

## case2

Aluno: Henrique Fuga Duran

```
populational_stdev <- function(data) {  
  mean <- mean(data)  
  square_sum <- sum((data - mean) ^ 2)  
  result <- sqrt(square_sum/length(data))  
  return(result)  
}  
  
sample_stdev <- function(data) {  
  mean <- mean(data)  
  square_sum <- sum((data - mean) ^ 2)  
  result <- sqrt(square_sum / (length(data) - 1))  
  return(result)  
}  
  
populational_var <- function(data) {  
  mean <- mean(data)  
  square_sum <- sum((data - mean) ^ 2)  
  result <- square_sum/length(data)  
  return(result)  
}  
  
sample_var <- function(data) {  
  mean <- mean(data)  
  square_sum <- sum((data - mean) ^ 2)  
  result <- square_sum / (length(data) - 1)  
  return(result)  
}  
  
library(ggplot2)
```

```

df <- readRDS("cases/spotify_modificada.rds")

samples_measures <- data.frame(
  mean = numeric(),
  var = numeric()
)

K <- 200 # size of the sample
M <- 450 # number of samples

for(i in 1:M) {
  sample_indexes <- sample(nrow(df), K, replace=TRUE)
  new_sample <- df[sample_indexes[1:K],]
  mean <- mean(new_sample$duration_ms)
  var <- sample_var(new_sample$duration_ms)
  new_row <- data.frame(mean = mean, var = var)
  samples_measures <- rbind(samples_measures, new_row)
}

mean_of_sample_means <- mean(samples_measures$mean)
mean_of_sample_vars <- mean(samples_measures$var)

population_mean <- mean(df$duration_ms)
population_var <- populational_var(df$duration_ms)

display_means_info <- function() {
  print(paste("Mean of Sample Means:", mean_of_sample_means))
  print(paste("Population Mean:", population_mean))

  difference <- mean_of_sample_means - population_mean
  print(paste("Difference:", round(difference, digits = 2)))

  error <- (mean_of_sample_means - population_mean) / population_mean
  print(paste("Error:", round(error * 100, digits = 1), "%"))
}

display_vars_info <- function() {
  print(paste("Mean of Sample Vars:", mean_of_sample_vars))

```

```

print(paste("Population Var:", population_var))

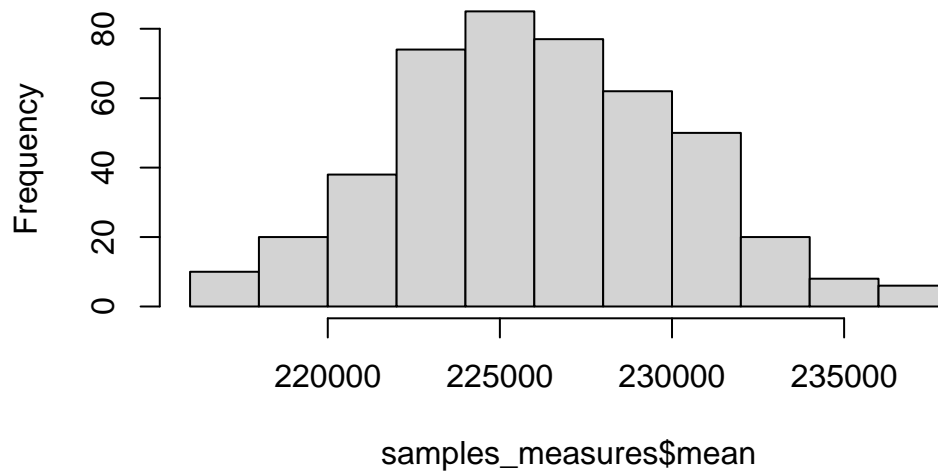
difference <- mean_of_sample_vars - population_var
print(paste("Difference:", round(difference, digits = 2)))

error <- (mean_of_sample_vars - population_var) / population_var
print(paste("Error:", round(error * 100, digits = 1), "%"))
}

hist(samples_measures$mean)

