

bootstrap

March 14, 2025

```
[1]: # Load necessary library
library(MASS)

# Read data from file
data <- read.table("pep1_ws17-pep2_ss18-nozeros.txt", header = TRUE)

# Calculate the original correlation coefficient and its confidence interval
original_cor <- cor.test(data[,1], data[,2])
original_ci <- c(original_cor$conf.int[1], original_cor$conf.int[2])

# Total number of data points (students)
n <- nrow(data)

# Define the number of bootstrap samples to use
# (using the suggestion in the lecture notes)
# (but you can try different numbers to see the effects)
n_bootstrap <- ceiling(n * log(n)^2)

# Initialize vector to store bootstrap correlations
bootstrap_correlations <- numeric(n_bootstrap)

[2]: # Here's how to check the runtime of your code if you want to:
start_time <- Sys.time()

# Bootstrap procedure
for (i in 1:n_bootstrap) {
  # Resample data with replacement
  resampled_data <- data[sample(nrow(data), replace = TRUE), ]
  # Calculate correlation for resampled data
  bootstrap_correlations[i] <- cor(resampled_data[,1], resampled_data[,2])
}

end_time <- Sys.time()
# Calculate the difference
execution_time <- end_time - start_time

# Print the execution time
```

```
print(paste("Bootstrap procedure execution time:", round(execution_time,2),  
↪ "seconds"))
```

```
[1] "Bootstrap procedure execution time: 3.9 seconds"
```

```
[3]: start_time <- Sys.time()

# An alternative, faster, Bootstrap procedure (avoids for loop):
# In the case of sampling with replacement each draw is independent. Hence,
# one can do the drawing in one go, and not having to do it in blocks of size n.
ij <- matrix(sample.int(n,n*n_bootstrap,replace=TRUE), nrow=n,  
↪ ncol=n_bootstrap)
bootstrap_correlations <- apply(ij, 2, function(x) cor(data[x,1],data[x,2]))
ci_bootstrap_correlations <- quantile(bootstrap_correlations, c(0.025, 1-0.025))

end_time <- Sys.time()

# Calculate the difference
execution_time <- end_time - start_time

# Print the execution time
print(paste("Alternative Bootstrap procedure execution time:",  
↪ round(execution_time,2), "seconds"))
```

```
[1] "Alternative Bootstrap procedure execution time: 2.04 seconds"
```

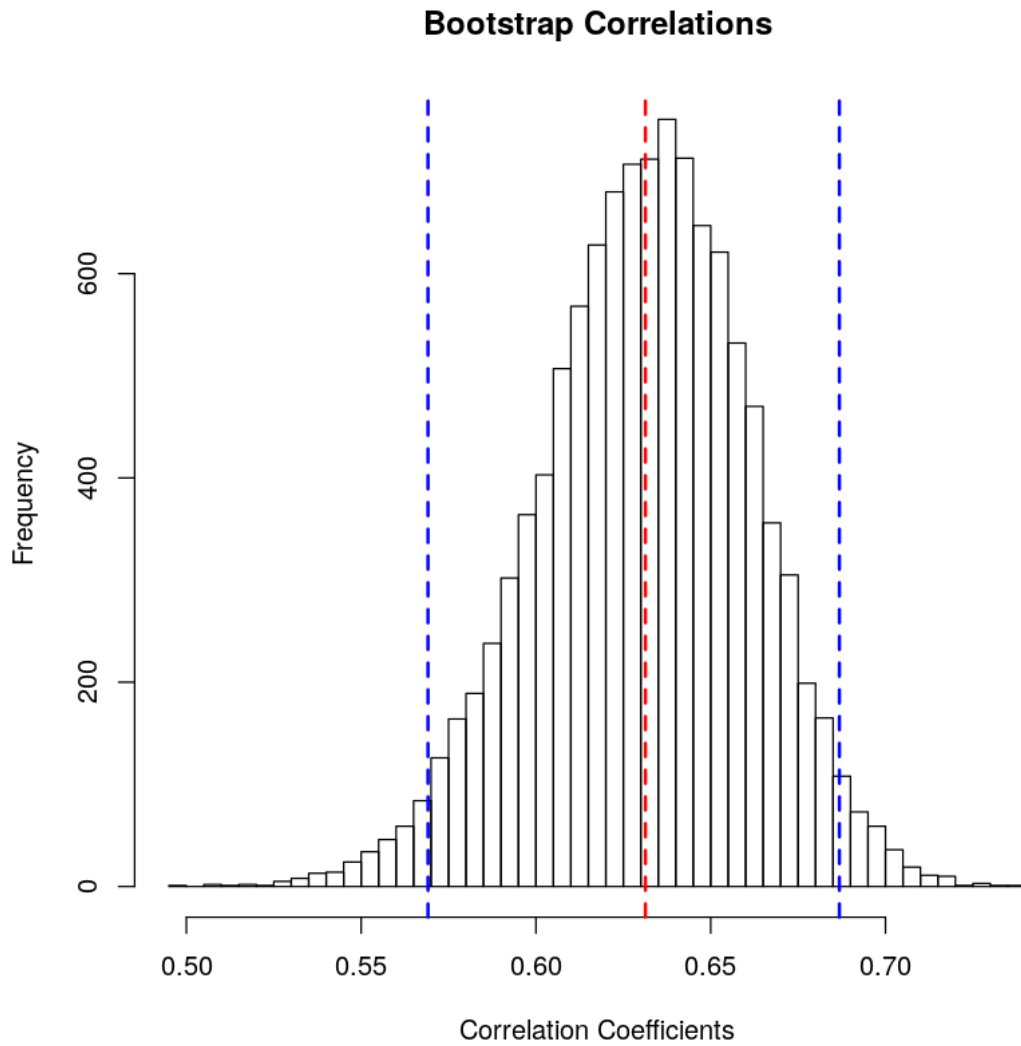
```
[4]: # Calculate bootstrap confidence intervals
ci_lower <- quantile(bootstrap_correlations, 0.025)
ci_upper <- quantile(bootstrap_correlations, 0.975)
```

```
[5]: # Print results
cat("Original correlation coefficient      : ", round(original_cor$estimate,  
↪ 3), "\n")
cat("95% confidence interval original sample : ", round(original_ci, 3), "\n")
cat("Correlation coefficient bootstrap sample: ",  
↪ round(mean(bootstrap_correlations), 3), "\n")
cat("95% confidence interval bootstrap sample: ", round(ci_lower, 3),  
↪ round(ci_upper, 3), "\n")
```

```
Original correlation coefficient      : 0.631
95% confidence interval original sample : 0.561 0.692
Correlation coefficient bootstrap sample: 0.631
95% confidence interval bootstrap sample: 0.569 0.687
```

```
[6]: # Plot histogram of bootstrap correlations
hist(bootstrap_correlations, breaks = 50, main = "Bootstrap Correlations", xlab=  
↪ "Correlation Coefficients")
```

```
abline(v = original_cor$estimate, col = "red", lwd = 2, lty = 2)
abline(v = ci_lower, col = "blue", lwd = 2, lty = 2)
abline(v = ci_upper, col = "blue", lwd = 2, lty = 2)
```



[]: