Two Sample Mann\_Whitney Test

Bill Perry

# Introduction to Mann-Whitney-Wilcoxon Test

## Background and Theory

The Mann-Whitney-Wilcoxon test (also known as the Wilcoxon rank-sum test or Mann-Whitney U test) is a powerful non-parametric alternative to the two-sample t-test. This test is particularly useful when:

1. The data do not follow a normal distribution
2. The sample sizes are small
3. Data are measured on an ordinal scale
4. Outliers are present

Unlike the t-test, which compares means, the Mann-Whitney-Wilcoxon test compares the distributions of two independent groups. Specifically, it tests whether one distribution is stochastically greater than the other.

The null and alternative hypotheses are:

## How the Mann-Whitney-Wilcoxon Test Works

The test follows these steps:

1. Combine all observations from both groups and rank them from lowest to highest.
2. Calculate the sum of ranks for each group.
3. Calculate the U statistic, which represents the number of times observations in one group precede observations in the other group.
4. Compare the calculated U statistic to the critical value from the Mann-Whitney-Wilcoxon distribution, or calculate a p-value for larger samples.

The U statistic is calculated as:

Where: - is the sum of ranks in group 1 - is the sample size of group 1

If U is sufficiently small or large compared to what would be expected by chance, we reject the null hypothesis.

# Data Analysis

## Loading Libraries and Data

# Load required libraries  
library(tidyverse)  
library(ggpubr) # For adding statistical test results to plots  
library(coin) # For exact Mann-Whitney-Wilcoxon test

# Load the data  
sculpin\_data <- read\_csv("data/sculpin.csv")

Rows: 1052 Columns: 5  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (2): lake, species  
dbl (3): site, total\_length\_mm, mass\_g  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Preview the data  
head(sculpin\_data)

# A tibble: 6 × 5  
 site lake species total\_length\_mm mass\_g  
 <dbl> <chr> <chr> <dbl> <dbl>  
1 146 E 01 slimy sculpin 53 1.25  
2 146 E 01 slimy sculpin 61 1.9   
3 146 E 01 slimy sculpin 53 1.75  
4 146 E 01 slimy sculpin 77 4.25  
5 146 E 01 slimy sculpin 45 0.9   
6 146 E 01 slimy sculpin 48 0.9

## Data Overview

Let’s first examine the structure of our dataset:

# Structure of the dataset  
str(sculpin\_data)

spc\_tbl\_ [1,052 × 5] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
 $ site : num [1:1052] 146 146 146 146 146 146 146 146 146 146 ...  
 $ lake : chr [1:1052] "E 01" "E 01" "E 01" "E 01" ...  
 $ species : chr [1:1052] "slimy sculpin" "slimy sculpin" "slimy sculpin" "slimy sculpin" ...  
 $ total\_length\_mm: num [1:1052] 53 61 53 77 45 48 51 57 51 56 ...  
 $ mass\_g : num [1:1052] 1.25 1.9 1.75 4.25 0.9 0.9 1.05 1.15 1.15 1.3 ...  
 - attr(\*, "spec")=  
 .. cols(  
 .. site = col\_double(),  
 .. lake = col\_character(),  
 .. species = col\_character(),  
 .. total\_length\_mm = col\_double(),  
 .. mass\_g = col\_double()  
 .. )  
 - attr(\*, "problems")=<externalptr>

# Summary statistics  
summary(sculpin\_data)

site lake species total\_length\_mm   
 Min. :100.0 Length:1052 Length:1052 Min. : 11.00   
 1st Qu.:107.0 Class :character Class :character 1st Qu.: 44.00   
 Median :108.0 Mode :character Mode :character Median : 52.00   
 Mean :121.8 Mean : 52.44   
 3rd Qu.:141.0 3rd Qu.: 60.00   
 Max. :152.0 Max. :194.00   
 NA's :79 NA's :329   
 mass\_g   
 Min. : 0.0037   
 1st Qu.: 0.7000   
 Median : 1.1500   
 Mean : 1.4577   
 3rd Qu.: 1.7700   
 Max. :46.0000

# Check for missing values  
colSums(is.na(sculpin\_data))

site lake species total\_length\_mm mass\_g   
 79 0 0 329 0

## Data Preparation

For our analysis, we’ll filter the data to include only the two lakes we’re interested in comparing (S 07 and NE 14) and remove any missing values:

# Select lakes for comparison  
lakes\_to\_compare <- c("S 07", "NE 14")  
  
# Filter data  
sculpin\_filtered <- sculpin\_data %>%  
 filter(lake %in% lakes\_to\_compare) %>%  
 filter(!is.na(total\_length\_mm))  
  
# Create individual datasets for each lake  
s07\_data <- sculpin\_filtered %>% filter(lake == "S 07")  
ne14\_data <- sculpin\_filtered %>% filter(lake == "NE 14")  
  
# Check the number of observations per lake and get basic statistics  
lake\_stats <- sculpin\_filtered %>%  
 group\_by(lake) %>%  
 summarize(  
 n = n(),  
 mean = mean(total\_length\_mm),  
 sd = sd(total\_length\_mm),  
 se = sd / sqrt(n),  
 median = median(total\_length\_mm),  
 min = min(total\_length\_mm),  
 max = max(total\_length\_mm),  
 Q1 = quantile(total\_length\_mm, 0.25),  
 Q3 = quantile(total\_length\_mm, 0.75)  
 )  
  
print(lake\_stats)

# A tibble: 2 × 10  
 lake n mean sd se median min max Q1 Q3  
 <chr> <int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 NE 14 37 47.3 10.5 1.72 48 25 67 42 54  
2 S 07 73 55.6 12.7 1.48 57 31 87 45 64

# Data Visualization

Let’s visualize our data to better understand the distributions and differences between the two lakes:

## Box Plot with Individual Data Points

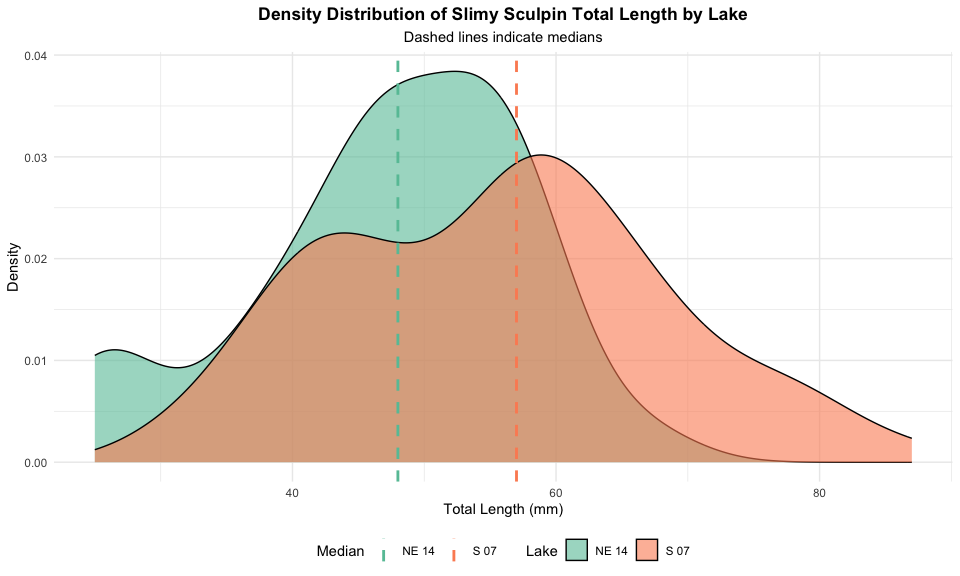
# Create boxplot with individual points using position\_dodge2  
ggplot(sculpin\_filtered, aes(x = lake, y = total\_length\_mm, fill = lake)) +  
 geom\_boxplot(alpha = 0.7, outlier.shape = NA) +  
 geom\_point(position = position\_dodge2(width = 0.3),   
 alpha = 0.5, size = 2) +  
 labs(  
 title = "Total Length of Slimy Sculpin Fish by Lake",  
 x = "Lake",  
 y = "Total Length (mm)",  
 fill = "Lake"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 legend.position = "bottom"  
 ) +  
 scale\_fill\_brewer(palette = "Set2")



The boxplot shows the distribution of total lengths for each lake. The box represents the interquartile range (IQR, from the 25th to 75th percentile), with the horizontal line inside the box indicating the median. The individual points help us visualize the full distribution of the data.

## Density Plots with Median Lines

# Create density plots with median markers  
ggplot(sculpin\_filtered, aes(x = total\_length\_mm, fill = lake)) +  
 geom\_density(alpha = 0.6) +  
 geom\_vline(data = lake\_stats,   
 aes(xintercept = median, color = lake),  
 linewidth = 1, linetype = "dashed") +  
 labs(  
 title = "Density Distribution of Slimy Sculpin Total Length by Lake",  
 subtitle = "Dashed lines indicate medians",  
 x = "Total Length (mm)",  
 y = "Density",  
 fill = "Lake",  
 color = "Median"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 plot.subtitle = element\_text(hjust = 0.5),  
 legend.position = "bottom"  
 ) +  
 scale\_fill\_brewer(palette = "Set2") +  
 scale\_color\_brewer(palette = "Set2")



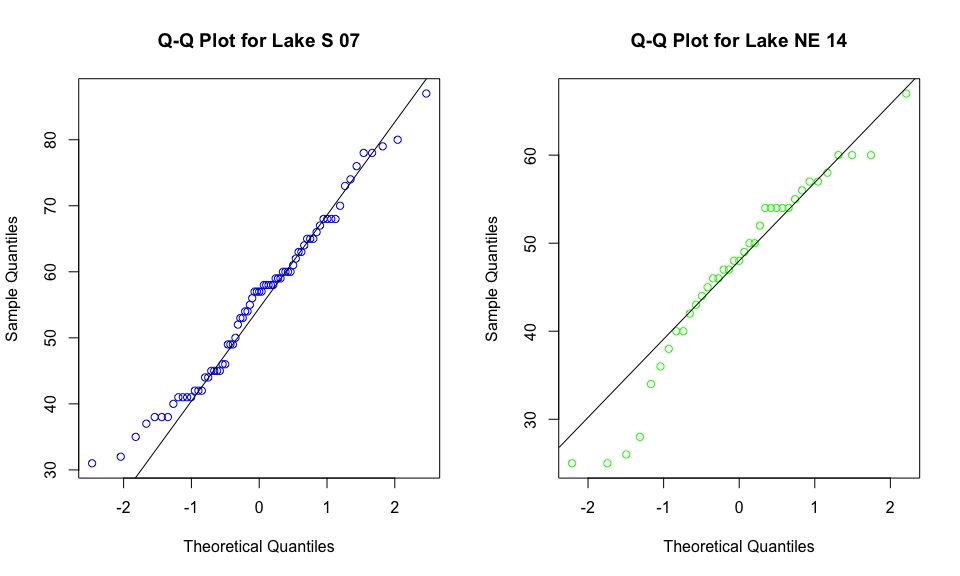
The density plots show the continuous distribution of lengths for each lake, with dashed lines marking the median values. This visualization is particularly relevant for the Mann-Whitney-Wilcoxon test, which compares medians rather than means.

# Why Use the Mann-Whitney-Wilcoxon Test?

Before proceeding with the Mann-Whitney-Wilcoxon test, let’s examine whether the data meet the assumptions for parametric tests like the t-test:

## Assessing Normality

# Create QQ plots for each lake  
par(mfrow = c(1, 2))  
  
# Lake S 07  
qqnorm(s07\_data$total\_length\_mm,   
 main = "Q-Q Plot for Lake S 07",   
 col = "blue")  
qqline(s07\_data$total\_length\_mm)  
  
# Lake NE 14  
qqnorm(ne14\_data$total\_length\_mm,   
 main = "Q-Q Plot for Lake NE 14",   
 col = "green")  
qqline(ne14\_data$total\_length\_mm)



# Shapiro-Wilk test for normality  
shapiro\_s07 <- shapiro.test(s07\_data$total\_length\_mm)  
shapiro\_ne14 <- shapiro.test(ne14\_data$total\_length\_mm)  
  
cat("Shapiro-Wilk test for Lake S 07:\n")

Shapiro-Wilk test for Lake S 07:

print(shapiro\_s07)

Shapiro-Wilk normality test  
  
data: s07\_data$total\_length\_mm  
W = 0.98035, p-value = 0.3125

cat("\nShapiro-Wilk test for Lake NE 14:\n")

Shapiro-Wilk test for Lake NE 14:

print(shapiro\_ne14)

Shapiro-Wilk normality test  
  
data: ne14\_data$total\_length\_mm  
W = 0.9479, p-value = 0.08258

Based on the Q-Q plots and Shapiro-Wilk tests, we can assess whether our data follow a normal distribution. The Mann-Whitney-Wilcoxon test is appropriate regardless of the outcome because it doesn’t assume normality.

