

Lista 03 - Simulações

Artur Damião

2 de outubro de 2025

Número USP: 10701251. E-mail: arturcardoso@usp.br

1 Começando a simular dados

1.

```
?rnorm  
?rbinom
```

A função `rnorm` refere-se à **distribuição normal**, ao passo que a função `rbinom` refere-se à uma **distribuição binomial**. Na primeira, os principais parâmetros são: o número de observações a serem geradas, a média da distribuição (`mean()`, com *default* igual a 0) e o desvio padrão (`sd()`, com *default* igual a 1). Na segunda, os principais parâmetros são o número de observações, o número de tentativas em cada observação (`size()`) e a probabilidade de sucesso (`prob()`).

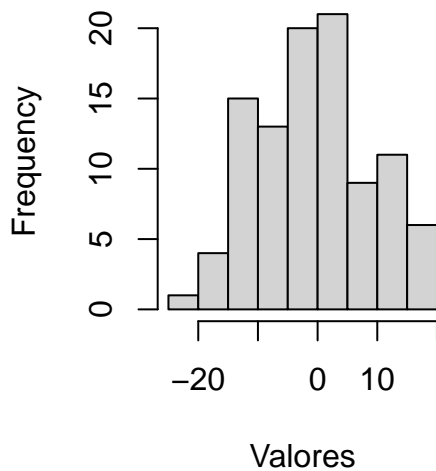
2.

Para a **distribuição normal**, a $\mathbb{E} = 0$ e a variância é o σ ao quadrado, ou seja, 100. Para a **distribuição binomial**, a $\mathbb{E}(X)$ é calculada a a partir de $n \times p$, logo é 70. A variância é calculada como $n \times p \times (1 - p)$, ou seja, 21.

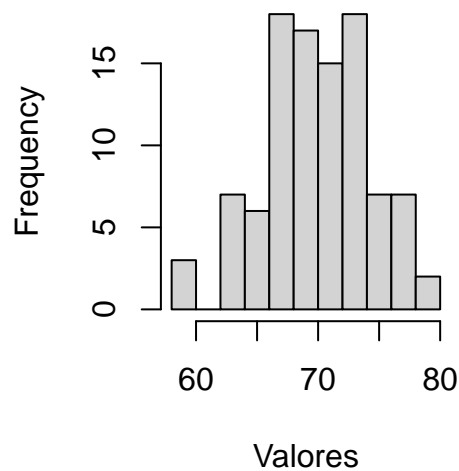
3.

```
vetor_normal <- rnorm(n = 100, mean = 0, sd = 10)  
vetor_binomial <- rbinom(n = 100, size = 100, prob = 0.7)  
  
# Plotar lado a lado  
par(mfrow = c(1, 2))  
  
hist(vetor_normal, main = "Histograma da Normal", xlab = "Valores")  
hist(vetor_binomial, main = "Histograma da Binomial", xlab = "Valores")
```

Histograma da Normal



Histograma da Binomial



4.

```
media_normal_1 <- mean(vetor_normal)
var_normal_1 <- var(vetor_normal)

media_binomial_1 <- mean(vetor_binomial)
var_binomial_1 <- var(vetor_binomial)
```

A média amostral da distribuição normal é -0.62 e a variância é 90.41. A média amostral da distribuição binomial é 70.41, enquanto a variância é 19.8. Os valores não são exatamente iguais, porque a **Lei dos Grandes Números** ($n = 100$, no nosso caso) coloca que os valores amostrais tendem a se convergir para os valores da população à medida que a amostra aumenta.

5.

```
vetor_normal_2 <- rnorm(n = 100, mean = 0, sd = 10)
vetor_binomial_2 <- rbinom(n = 100, size = 100, prob = 0.7)

media_normal_2 <- mean(vetor_normal_2)
var_normal_2 <- var(vetor_normal_2)

media_binomial_2 <- mean(vetor_binomial_2)
var_binomial_2 <- var(vetor_binomial_2)
```

A nova média amostral da distribuição normal é -0.33 e a variância é 132.13. A nova média amostral da distribuição binomial é 69.17 e a variância é 25.17. Os valores não são iguais aos teóricos nem aos valores do item anterior porque a amostragem é aleatória (variabilidade). Cada amostra terá sua própria média e variância.

