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)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.2 ✔ tibble 3.3.0  
✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
✔ purrr 1.1.0   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(car) # For Levene's test

Loading required package: carData  
  
Attaching package: 'car'  
  
The following object is masked from 'package:dplyr':  
  
 recode  
  
The following object is masked from 'package:purrr':  
  
 some

# library(ggpubr) # For adding p-values to plots  
library(coin) # For permutation tests

Loading required package: survival

library(skimr)  
library(rcompanion) # For plotNormalHistogram

# Load the data  
sculpin\_df <- read\_csv("data/t\_test\_sculpin\_s07\_ne14.csv")

Rows: 110 Columns: 5  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: ","  
chr (2): lake, species  
dbl (3): site, 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\_df)

# A tibble: 6 × 5  
 site lake species length\_mm mass\_g  
 <dbl> <chr> <chr> <dbl> <dbl>  
1 109 NE 14 slimy sculpin 47 0.7   
2 109 NE 14 slimy sculpin 49 0.9   
3 109 NE 14 slimy sculpin 46 0.7   
4 109 NE 14 slimy sculpin 28 0.15  
5 109 NE 14 slimy sculpin 45 0.65  
6 109 NE 14 slimy sculpin 40 0.3

## Data Overview

Let’s first examine the structure of our dataset:

# Summary statistics  
sculpin\_df %>% group\_by(lake) %>% skim()

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 110 |
| Number of columns | 5 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 1 |
| numeric | 3 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | lake |

**Variable type: character**

| skim\_variable | lake | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| species | NE 14 | 0 | 1 | 13 | 13 | 0 | 1 | 0 |
| species | S 07 | 0 | 1 | 13 | 13 | 0 | 1 | 0 |

**Variable type: numeric**

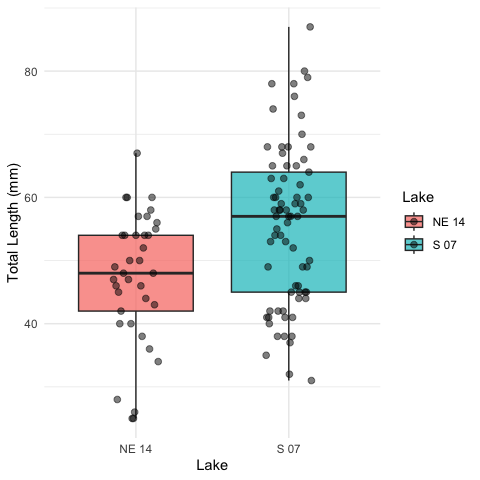
| skim\_variable | lake | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| site | NE 14 | 0 | 1 | 109.00 | 0.00 | 109.00 | 109.00 | 109.00 | 109.00 | 109.00 | ▁▁▇▁▁ |
| site | S 07 | 0 | 1 | 152.00 | 0.00 | 152.00 | 152.00 | 152.00 | 152.00 | 152.00 | ▁▁▇▁▁ |
| length\_mm | NE 14 | 0 | 1 | 47.27 | 10.49 | 25.00 | 42.00 | 48.00 | 54.00 | 67.00 | ▂▃▇▇▂ |
| length\_mm | S 07 | 0 | 1 | 55.56 | 12.65 | 31.00 | 45.00 | 57.00 | 64.00 | 87.00 | ▅▅▇▃▂ |
| mass\_g | NE 14 | 0 | 1 | 0.89 | 0.52 | 0.10 | 0.45 | 0.85 | 1.25 | 2.30 | ▇▇▇▂▁ |
| mass\_g | S 07 | 0 | 1 | 1.66 | 1.23 | 0.25 | 0.80 | 1.45 | 2.10 | 7.37 | ▇▃▁▁▁ |

# Data Visualization

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

## Box Plot with Individual Data Points

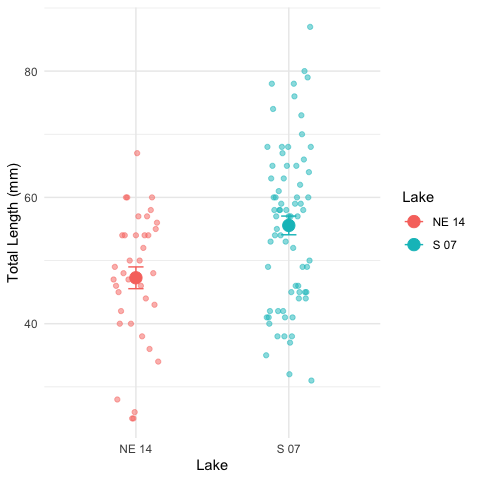
# Create boxplot with individual points  
ggplot(sculpin\_df, aes(x = lake, y = length\_mm, fill = lake)) +  
 geom\_boxplot(alpha = 0.7, outlier.shape = NA) +  
 geom\_point(position = position\_dodge2(width = 0.3),   
 alpha = 0.5, size = 2) +  
 labs(  
 x = "Lake",  
 y = "Total Length (mm)",  
 fill = "Lake"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 legend.position = "right"  
 )



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 show the actual measurements, helping us visualize the full distribution of the data.

