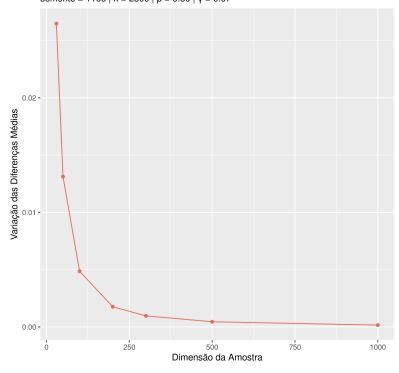
Consideremos que foi fixada uma semente igual a 1166. O objetivo deste exercício é o de gerar 2500 amostras de tamanho n, para cada $n \in \{30, 50, 100, 200, 300, 500, 1000\}$, de uma distribuição de Bernoulli com parâmetro igual a 0.5. De seguida, usar dois métodos distintos que calculam intervalos de confiança de aproximação para o parâmetro mencionado, o que permite obter a diferença entre as amplitudes desses intervalos. Por fim, calcular a média das 2500 diferenças para cada n. Para tal, recorreu-se ao seguinte trecho de código em \mathbb{R} :

```
library("ggplot2")
library("Rlab")
        SEED <- 1166
            AMPLE_COUNT
        BERNOULLI_P <- 0.5
CONF_LEVEL <- 0.97
N <- c(30, 50, 100
                                                             100, 200, 300, 500, 1000)
        set.seed(SEED)
       method 1 <- function(samples, conf level) {</pre>
11
                       len <- length(samples)
mean <- mean(samples)
13
                       z <- qnorm((1 + conf_level) / 2)
denom <- 2 * (1 + z**2 / len)
upper <- ((2 * mean + z**2 / len) + sqrt(4 * mean * z**2 * (1 - mean) / len + z**4 / len**2)
15
16
                                   denom
                                                       ((2 * mean + z**2 / len) - sqrt(4 * mean * z**2 * (1 - mean) / len + z**4 / len**2)
17
                                   denom
                       return(abs(upper - lower))
18
19
       }
20
       method_2 <- function(samples, conf_level) {
   len <- length(samples)
   mean <- mean(samples)</pre>
22
23
                       upper <- mean + qnorm(1 - (1 - conf_level) / 2) * sqrt(mean * (1 - mean) / lower <- mean - qnorm(1 - (1 - conf_level) / 2) * sqrt(mean * (1 - mean) /
24
25
                                                                                      - lower))
26
                       return(abs(upper
27
        df <- data.frame()</pre>
29
                      (n in N) {
method_diffs <- c()
for (i in 1:SAMPLE_COUNT)</pre>
         for (n in N)
31
                                     samples <- rbern(n, BERNOULLI_P)
diff <- method_2(samples, CONF_LEVEL) - me
method_diffs <- append(method_diffs, diff)
33
34
                                                                                                                                                                                        method_1(samples, CONF_LEVEL)
35
36
37
                       mean_diffs <- mean(method_diffs)</pre>
                       df <- rbind(df, data.frame(n = n, difference = mean_diffs))</pre>
38
40
       ggplot(df, aes(x = n, y = difference)) +
geom_line(color = "#e76f51") +
geom_point(color = "#e76f51") +
xlab("Dimensão da Amostra") +
ylab("Variação das Diferenças Médias")
labs(title = "Relação entre Variação da Diference Variação da Di
41
42
44
45
                                                       = "Relação entre Variação das Diferenças Médias e Dimensão da Amostra",
= sprintf("semente = %d | k = %d | p = %.2f | = %.2f",
46
                              SEED, SAMPLE_COUNT, BERNOULLI_P, CONF_LEVEL))
47
```

O gráfico obtido permite concluir que a relação entre as duas variáveis em causa é inversamente proporcional. Deste modo, à medida que o tamanho das 2500 amostras aumenta, os métodos 1 e 2 apresentam, em média, intervalos de confiança com uma amplitude cada vez mais próxima.

As diferenças médias foram obtidas subtraindo a amplitude do método 2 pela a do método 1 para cada amostra e posteriormente calculando a média dessas diferenças. Assim, já que todos os valores no eixo dos yy são positivos, conclui-se que, em geral, para amostras de dimensões mais pequenas o método 1 é mais favorável para aproximar o parâmetro p, pois o seu intervalo de confiança tem menor amplitude. Porém, para amostras de dimensões maiores, ambos os métodos garantem um

Relação entre Variação das Diferenças Médias e Dimensão da Amostra semente = 1166 | k = 2500 | p = 0.50 | y = 0.97



intervalo de confiança com amplitude semelhante.