2 Simulando a Normal múltiplas vezes

1.

```
medias_10 <- numeric(10)
for (i in 1:10) {
  amostra <- rnorm(n = 10, mean = 0, sd = 1)
  medias_10[i] <- mean(amostra)
}

print(medias_10)
```

```
[1] -0.37571014  0.14337809 -0.51182327  0.18031924  0.16980822  0.27952144
[7]  0.43146362 -0.05162519 -0.46679790  0.19207058
```

2.

```
medias_50 <- numeric(50) # Vetor vazio
for (i in 1:50) {
  amostra <- rnorm(n = 10, mean = 0, sd = 1)
  medias_50[i] <- mean(amostra)
}

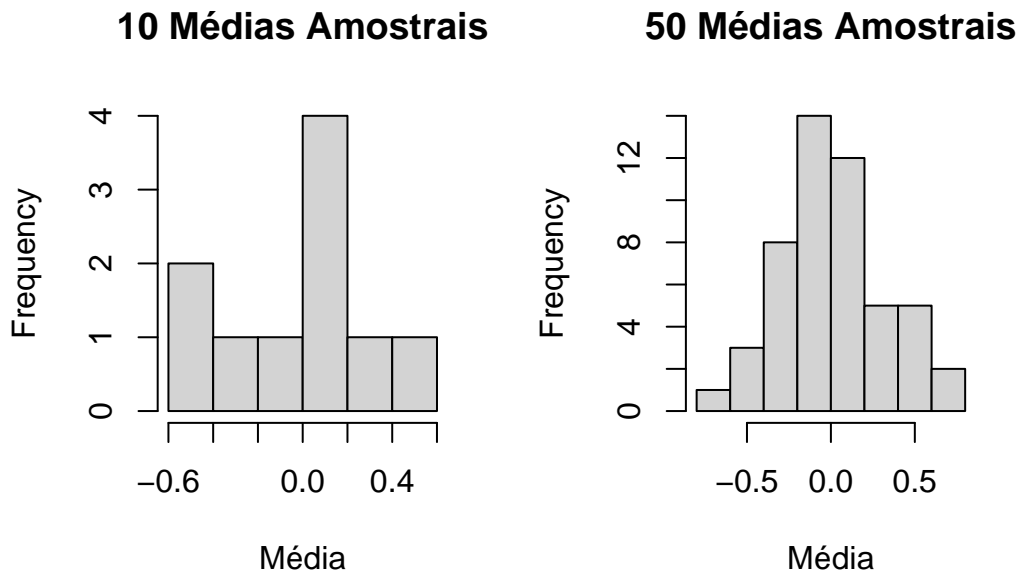
print(medias_50)
```

```
[1] -0.01714487  0.52115330  0.13578916 -0.05866797  0.01939570 -0.43690491
[7] -0.06110792 -0.19506101 -0.41428892 -0.06057026  0.25594689 -0.19053654
[13]  0.42936602 -0.54372720  0.06483393  0.11431034  0.61784587 -0.17645259
[19]  0.07276221 -0.29814585  0.33993158 -0.17688038  0.30139546  0.48048308
[25] -0.21097242 -0.07912908  0.02113591  0.04847994  0.07521042 -0.06105894
[31] -0.17327730  0.20531714 -0.29556364  0.13080520 -0.02114567  0.06696994
[37] -0.26806906 -0.63123199 -0.27945055  0.41324308  0.08280845 -0.02196960
[43] -0.22962305 -0.33902530  0.20557123 -0.20647884  0.10424058  0.73803863
[49]  0.45189008 -0.05258725
```

3.

```
par(mfrow = c(1, 2))

hist(medias_10, main = "10 Médias Amostras", xlab = "Média")
hist(medias_50, main = "50 Médias Amostras", xlab = "Média")
```



Cada histograma representa a média das distribuições amostrais. A média calculada é a média da média. A média da média amostral é a estimativa da média populacional (teoricamente 0).

4¹.

```
simulacoes <- c(10, 30, 50, 100, 1000)

lista_de_medias <- list()

for (n_sim in simulacoes) {
  vetor_temp <- numeric(n_sim)
  for (i in 1:n_sim) {
    vetor_temp[i] <- mean(rnorm(n = 10, mean = 0, sd = 1))
  }
  lista_de_medias[[as.character(n_sim)]] <- vetor_temp
}
```

5.

Para esta operação, utilizei a função `sapply()` para calcular todas as médias de uma vez e o pacote `tidyverse` para visualização.

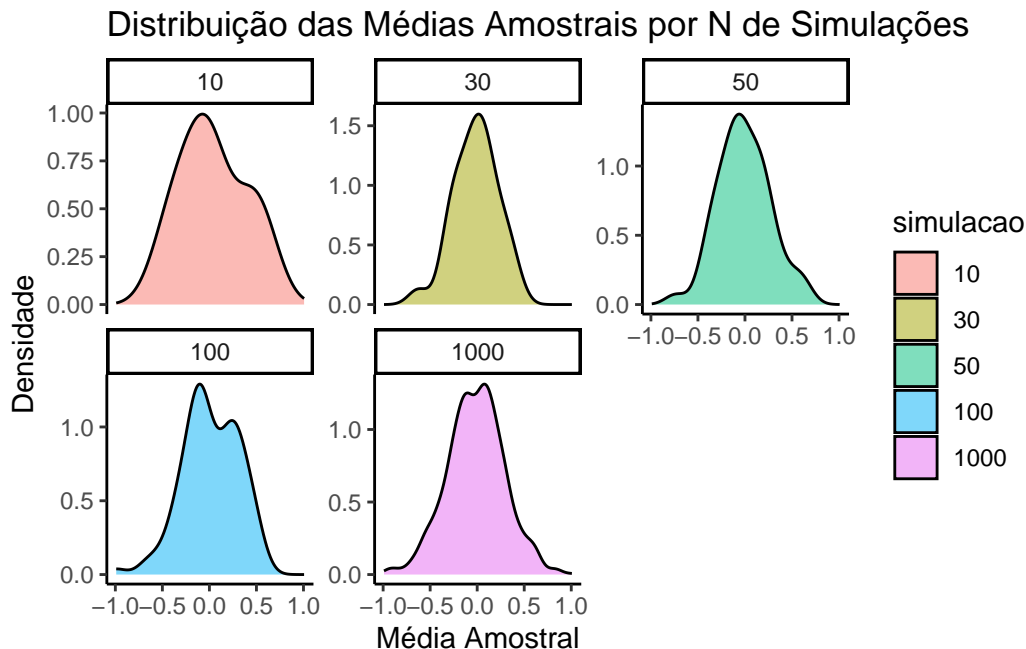
```
# Calculando médias
medias_das_simulacoes <- sapply(lista_de_medias, mean)

# Empilhando dados para fazer o plot
df_densidade <- enframe(lista_de_medias, name = "simulacao", value = "media") %>%
  unnest(cols = c(media)) %>%
  mutate(simulacao = factor(simulacao, levels = as.character(simulacoes)))

# Plotando gráfico de densidade
ggplot(df_densidade, aes(x = media, fill = simulacao)) +
  geom_density(alpha = 0.5) +
  facet_wrap(~simulacao, scales = "free_y") +
```

¹Para este exercício, consultei o ChatGPT. Inicialmente, criei um objeto para cada simulação, mas não gosto de criar muitos objetos no R. Como não consegui entender como deixar um código reduzido, recorri à IA.

```
labs(title = "Distribuição das Médias Amostrais por N de Simulações",
     x = "Média Amostral",
     y = "Densidade") +
theme_classic()
```



Percebe que, à medida que o número de simulações aumenta (de 10 para 1000), a média das médias amostrais tende a se aproximar de 0, que é a estimativa de média populacional mencionada no exercício 2.3. O gráfico com 1000 simulações é o que mais se aproxima de uma **distribuição normal**. É uma demonstração do **Teorema do Limite Central**.

3 Um pouco mais sobre variáveis aleatórias

1.

$U \sim Unif(0, 1)$ é uma distribuição uniforme com intervalo entre 0 e 1. A função no R é `runif()`². $\mathbb{E}(U) = 0,5$.

$B \sim Binom(100, 0.6)$ é uma distribuição binomial com 100 tentativas e probabilidade de sucesso de 0.6. $\mathbb{E}(B) = 60$.

2.

```
iteracoes <- c(10, 100, 1000)
lista_medias_unif <- list()
lista_medias_binom <- list()

# distribuição Uniforme
for (n_iter in iteracoes) {
  medias_temp <- numeric(n_iter)
  for (i in 1:n_iter) {
```

²Consultei a família de distribuições do pacote ‘{stats}’.

```

    medias_temp[i] <- mean(runif(n = 100, min = 0, max = 1))
  }
  lista_medias_unif[[as.character(n_iter)]] <- medias_temp
}

# Loop para a distribuição Binomial
for (n_iter in iteracoes) {
  medias_temp <- numeric(n_iter)
  for (i in 1:n_iter) {
    medias_temp[i] <- mean(rbinom(n = 100, size = 100, prob = 0.6))
  }
  lista_medias_binom[[as.character(n_iter)]] <- medias_temp
}

```

3.

Calcular as médias:

```

# Médias uniformes
sapply(lista_medias_unif, mean)

```

```

      10      100     1000
0.4943006 0.5010807 0.5013651

```

```

# Médias Binomial
sapply(lista_medias_binom, mean)

```

```

      10      100     1000
60.03400 59.94130 59.96548

```

Plotando Histogramas:

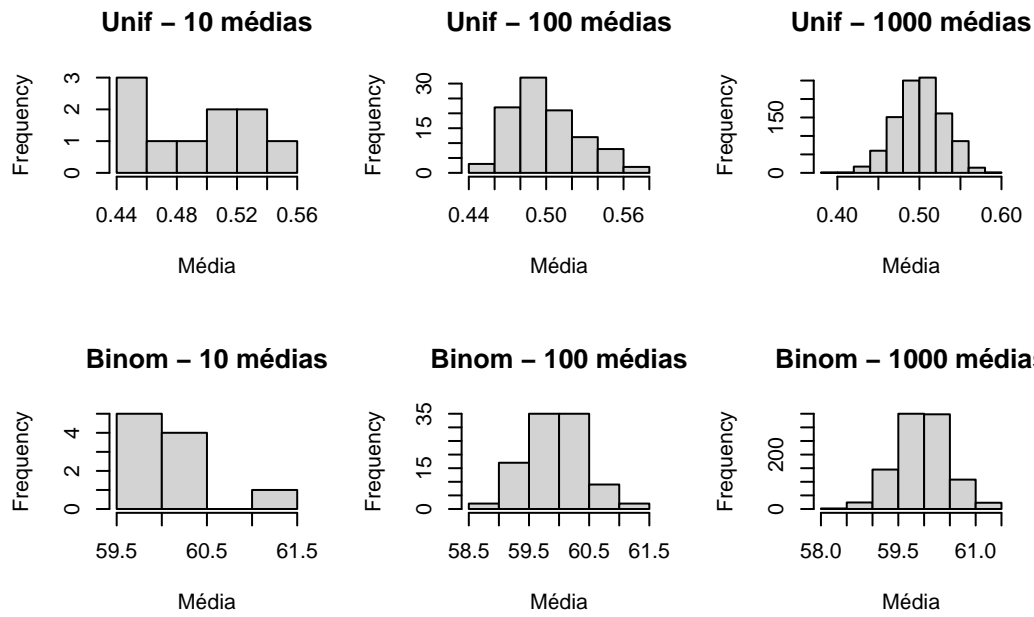
```

par(mfrow = c(2, 3))

# Histogramas para a Uniforme
for (name in names(lista_medias_unif)) {
  hist(lista_medias_unif[[name]], main = paste("Unif -", name, "médias"), xlab = "Média")
}

# Histogramas para a Binomial
for (name in names(lista_medias_binom)) {
  hist(lista_medias_binom[[name]], main = paste("Binom -", name, "médias"), xlab = "Média")
}

```



A maioria dos histogramas aparentam se aproximar de uma distribuição normal, mesmo sendo distribuições de tipos diferentes.