

Exercicio09-09-2024

higor lucas de Araujo Freitas

2024-09-08

```
knitr::opts_chunk$set(echo = TRUE)
```

Probabilidade

1 - Uma caixa contém 20 canetas pretas e 30 canetas verdes. Calcule a probabilidade de tirar uma caneta que não seja verde.

```
canetas_pretas = 20
canetas_verdes = 30

total_canetas = canetas_pretas + canetas_verdes

canetas_nao_verdes = canetas_pretas / total_canetas

sprintf("A probabilidade de não sair canetas verdes são : %s", canetas_nao_verdes)
```

```
## [1] "A probabilidade de não sair canetas verdes são : 0.4"
```

2 - Em um grupo de 200 pessoas, 80 jogam xadrez, 120 jogam damas, e 30 jogam ambos os jogos. Calcule a probabilidade de uma pessoa escolhida aleatoriamente jogar pelo menos um dos jogos (xadrez ou damas)

```
pessoas_xadrez = 80
pessoas_damas = 120

uniao = (80 + 120) - 30

total_pessoas = 200

probabilidade = uniao / total_pessoas

sprintf("A probabilidade de sair uma pessoa que joga ambos os jogos são : %s", probabilidade)
```

```
## [1] "A probabilidade de sair uma pessoa que joga ambos os jogos são : 0.85"
```

3 - De todos os funcionários de uma empresa, 70% usam o transporte público, 50% participam de um programa de carona compartilhada, e 30% usam ambos. Se um funcionário é escolhido ao acaso, qual a probabilidade de que ele participe do programa de carona compartilhada, dado que já sabemos que ele usa o transporte público?

```

transporte_publico <- 0.7
transporte_compartilhado <- 0.5
ambos <- 0.3

result <- ambos / transporte_publico

sprintf("A probabilidade de sair uma pessoa que já sabemos que ele usa o transporte público é : %s", r

```

```
## [1] "A probabilidade de sair uma pessoa que já sabemos que ele usa o transporte público é : 0.428571428571429"
```

4 - Em um estudo clínico, um novo teste para detectar uma doença rara foi aplicado a um grupo de 1000 pessoas. Dentre estas, sabemos que 10 pessoas realmente têm a doença. O teste diagnosticou corretamente 9 dessas 10 pessoas como positivas para a doença (verdadeiro positivo) e identificou incorretamente 40 pessoas saudáveis como positivas para a doença (falso positivo). Com base nesses dados experimentais, calcule a probabilidade de uma pessoa realmente ter a doença dado que ela recebeu um resultado positivo no teste.

Dados Observados: Número total de pessoas testadas: 1000 Pessoas com a doença: 10 Verdadeiros positivos: 9 Falsos positivos: 40

```

total_pessoas_doenca <- 1000
pessoa_com_doenca <- 10
falso_positivo <- 40
verdadeiro_positivo <- 9

#P é Probabilidade
P_pessoa_doente <- pessoa_com_doenca/total_pessoas_doenca
P_pessoa_sem_doenca <- (total_pessoas_doenca - pessoa_com_doenca) / total_pessoas_doenca
P_teste_positivo_doenca_verdadeiro <- verdadeiro_positivo/pessoa_com_doenca
P_teste_positivo_sem_doenca_verdadeiro <- falso_positivo / (total_pessoas_doenca - pessoa_com_doenca)

#Probabilidade do test ser positivo
forumula_P_B <- function(P_OR_B_A,P_A,P_OR_B_A_NEGATIVO,P_A_NEGATIVO){
  return((P_OR_B_A*P_A)+(P_OR_B_A_NEGATIVO*P_A_NEGATIVO))
}

#Probabilidade de ser realmente positivo
formula_bayes <- function(P_OR_B_A,P_A,P_B){
  return ((P_OR_B_A * P_A ) / P_B)
}

P_teste_ser_positivo <- forumula_P_B(P_teste_positivo_doenca_verdadeiro,P_pessoa_doente,P_teste_positivo_sem_doenca_verdadeiro)

P_ter_doenca_test_positivo <- formula_bayes(P_pessoa_doente,P_teste_positivo_doenca_verdadeiro,P_teste_ser_positivo)
(P_pessoa_doente * P_teste_positivo_doenca_verdadeiro ) / P_teste_ser_positivo