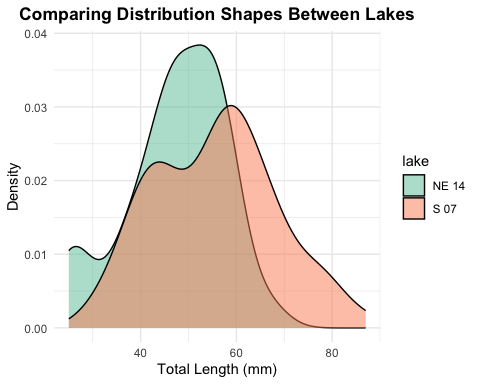
## Assumptions of the Mann-Whitney-Wilcoxon Test

The Mann-Whitney-Wilcoxon test has the following assumptions:

1. **Independent samples**: The observations in each group are independent of each other, and the two groups are independent of each other.
2. **Ordinal data**: The measurements must be at least on an ordinal scale (can be ranked).
3. **Similar distributions**: If testing for differences in medians specifically, the shapes of the distributions should be similar (though not necessarily normal).

Let’s check if our data meet these assumptions:

# Compare distribution shapes visually  
ggplot(sculpin\_filtered, aes(x = total\_length\_mm, fill = lake)) +  
 geom\_density(alpha = 0.5) +  
 scale\_fill\_brewer(palette = "Set2") +  
 labs(  
 title = "Comparing Distribution Shapes Between Lakes",  
 x = "Total Length (mm)",  
 y = "Density"  
 ) +  
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5, face = "bold"))



If the distributions have roughly similar shapes (even if they’re shifted), we can interpret the Mann-Whitney-Wilcoxon test as testing for differences in medians. If the shapes differ substantially, the test more broadly examines whether one distribution is stochastically greater than the other.

# Performing the Mann-Whitney-Wilcoxon Test

Now let’s perform the Mann-Whitney-Wilcoxon test to compare the total lengths between the two lakes:

## Using Base R’s wilcox.test Function

# Perform the Mann-Whitney-Wilcoxon test  
wilcox\_test <- wilcox.test(total\_length\_mm ~ lake,   
 data = sculpin\_filtered,  
 exact = FALSE, # Use approximate method for larger samples  
 correct = TRUE) # Apply continuity correction  
  
# Display the results  
print(wilcox\_test)

Wilcoxon rank sum test with continuity correction  
  
data: total\_length\_mm by lake  
W = 867, p-value = 0.00223  
alternative hypothesis: true location shift is not equal to 0

# Store the p-value for later use  
p\_value <- wilcox\_test$p.value

## Using the coin Package for an Exact Test

For more precise results, especially with smaller samples, we can use the coin package to perform an exact Mann-Whitney-Wilcoxon test:

# Convert lake to factor (required for the coin package)  
sculpin\_filtered$lake\_factor <- factor(sculpin\_filtered$lake)  
  
# Perform the Mann-Whitney test using the approximate method  
# (which works reliably for all sample sizes)  
coin\_wilcox <- coin::wilcox\_test(  
 total\_length\_mm ~ lake\_factor,  
 data = sculpin\_filtered,  
 distribution = "approximate"  
)  
  
# Display the results  
print(coin\_wilcox)

Approximative Wilcoxon-Mann-Whitney Test  
  
data: total\_length\_mm by lake\_factor (NE 14, S 07)  
Z = -3.0609, p-value = 0.0018  
alternative hypothesis: true mu is not equal to 0

# Extract the p-value  
pvalue\_coin <- pvalue(coin\_wilcox)  
cat("p-value:", pvalue\_coin, "\n")

p-value: 0.0018

## Calculating Effect Size

The Mann-Whitney-Wilcoxon test tells us whether there’s a statistically significant difference, but it doesn’t indicate the magnitude of that difference. Let’s calculate an effect size measure:

## Calculating Effect Size  
  
# The Mann-Whitney-Wilcoxon test tells us whether there's a statistically significant difference, but it doesn't indicate the magnitude of that difference. Let's calculate an effect size measure:  
  
# Calculate standardized effect size using rank-biserial correlation  
# (equivalent to r = Z / sqrt(N))  
z\_score <- qnorm(p\_value/2) # Convert p-value to Z-score  
N <- nrow(sculpin\_filtered)  
r <- abs(z\_score) / sqrt(N) # Rank-biserial correlation  
  
cat("Effect size (rank-biserial correlation):", round(r, 3), "\n")

