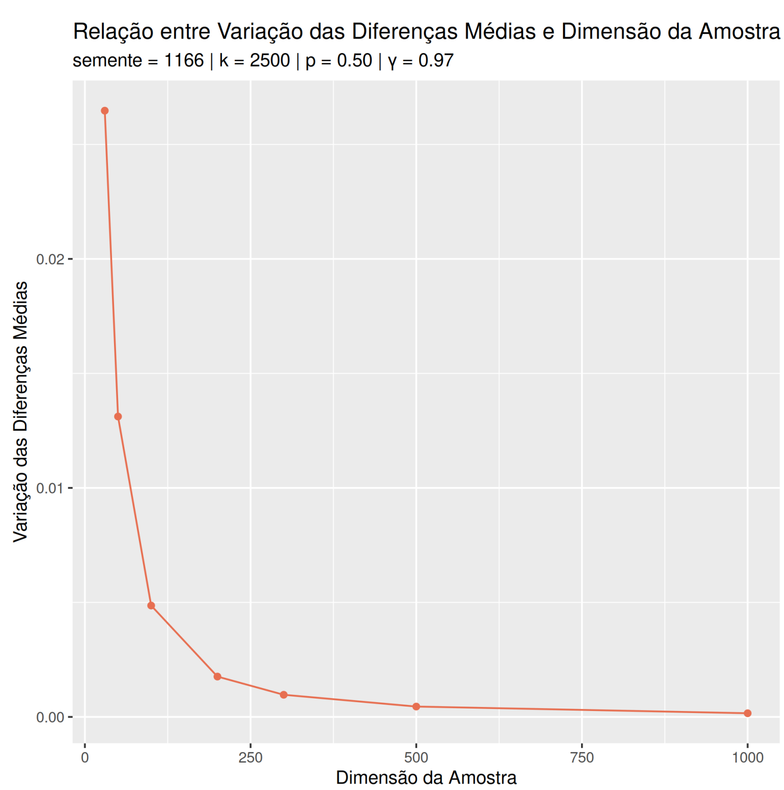
Consideremos que foi fixada uma semente igual a `1166`. O objetivo deste exercício é o de gerar `2500` amostras de tamanho n, para cada $n \in \{30, 50, 100, 200, 300, 500, 1000\}$, de uma distribuição de Bernoulli com parâmetro igual a `0.5`. De seguida, usar dois métodos distintos que calculam intervalos 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 `R`:

```r
library("ggplot2")
library("Rlab")

SEED <- 1166
SAMPLE_COUNT <- 2500
BERNOULLI_P <- 0.5
CONF_LEVEL <- 0.97
N <- c(30, 50, 100, 200, 300, 500, 1000)
set.seed(SEED)

method_1 <- function(samples, conf_level) {
    len <- length(samples)
    mean <- mean(samples)
    z <- qnorm((1 + conf_level) / 2)
    sols <- polyroot(c(mean**2, -2 * mean - z**2 / len, 1 + z**2 / len))
    return(abs(sols[2] - sols[1]))
}

method_2 <- function(samples, conf_level) {
    len <- length(samples)
    mean <- mean(samples)
    upper <- mean + qnorm(1 - (1 - conf_level) / 2) * sqrt(mean * (1 - mean) / len)
    lower <- mean - qnorm(1 - (1 - conf_level) / 2) * sqrt(mean * (1 - mean) / len)
    return(abs(upper - lower))
}

df <- data.frame()
for (n in N) {
    method_diffs <- c()
    for (i in 1:SAMPLE_COUNT) {
        samples <- rbern(n, BERNOULLI_P)
        diff <- method_2(samples, CONF_LEVEL) - method_1(samples, CONF_LEVEL)
        method_diffs <- append(method_diffs, diff)
    }
    mean_diffs <- mean(method_diffs)
    df <- rbind(df, data.frame(n = n, difference = mean_diffs))
}

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 das Diferenças Médias e Dimensão da Amostra",
    subtitle = sprintf("semente = %d | k = %d | p = %.2f |   = %.2f",
      SEED, SAMPLE_COUNT, BERNOULLI_P, CONF_LEVEL))
```



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 intervalo de confiança com amplitude semelhante.