```

```
## [1] 0.1836735
```

```

sprintf("A probabilidade de sair uma pessoa que realmente tem a doença dado que ela recebeu um resultado

```

```
## [1] "A probabilidade de sair uma pessoa que realmente tem a doença dado que ela recebeu um resultado"
```

Parte 2 - Distribuição de Probabilidades

Objetivo: Explorar a forma das distribuições das médias amostrais para diferentes tipos de distribuições de populações. Investigar como a distribuição da média amostral se comporta para diferentes tipos de distribuições populacionais.

1) Gerar uma população com 10.000 valores seguindo uma distribuição normal com média $\mu=50$ e desvio padrão $\sigma=10$

```
library(ggplot2)
# 1 Gerar uma população com 10.000 valores seguindo uma distribuição normal com média  $\mu=50$  e desvio padrão  $\sigma=10$ 
set.seed(123) # Para garantir a reprodutibilidade
populacao <- rnorm(10000, mean = 50, sd = 10)
head(populacao, 10)
```

```
## [1] 44.39524 47.69823 65.58708 50.70508 51.29288 67.15065 54.60916 37.34939
## [9] 43.13147 45.54338
```

2 Gerar uma população com 10.000 valores seguindo uma distribuição uniforme entre 20 e 80

```
populacao_2 <- runif(10000, min = 20, max = 80)
head(populacao_2, 10)
```

```
## [1] 79.46740 38.13384 46.02554 29.63125 69.38160 32.48543 37.09796 52.77253
## [9] 55.34272 46.53519
```

3 Gerar uma população com 10.000 valores seguindo uma distribuição exponencial com taxa $\lambda=1/30$

```
populacao_3 <- rexp(10000, rate = 1/30)
head(populacao_3, 10)
```

```
## [1] 74.571549 16.776689 24.534494 2.981942 58.294476 31.527195 20.078604
## [8] 32.476253 13.925390 52.637491
```

4 Para cada uma das populações acima, extrair 1.000 amostras, cada uma com 50 observações

```
numero_amostras <- 1000
tamanho_amostra <- 50

extrairAmostra <- function(populacao, numero_amostras, tamanho_amostra){
  guardar_amostras <- matrix(nrow = numero_amostras, ncol = tamanho_amostra)

  for(i in 1:numero_amostras) {
    guardar_amostras[i, ] <- sample(populacao, tamanho_amostra, replace = TRUE)
  }

  return(guardar_amostras)
}
```

```

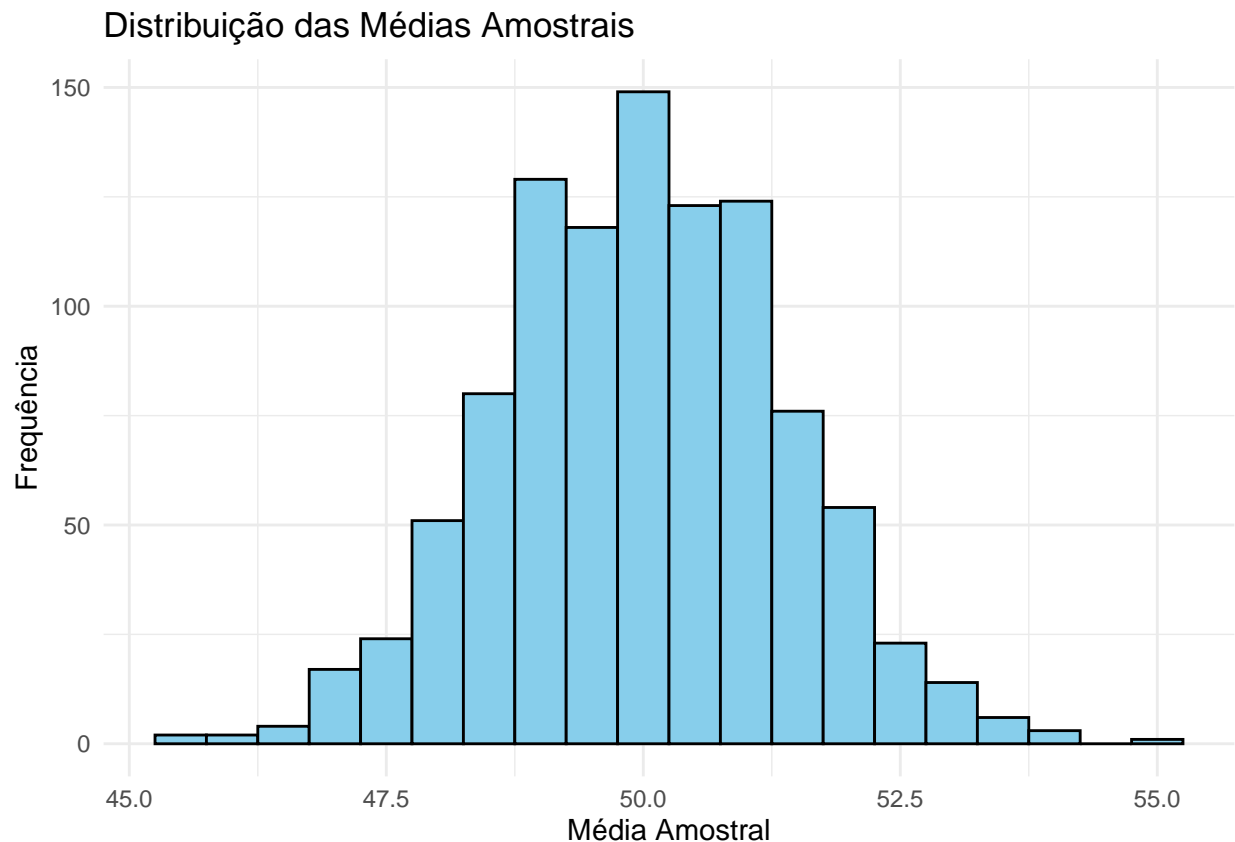
amostra_1 <- extrairAmostra(populacao,numero_amstras,tamanho_amostra)
amostra_2 <- extrairAmostra(populacao_2,numero_amstras,tamanho_amostra)
amostra_3 <- extrairAmostra(populacao_3,numero_amstras,tamanho_amostra)