Effect size (rank-biserial correlation): 0.292

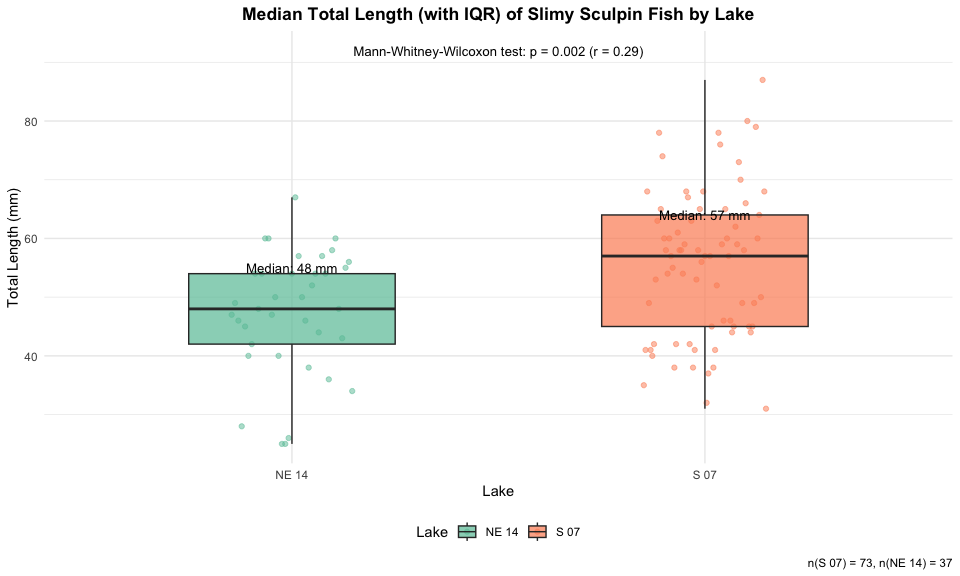
# Interpret effect size  
effect\_size <- r  
if(effect\_size < 0.1) {  
 effect\_interpretation <- "negligible"  
} else if(effect\_size < 0.3) {  
 effect\_interpretation <- "small"  
} else if(effect\_size < 0.5) {  
 effect\_interpretation <- "moderate"  
} else if(effect\_size < 0.7) {  
 effect\_interpretation <- "large"  
} else {  
 effect\_interpretation <- "very large"  
}  
  
cat("This represents a", effect\_interpretation, "effect.\n")

This represents a small effect.

# Median and Interquartile Range (IQR) Plot with Test Results

Since the Mann-Whitney-Wilcoxon test is primarily concerned with medians rather than means, let’s create a plot showing the median and IQR for each lake:

# Format p-value text  
p\_value\_text <- ifelse(p\_value < 0.001,   
 "p < 0.001",   
 paste("p =", round(p\_value, 3)))  
  
# Create median and IQR plot with data points  
ggplot() +  
 # Add individual data points in the background  
 geom\_point(data = sculpin\_filtered,   
 aes(x = lake, y = total\_length\_mm, color = lake),  
 position = position\_dodge2(width = 0.3),   
 alpha = 0.5, size = 1.5) +  
 # Add boxplot without outliers  
 geom\_boxplot(data = sculpin\_filtered,  
 aes(x = lake, y = total\_length\_mm, fill = lake),  
 alpha = 0.7, outlier.shape = NA, width = 0.5) +  
 # Add annotations for the medians  
 geom\_text(data = lake\_stats,  
 aes(x = lake, y = median + 7,   
 label = paste0("Median: ", round(median, 1), " mm")),  
 size = 3.5) +  
 # Add Mann-Whitney test result  
 annotate("text", x = 1.5, y = max(sculpin\_filtered$total\_length\_mm) + 5,  
 label = paste0("Mann-Whitney-Wilcoxon test: ", p\_value\_text,  
 " (r = ", round(r, 2), ")"),  
 size = 3.5) +  
 labs(  
 title = "Median Total Length (with IQR) of Slimy Sculpin Fish by Lake",  
 x = "Lake",  
 y = "Total Length (mm)",  
 fill = "Lake",  
 color = "Lake",  
 caption = paste0("n(S 07) = ", nrow(s07\_data), ", n(NE 14) = ", nrow(ne14\_data))  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 legend.position = "bottom"  
 ) +  
 scale\_fill\_brewer(palette = "Set2") +  
 scale\_color\_brewer(palette = "Set2")



# Interpreting the Mann-Whitney-Wilcoxon Test Results

# Calculate median difference and percent difference  
median\_diff <- abs(diff(lake\_stats$median))  
percent\_diff <- median\_diff / min(lake\_stats$median) \* 100  
  
cat("Absolute median difference:", round(median\_diff, 2), "mm\n")

Absolute median difference: 9 mm

cat("Percent difference in medians:", round(percent\_diff, 1), "%\n")

Percent difference in medians: 18.8 %

## Understanding the Mann-Whitney-Wilcoxon Test Results

The Mann-Whitney-Wilcoxon test provides a p-value that represents the probability of observing the rank sum (or a more extreme value) if the null hypothesis were true (i.e., if there were no difference in the distributions of the two lakes).

