ForexamSample Spearman Rank-Order Correlation (Small\_Data\_Samples \_with \_Ties)

Kyaw Min Khaing

2025-01-29

## Introduction

This analysis applies the **Spearman Rank-Order Correlation** to examine the relationship between **Number of Visits** and **Mean Heart Rate** for a small dataset with ties.

## Step 1: State the Hypotheses

* **Null Hypothesis (H₀):** There is no significant correlation between the number of visits and the mean heart rate.
* **Alternative Hypothesis (H₁):** There is a significant correlation between the number of visits and the mean heart rate.

## Step 2: Set the Level of Significance

We set **α = 0.05** as the significance level.

data <- tibble(  
 Participant = 1:13,  
 Visits = c(5, 12, 7, 14, 3, 8, 15, 12, 2, 16, 12, 7, 17),  
 HeartRate = c(96, 63, 78, 66, 79, 95, 67, 64, 99, 62, 65, 76, 61)  
)  
print(data)

## # A tibble: 13 × 3  
## Participant Visits HeartRate  
## <int> <dbl> <dbl>  
## 1 1 5 96  
## 2 2 12 63  
## 3 3 7 78  
## 4 4 14 66  
## 5 5 3 79  
## 6 6 8 95  
## 7 7 15 67  
## 8 8 12 64  
## 9 9 2 99  
## 10 10 16 62  
## 11 11 12 65  
## 12 12 7 76  
## 13 13 17 61

## Rank the data with ties correction

You can also embed plots, for example:

## # A tibble: 13 × 9  
## Participant Visits HeartRate Rank\_Visits Rank\_HeartRate D D\_squared  
## <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 5 96 3 12 -9 81   
## 2 2 12 63 8 3 5 25   
## 3 3 7 78 4.5 9 -4.5 20.2  
## 4 4 14 66 10 6 4 16   
## 5 5 3 79 2 10 -8 64   
## 6 6 8 95 6 11 -5 25   
## 7 7 15 67 11 7 4 16   
## 8 8 12 64 8 4 4 16   
## 9 9 2 99 1 13 -12 144   
## 10 10 16 62 12 2 10 100   
## 11 11 12 65 8 5 3 9   
## 12 12 7 76 4.5 8 -3.5 12.2  
## 13 13 17 61 13 1 12 144   
## # ℹ 2 more variables: Ties\_Visits <lgl>, Ties\_HeartRate <lgl>

## Step 4: Compute Spearman’s Rank Correlation Coefficient with Ties Correction

The formula for Spearman’s Rank Correlation Coefficient with ties correction is:

where: - is the difference between ranks, - is the number of observations, - and are tie correction factors for Visits and Heart Rate.

n <- nrow(ranked\_data)  
D\_squared\_sum <- sum(ranked\_data$D\_squared)  
  
numerator <- (n^3 - n) - 6 \* D\_squared\_sum - (T\_visits + T\_heart\_rate) / 2  
denominator <- sqrt((n^3 - n)^2 - (T\_visits + T\_heart\_rate) \* (n^3 - n) + T\_visits \* T\_heart\_rate)  
  
r\_s <- numerator / denominator  
  
cat("Spearman Rank Correlation Coefficient (ρ):", round(r\_s, 4), "\n")

## Spearman Rank Correlation Coefficient (ρ): -0.8486

## Step 5: Determine the Critical Value

For **n = 13** and **α = 0.05 (two-tailed test)**, the critical value from the **Spearman correlation table** is **±0.552**.

## Step 6: Compare the Obtained Value with the Critical Value

if (abs(r\_s) >= 0.552) {  
 cat("The correlation is statistically significant. We reject H₀ and conclude there is a significant correlation.\n")  
} else {  
 cat("The correlation is not statistically significant. We fail to reject H₀.\n")  
}

## The correlation is statistically significant. We reject H₀ and conclude there is a significant correlation.

## Step 7: Interpret the Results

If **|ρ| ≥ 0.552**, the result is significant.

## Step 8: Reporting the Results

*A Spearman rank-order correlation was conducted to determine the relationship between the number of visits and mean heart rate. The analysis yielded ρ = -0.8486. Since the obtained ρ is greater than the critical value of 0.552, we reject the null hypothesis. This suggests that there is significant correlation between the two variables.*

## Step 9: Data Visualization

To better understand the ranked data distribution, we use pivot\_longer to reshape the data and visualize it using ggplot2.

# Convert data to long format for visualization  
data\_long <- ranked\_data %>%  
 pivot\_longer(cols = c(Rank\_Visits, Rank\_HeartRate), names\_to = "Variable", values\_to = "Rank")  
  
# Create visualization  
rank\_plot <- ggplot(data\_long, aes(x = Variable, y = Rank, fill = Variable)) +  
 geom\_boxplot(outlier.color = "red", outlier.shape = 8) +  
 geom\_jitter(width = 0.2, size = 2, alpha = 0.7) +  
 labs(  
 title = "Rank Distribution of Visits and Heart Rate",  
 x = "Variable",  
 y = "Rank"  
 ) +  
 theme\_minimal()  
  
# Print the plot  
print(rank\_plot)

