

Exercício 6 - Projeto Computacional PE 2022

Diogo Gaspar, 99207

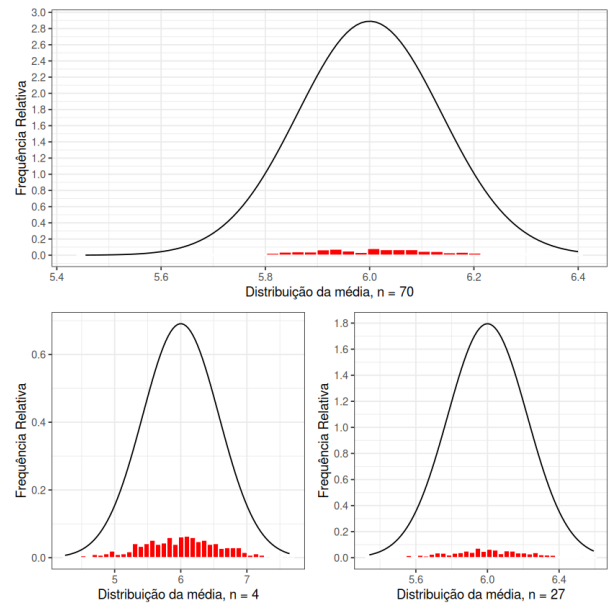
Consideremos como premissas que foram fixadas uma semente em 1896, três tamanhos de amostras (respetivamente 4, 27, e 70) e um intervalo contínuo [4,8]. O objetivo deste exercício passa por gerar 500 amostras com distribuição uniforme contínua para cada tamanho supra-mencionado. De seguida, calcular a média de cada uma das amostras (obtendo, assim, os valores da distribuição da média associada às amostras de cada tamanho), e ilustrar através de um histograma de frequência relativa, a relação entre os valores obtidos para a distribuição das médias e a distribuição normal das mesmas, considerando valor esperado $E(X)$ e a variância $Var(X)/n$.

$$E(X) = \frac{(4 + 8)}{2}$$

$$\forall_{n \in \{4, 27, 70\}}, \quad \sigma^2 = \frac{Var(X)}{n} \leftrightarrow \sigma = \sqrt{Var(X)/n} = \sqrt{\frac{(8 - 4)^2}{12n}}$$

Para tal, recorreu-se ao seguinte trecho de código R (utilizando as bibliotecas `ggplot2`, `gridExtra` e `purrr`):

```
1  set.seed(1896)
2  n_values <- c(4, 27, 70)
3  sample_amount <- 500
4  a <- 4
5  b <- 8
6  expected_value <- (b + a) / 2
7
8  calc_sd <- function(n) {
9    return (sqrt(((b - a)**2) / (n * 12)))
10 }
11 calc_means <- function(n) {
12   means <- c()
13   for (i in 1:sample_amount) {
14     means <- c(means, mean(runif(n, 4, 8)))
15   }
16   return (means)
17 }
18
19 calc_plot <- function(n) {
20   means <- calc_means(n)
21   df <- data.frame(means)
22   ggplot(df, aes(x = means)) +
23     geom_histogram(aes(y=after_stat(count / sum(count))), color="white", fill="red", bins=40) +
24     stat_function(fun=dnorm, args=list(mean=expected_value, sd=calc_sd(n))) +
25     theme_bw() +
26     labs(x = paste("Distribuição da média, n =", n), y = "Frequência Relativa") +
27     scale_y_continuous(breaks = seq(0, 5, .2))
28 }
29
30 plots <- map(n_values, calc_plot)
31 grid.arrange(grobs = plots, layout_matrix = matrix(c(3,1,3,2), nrow = 2))
```



Note-se, tal como esperado, que o pico da distribuição normal encontra-se, em todos os histogramas, no valor 6. Mais, o tamanho da amostra aparenta ser inversamente proporcional ao desvio padrão, e que o tamanho da amostra não parece ter particular impacto nas frequências relativas dos valores da distribuição média (sendo que todos, no seu pico, aproximam-se de 0.1 sem nunca lá chegar, tendo decaimentos suaves para a esquerda e direita do pico).