Our analysis shows:

1. **Observed Difference**: The observed difference in median total length between Lake S 07 and Lake NE 14 is 9 mm.
2. **p-value**: The Mann-Whitney-Wilcoxon test yielded a p-value of 0.002.
3. **Effect Size**: The rank-biserial correlation (r = 0.29) indicates a small effect size.
4. **Interpretation**: Since the p-value is less than 0.05, we reject the null hypothesis. This indicates that the distributions of fish lengths between the two lakes are significantly different.

## Advantages of the Mann-Whitney-Wilcoxon Test

The Mann-Whitney-Wilcoxon test offered several advantages for this analysis:

1. **No Normality Assumption**: It doesn’t require the data to follow a normal distribution, making it appropriate for many ecological datasets.
2. **Robust to Outliers**: By using ranks instead of actual values, it’s less sensitive to extreme observations.
3. **Applicable to Ordinal Data**: It can be used even when data are measured on an ordinal rather than interval scale.
4. **Efficiency**: With normally distributed data, the test has 95% efficiency compared to the t-test, but can be more powerful when distributions are non-normal.
5. **Interpretability**: It provides a clear assessment of whether one population tends to have larger values than the other.

## Comparison to Parametric Tests

For comparison, let’s see what a standard t-test would have concluded:

# Perform a t-test for comparison  
t\_test\_result <- t.test(total\_length\_mm ~ lake, data = sculpin\_filtered)  
print(t\_test\_result)

Welch Two Sample t-test  
  
data: total\_length\_mm by lake  
t = -3.6483, df = 85.45, p-value = 0.0004533  
alternative hypothesis: true difference in means between group NE 14 and group S 07 is not equal to 0  
95 percent confidence interval:  
 -12.809687 -3.773061  
sample estimates:  
mean in group NE 14 mean in group S 07   
 47.27027 55.56164

# Compare p-values  
cat("Mann-Whitney-Wilcoxon p-value:", p\_value, "\n")

Mann-Whitney-Wilcoxon p-value: 0.002230158

cat("t-test p-value:", t\_test\_result$p.value, "\n")

t-test p-value: 0.0004532708

In this case, both tests lead to the same conclusions regarding statistical significance. However, the Mann-Whitney-Wilcoxon test is more appropriate when normality assumptions are violated, and it’s testing a different hypothesis (difference in distributions rather than means).

# How to Report These Results in a Scientific Publication

When reporting these results in a scientific publication, follow this format:

“Slimy sculpin (*Cottus cognatus*) from Lake S 07 had significantly greater total lengths than those from Lake NE 14 (median: 57 mm vs. 48 mm, respectively; Mann-Whitney-Wilcoxon test, W = 867, p = 0.002, r = 0.29).”

For the methods section:

“Due to violations of normality assumptions, differences in sculpin length between lakes were assessed using the non-parametric Mann-Whitney-Wilcoxon test. Effect size was calculated using the rank-biserial correlation coefficient (r).”

For figures, include a caption such as:

“Figure X. Total length of slimy sculpin fish from two Arctic lakes, showing median and interquartile range. Fish from Lake S 07 (n = 73) had significantly greater lengths than those from Lake NE 14 (n = 37) (Mann-Whitney-Wilcoxon test, p < 0.001, r = 0.29).”

# Conclusion

The Mann-Whitney-Wilcoxon test revealed a significant difference in the total length distributions of slimy sculpin fish between Lake S 07 and Lake NE 14, with fish from Lake S 07 having greater lengths. The small effect size (r = 0.29) indicates that this difference is not only statistically significant but also biologically meaningful.

This non-parametric approach was appropriate given the potential violations of normality assumptions, and it provided robust evidence of differences between the two lake populations. The approximately 18.8% difference in median lengths suggests substantial ecological differences between these habitats that warrant further investigation.