioria dos valores agrupados em torno de 193.

A quase perfeita simetria do histograma é evidente, destacada pela proximidade entre os valores da média e mediana, que são 193.1 e 192.9, respectivamente. Essa simetria sugere uma distribuição balanceada das amostras em torno da média populacional.

Para realizar a análise dos Intervalos de Confiança (ICs), utilizamos o software R para otimizar a construção desses intervalos com base nos dados gerados e na distribuição t de Student (considerando variância desconhecida). O objetivo foi determinar quantos desses intervalos continham o valor da média populacional. Detalhes sobre os procedimentos adotados estão disponíveis no Apêndice A, enquanto neste relatório nos concentraremos na análise dos resultados.

Ao executar a análise, identificamos que 945 dos intervalos de confiança gerados continham a média populacional. Essa informação é crucial para avaliar a eficácia dos intervalos de confiança no encapsulamento do verdadeiro valor da média em uma proporção significativa das amostras. Esses resultados fornecem insights valiosos sobre a precisão das estimativas geradas a partir das amostras e podem ser fundamentais para inferências mais amplas sobre a distribuição da média populacional.

O intervalo de confiança é uma ferramenta essencial para quantificar a incerteza em torno de uma estimativa pontual. O coeficiente de confiança desse intervalo representa a probabilidade de conter o verdadeiro valor do parâmetro. Em outras palavras, ao escolher um intervalo de confiança com uma confiança de 95%, estamos expressando a expectativa de que, ao repetir o processo de amostragem muitas vezes e calcular os intervalos correspondentes, aproximadamente 95% deles irão incluir o verdadeiro valor do parâmetro. Esse conceito é fundamental para avaliar a robustez das estimativas e proporciona uma medida quantitativa da confiabilidade das inferências realizadas a partir da amostragem.

No contexto do nosso estudo, ao escolher um coeficiente de confiança de 95%, esperávamos que aproximadamente 95% dos intervalos de confiança contivessem a média populacional. Os resultados confirmaram essa expectativa, uma vez que constatamos que 94,5% dos intervalos possuíam essa característica. O código disponível no anexo permite reproduzir essa atividade e validar que, para o coeficiente de confiança selecionado, a porcentagem de intervalos que englobam a média populacional se aproxima consistentemente de 95

Para oferecer uma visualização mais intuitiva, selecionamos aleatoriamente 100

intervalos de confiança e os representamos graficamente. Cada intervalo é representado por uma barra, enquanto o ponto vermelho dentro de cada barra indica a média amostral correspondente. No segundo gráfico, adicionamos uma linha para destacar o valor da média populacional, simplificando a identificação dos intervalos que contêm esse valor.

É notável observar que, em vários momentos, a média da amostra (ponto vermelho) se aproxima da média verdadeira (linha verde), destacando a eficácia dos intervalos de confiança em proporcionar estimativas precisas da média populacional. Essa análise visual oferece uma compreensão mais tangível da distribuição das médias amostrais em relação à verdadeira média.

Figura 2: Gráfico de Barras com intervalos de confiança (IC) e marcadores para médias amostrais