# 5 Calcular a média de cada uma das 1.000 amostras para cada tipo de população.

calculo_media_1 <- replicate(numero_amstras, mean(sample(populacao, tamanho_amostra, replace = TRUE)))
calculo_media_2 <-replicate(numero_amstras, mean(sample(populacao_2, tamanho_amostra, replace = TRUE)))
calculo_media_3 <-replicate(numero_amstras, mean(sample(populacao_3, tamanho_amostra, replace = TRUE)))

# 6 Plotar a distribuição das médias amostrais para cada tipo de população.
ggplot(data.frame(Medias = calculo_media_1), aes(x = Medias)) +
  geom_histogram(binwidth = 0.5, fill = "skyblue", color = "black") +
  theme_minimal() +
  ggtitle("Distribuição das Médias Amostrais") +
  xlab("Média Amostral") +
  ylab("Frequência")

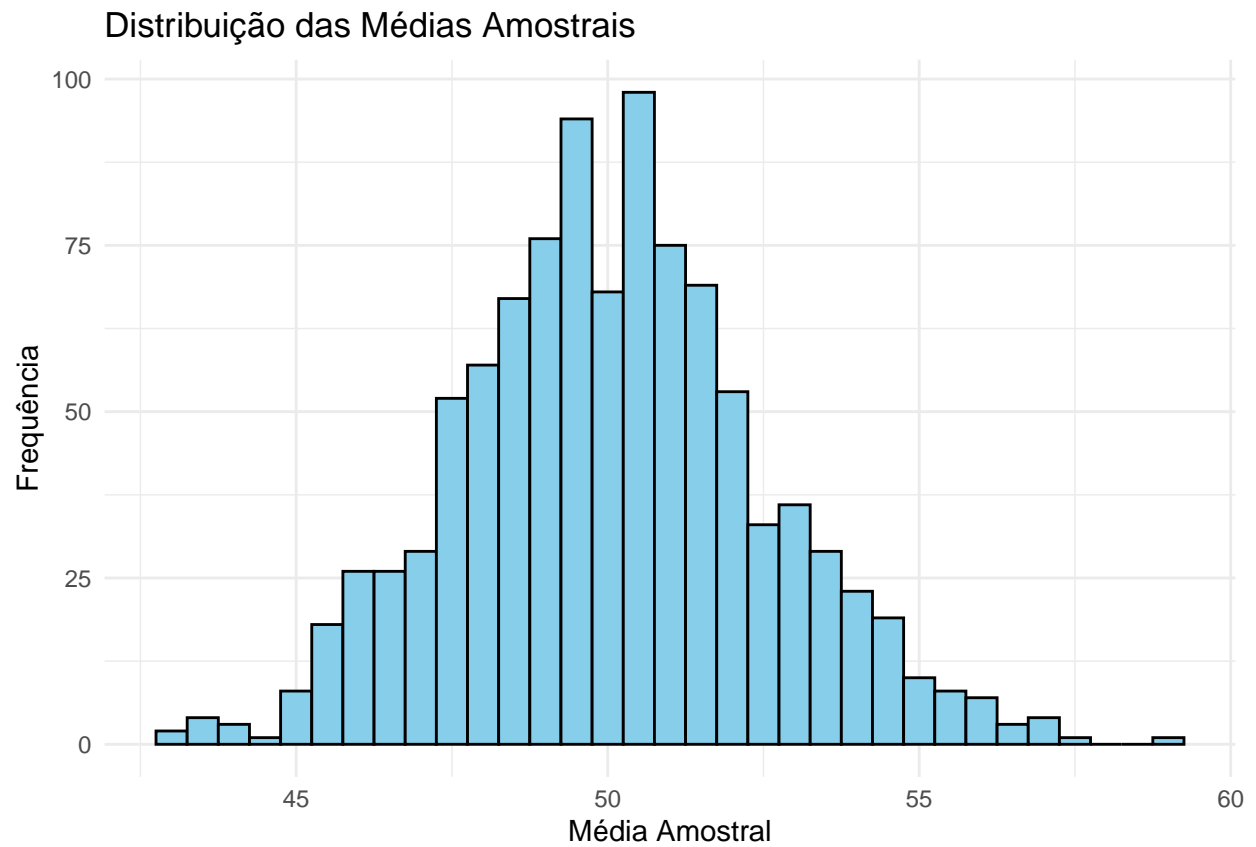
```