```

**Histogram of samples\_measures\$mean**

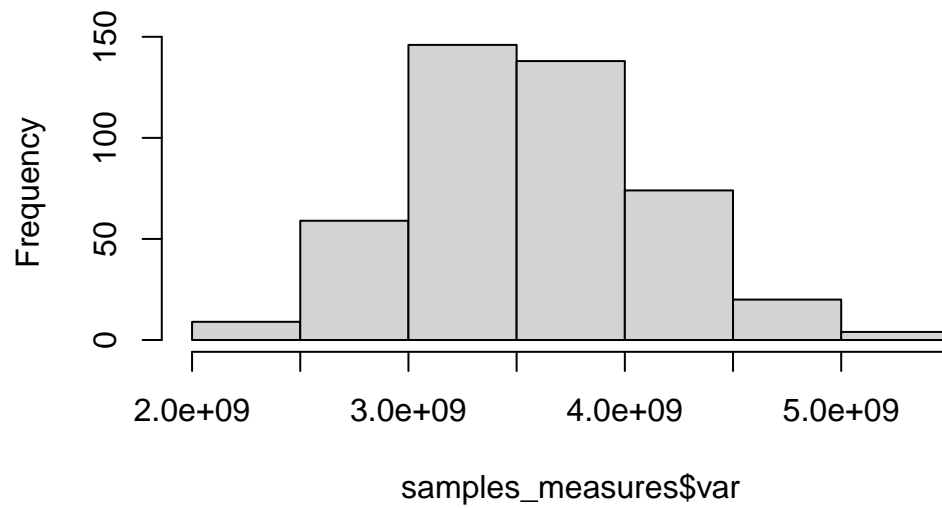


```

hist(samples_measures$var)

```

**Histogram of samples\_measures\$var**

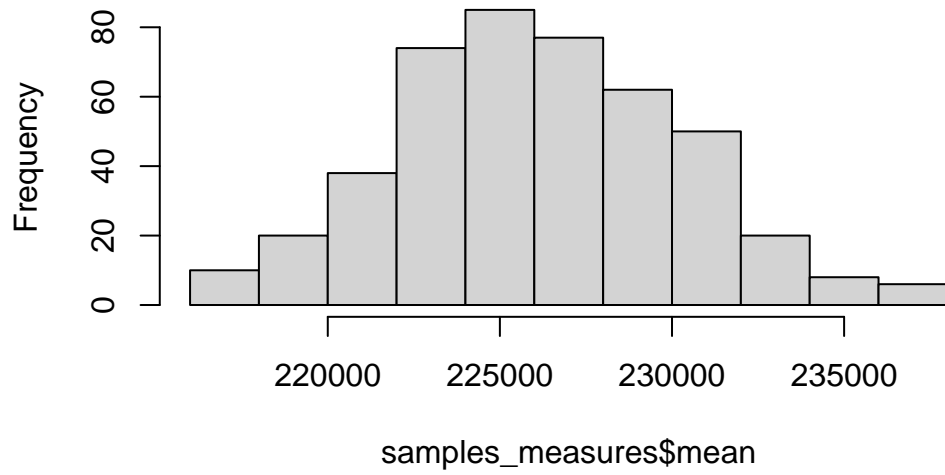


```
display_info <- function() {  
  display_means_info()  
  display_vars_info()  
  hist(samples_measures$mean)  
  hist(samples_measures$var)  
}
```

```
display_info()
```

```
[1] "Mean of Sample Means: 226132.6595"  
[1] "Population Mean: 225779.993932372"  
[1] "Difference: 352.67"  
[1] "Error: 0.2 %"  
[1] "Mean of Sample Vars: 3577994067.7784"  
[1] "Population Var: 3580056952.98649"  
[1] "Difference: -2062885.21"  
[1] "Error: -0.1 %"
```

**Histogram of samples\_measures\$mean**



**Histogram of samples\_measures\$var**