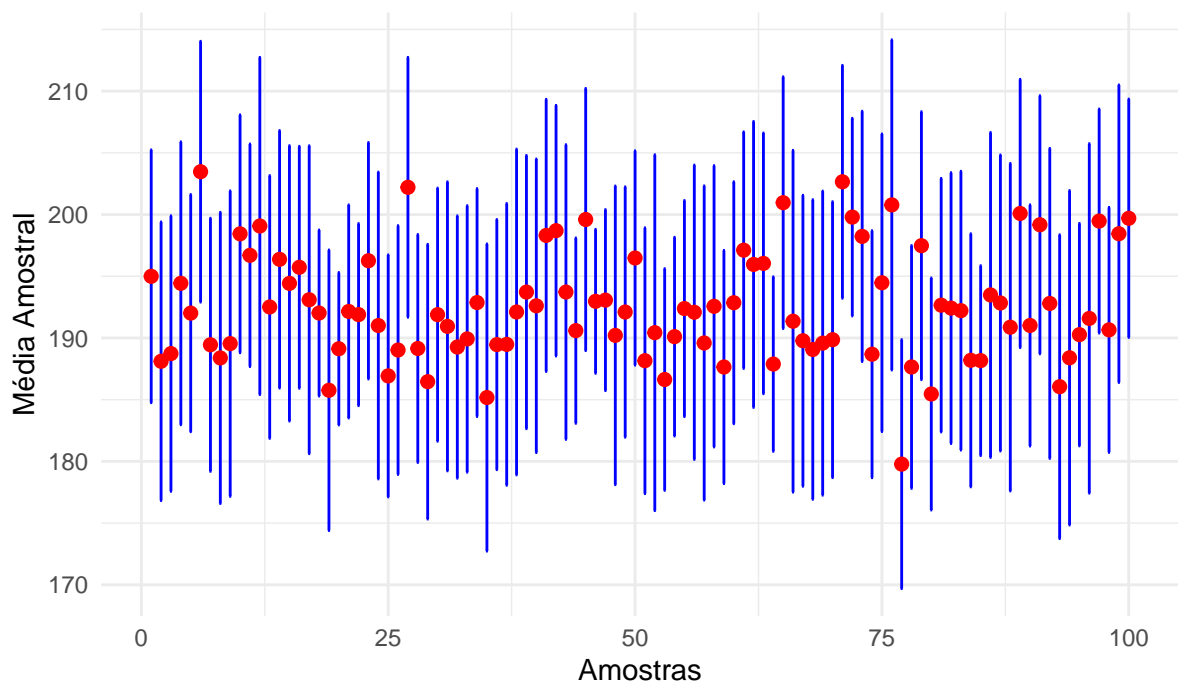
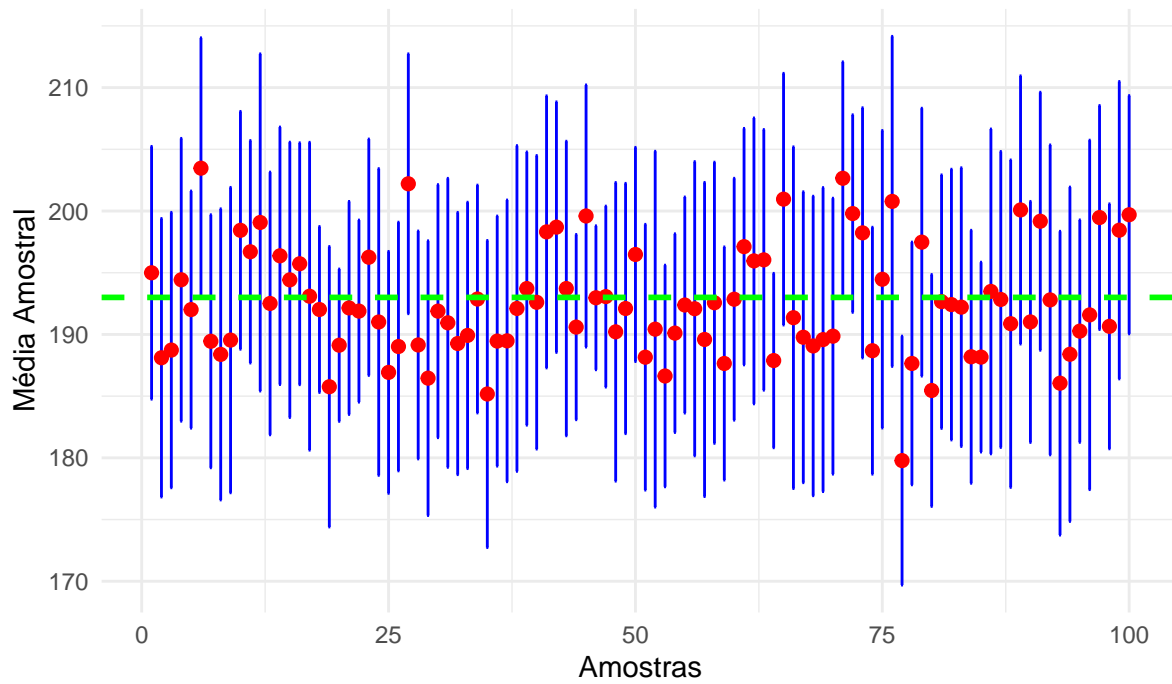


Figura 3: Gráfico de Barras com intervalos de confiança (IC) e marcadores para médias amostrais



Apêndice

```
caminho_lucas <- "resultados"

# Definindo o tamanho da amostra e o número de amostras
tamanho_amostra <- 15
numero_amostras <- 1000

# Gerando a média e o desvio padrão
media <- runif(1, min = 100, max = 200)
media <- as.integer(media)
cv <- 0.1 # 10% de coeficiente de variação
desvio_padrao <- media * cv

# Inicializando uma matriz para armazenar as amostras
amostras <- matrix(nrow = numero_amostras, ncol = tamanho_amostra)

# Gerando as amostras
for (i in 1:numero_amostras) {
  amostras[i,] <- rnorm(tamanho_amostra, mean = media, sd = desvio_padrao)
}

# Função para calcular o intervalo de confiança de 95%
calculate_ci <- function(x) {
  n <- length(x)
  media <- mean(x)
  desvio_padrao <- sd(x)
  erro_padrao <- desvio_padrao / sqrt(n)
  margem_erro <- qt(0.975, df = n - 1) * erro_padrao
  limite_inferior <- media - margem_erro
  limite_superior <- media + margem_erro
  return(c(media, desvio_padrao, limite_inferior, limite_superior))
}

# Calculando média, desvio padrão e intervalo de confiança para cada amostra
resultados <- apply(amostras, 1, calculate_ci)
```

```
# Convertendo a matriz de resultados em um data frame
resultados_df <- data.frame(t(resultados))

# Nomeando as colunas
colnames(resultados_df) <- c("Media_Amostral",
"Desvio_Padrao_Amostral",
"Limite_Inferior_CI",
"Limite_Superior_CI")

# Carregando a biblioteca para plotar gráficos
library(ggplot2)

# Medidas de posição
medidas_posicao <- summary(resultados_df$Media_Amostral)

# Criando um histograma
histograma <- ggplot(resultados_df, aes(x = Media_Amostral)) +
  geom_histogram(binwidth = 2, fill = "blue", color = "black", alpha = 0.7) +
  labs(x = "Média Amostral",
       y = "Frequência") +
  theme_minimal()
ggsave(filename = file.path(caminho_lucas, "histograma.pdf"),
width = 158,
height = 93,
units = "mm")

histograma
# Exibindo medidas de posição e histograma
print("Medidas de Posição:")
print(medidas_posicao)

print("Histograma:")
print(histograma)

# Contando quantos intervalos de confiança contêm a média populacional
contagem_contem_media <- sum(
resultados_df$Limite_Inferior_CI <= media &
resultados_df$Limite_Superior_CI >= media)
```



```
print(paste("Número de intervalos de confiança que contêm a média populacional:",
contagem_contem_media))

# Selecionando aleatoriamente 100 intervalos de confiança
amostra_aleatoria <- resultados_df[sample(nrow(resultados_df), 100), ]

# Criando um gráfico de barras para os intervalos de confiança
grafico_intervalos_confianca <- ggplot(amostra_aleatoria,
aes(x = 1:nrow(amostra_aleatoria))) +
  geom_errorbar(aes(ymin = Limite_Inferior_CI,
ymax = Limite_Superior_CI),
width = 0.2,
color = "blue") +
  geom_point(aes(y = Media_Amostral),
color = "red",
size = 2) +
  labs(title = "Intervalos de Confiança (100 amostras aleatórias)",
x = "Amostras",
y = "Média Amostral") +
  theme_minimal()
ggsave(filename = file.path(caminho_lucas, "intervalo_Confiança(100).pdf"),
width = 158,
height = 93,
units = "mm")

# Exibindo o gráfico
print(grafico_intervalos_confianca)

# Criando um gráfico de barras para os intervalos de confiança
grafico_intervalos_confianca2 <- ggplot(amostra_aleatoria,
aes(x = 1:nrow(amostra_aleatoria))) +
  geom_errorbar(aes(
ymin = Limite_Inferior_CI,
ymax = Limite_Superior_CI),
width = 0.2,
color = "blue") +
```

```
geom_point(
  aes(y = Media_Amostral),
  color = "red",
  size = 2) +
geom_hline(yintercept = media, linetype = "dashed",
  color = "green",
  size = 1) + # Adicionando linha para a média populacional
labs(title = "Intervalos de Confiança (100 amostras aleatórias)",
  x = "Amostras",
  y = "Média Amostral") +
theme_minimal()
ggsave(filename = file.path(caminho_lucas,
  "Intervalo_Confianca_mediaV.pdf"),
width = 158,
height = 93,
units = "mm")

# Exibindo o gráfico
print(grafico_intervalos_confianca2)
```