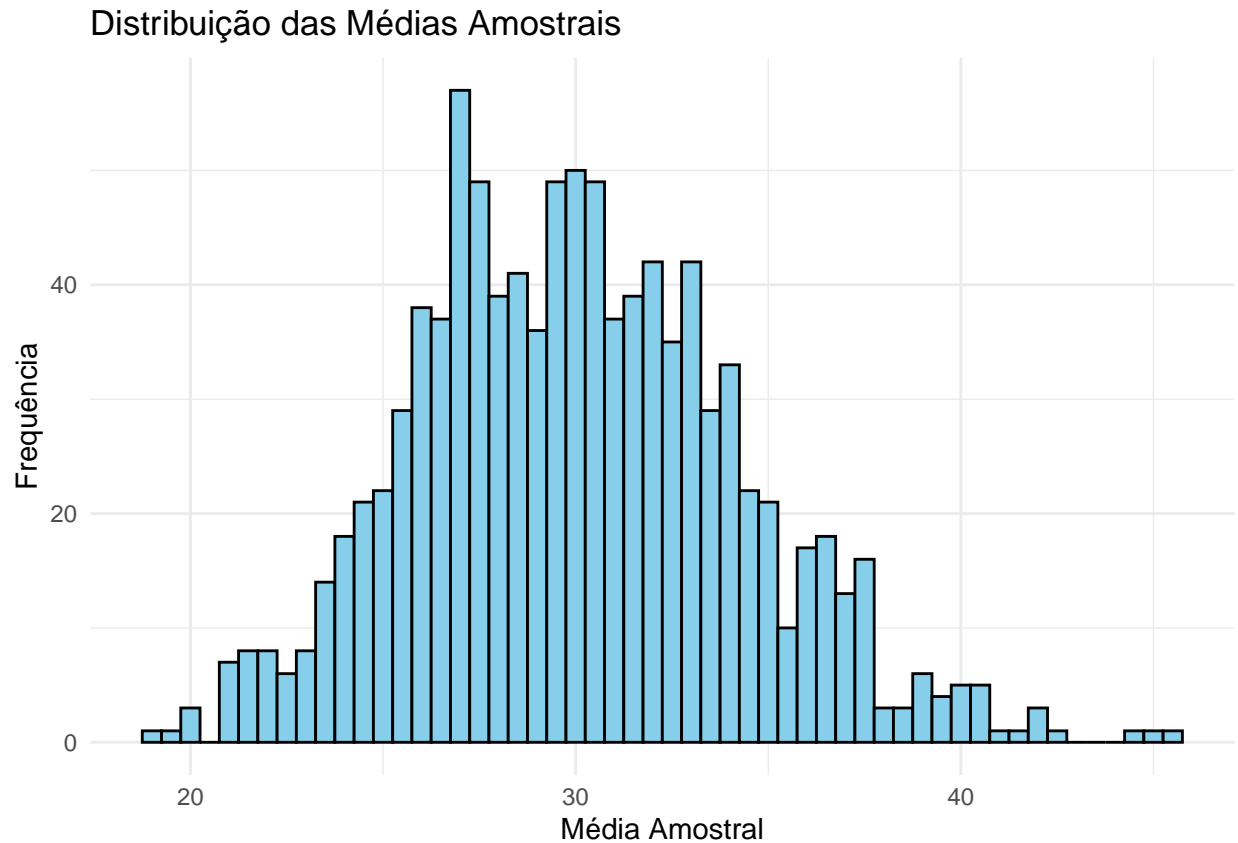
```

ggplot(data.frame(Medias = calculo_media_2), aes(x = Medias)) +
  geom_histogram(binwidth = 0.5, fill = "skyblue", color = "black") +
  theme_minimal() +
  ggtitle("Distribuição das Médias Amostrais") +
  xlab("Média Amostral") +
  ylab("Frequência")

```



```
ggplot(data.frame(Medias = calculo_media_3), aes(x = Medias)) +  
  geom_histogram(binwidth = 0.5, fill = "skyblue", color = "black") +  
  theme_minimal() +  
  ggtitle("Distribuição das Médias Amostrais") +  
  xlab("Média Amostral") +  
  ylab("Frequência")
```



7 Discutir como a forma da distribuição da média amostral se relaciona com a forma da distribuição da população original e como isto se relaciona com o Teorema Central do Limite

De acordo com o TCL (Teorema central do limite) afirma que a soma de N variáveis aleatórias independentes, com qualquer distribuição e variâncias semelhantes, é uma variável com distribuição que se aproxima da distribuição de Gauss quando N aumenta.

Isso quer dizer que quanto mais dados a amostra obtiver a média dela sempre tende a ficar numa normalidade, no seu maior pico.