Sample Runs Test for Small Data Samples

Kyaw Min Khaing

2025-02-03

Step 1: State the Null and Research Hypotheses Null Hypothesis : The sequence of categories (M/F) is random. Alternative Hypothesis : The sequence of categories (M/F) is not random (suggesting clustering or patterns). Step 2: Set the Level of Significance The significance level (α) is set at 0.05 (5%).

# Level of significance  
alpha <- 0.05

Step 3: Create the Dataset We define the sequence of Male (M) and Female (F) groupings and represent them as a vector.

## # A tibble: 13 × 2  
## Position Category  
## <int> <chr>   
## 1 1 M   
## 2 2 FF   
## 3 3 M   
## 4 4 F   
## 5 5 M   
## 6 6 F   
## 7 7 M   
## 8 8 FF   
## 9 9 M   
## 10 10 FFF   
## 11 11 MM   
## 12 12 F   
## 13 13 MMM

Step 4: Count the Total Runs in the Sequence A run is defined as a sequence of the same category (M or F) occurring consecutively.

# Convert sequence into binary format for easy computation  
binary\_seq <- data$Category %>% str\_extract("^M|^F")  
  
# Count number of runs  
num\_runs <- sum(binary\_seq[-1] != binary\_seq[-length(binary\_seq)]) + 1  
  
# Display number of runs  
num\_runs

## [1] 13

Compute the Expected Number of Runs and Standard Deviation

The expected number of runs () and the standard deviation () are given by:

where: - = number of **M** (Male occurrences) - = number of **F** (Female occurrences) -

# Count Males (M) and Females (F)  
n1 <- sum(str\_detect(data$Category, "M"))  
n2 <- sum(str\_detect(data$Category, "F"))  
n <- n1 + n2  
  
# Compute expected runs and standard deviation  
E\_R <- (2 \* n1 \* n2) / n + 1  
sigma\_R <- sqrt((2 \* n1 \* n2 \* (2 \* n1 \* n2 - n)) / (n^2 \* (n - 1)))  
  
# Display expected runs and standard deviation  
E\_R

## [1] 7.461538

sigma\_R

## [1] 1.714884

Step 6: Compute the Z-score

# Compute Z-score  
Z\_score <- (num\_runs - E\_R) / sigma\_R  
  
# Display Z-score  
Z\_score

## [1] 3.229641

Step 7: Determine the Critical Value Using a standard normal distribution table, the critical value for α = 0.05 (two-tailed test)

# Compute critical value  
critical\_value <- qnorm(1 - alpha/2)  
# Display critical value  
critical\_value

## [1] 1.959964

Compare the Obtained Value with the Critical Value If |Z-score| > critical value, we reject the null hypothesis.

# Decision based on comparison  
if (abs(Z\_score) > critical\_value) {  
 result <- "Reject the null hypothesis (the sequence is not random)."  
} else {  
 result <- "Fail to reject the null hypothesis (the sequence is random)."  
}  
  
# Display decision  
result

## [1] "Reject the null hypothesis (the sequence is not random)."

Conclusion

cat("Number of Runs (R):", num\_runs, "\n")

## Number of Runs (R): 13

cat("Expected Runs (E(R)):", E\_R, "\n")

## Expected Runs (E(R)): 7.461538

cat("Standard Deviation (σR):", sigma\_R, "\n")

## Standard Deviation (σR): 1.714884

cat("Computed Z-score:", Z\_score, "\n")

## Computed Z-score: 3.229641

cat("Critical Value (two-tailed):", critical\_value, "\n")

## Critical Value (two-tailed): 1.959964

cat("Conclusion:", result, "\n")

## Conclusion: Reject the null hypothesis (the sequence is not random).