## Mean and SE Individual Data Points

sculpin\_df %>%   
ggplot( aes(x = lake, y = length\_mm, color = lake)) +  
 # Add individual data points in the background  
 geom\_point(position = position\_dodge2(width = 0.3),   
 alpha = 0.5, size = 1.5) +  
 # Add mean and standard error  
 stat\_summary(fun = mean, geom = "point", size = 4) +  
 stat\_summary(fun.data = mean\_se, geom = "errorbar", width = 0.1) +  
 labs(  
 x = "Lake",  
 y = "Total Length (mm)",  
 color = "Lake"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 legend.position = "right"  
 )



# Why Use the Mann-Whitney-Wilcoxon Test?

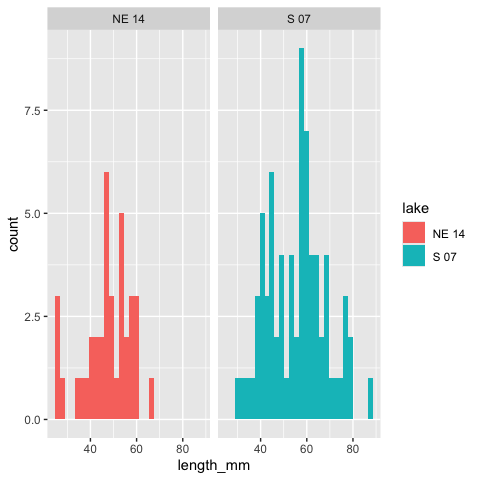
Before proceeding with the Mann-Whitney-Wilcoxon test, let’s examine whether the data meet the assumptions for parametric tests like the t-test:

### 2. Normality Assumption

* We’ll check normality using:
  + Visual methods: Histograms and Q-Q plots
  + Formal test: Shapiro-Wilk test

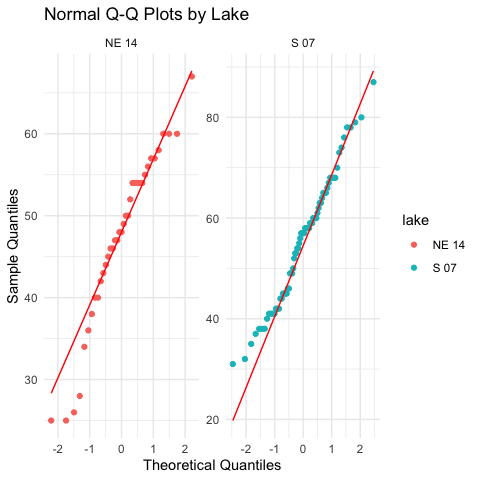
#### Histograms

sculpin\_df %>%   
 ggplot(aes(length\_mm, fill = lake))+  
 geom\_histogram()+  
 facet\_wrap(~lake)



#### QQ Plots

# QQ plot for lakes  
sculpin\_df %>%   
 ggplot( aes(sample = length\_mm, color=lake)) +  
 stat\_qq() +  
 stat\_qq\_line(color = "red") +  
 facet\_wrap(~lake, scales = "free") +  
 labs(title = "Normal Q-Q Plots by Lake",  
 x = "Theoretical Quantiles",  
 y = "Sample Quantiles") +  
 theme\_minimal()



#### Shapiro-Wilk Test

# Simple approach - just split by lake and run the test  
sculpin\_df %>%  
 filter(lake == "S 07") %>%  
 pull(length\_mm) %>%  
 shapiro.test()

Shapiro-Wilk normality test  
  
data: .  
W = 0.98035, p-value = 0.3125

sculpin\_df %>%  
 filter(lake == "NE 14") %>%  
 pull(length\_mm) %>%  
 shapiro.test()

Shapiro-Wilk normality test  
  
data: .  
W = 0.9479, p-value = 0.08258

Another way

sculpin\_df %>%  
 group\_by(lake) %>%  
 group\_walk(~ {  
 cat("Shapiro-Wilk test for Lake", .y$lake, ":\n")  
 test\_result <- shapiro.test(.x$length\_mm)  
 print(test\_result)  
 cat("\n")  
 })

Shapiro-Wilk test for Lake NE 14 :  
  
 Shapiro-Wilk normality test  
  
data: .x$length\_mm  
W = 0.9479, p-value = 0.08258  
  
  
Shapiro-Wilk test for Lake S 07 :  
  
 Shapiro-Wilk normality test  
  
data: .x$length\_mm  
W = 0.98035, p-value = 0.3125

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).

# 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(length\_mm ~ lake,   
 data = sculpin\_df,  
 exact = FALSE, # Use approximate method for larger samples  
 correct = TRUE) # Apply continuity correction  
  
# Display the results  
wilcox\_test

Wilcoxon rank sum test with continuity correction  
  
data: 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\_df$lake\_factor <- factor(sculpin\_df$lake)  
  
# Perform the Mann-Whitney test using the approximate method  
# (which works reliably for all sample sizes)  
coin\_wilcox <- coin::wilcox\_test(  
 length\_mm ~ lake\_factor,  
 data = sculpin\_df,  
 distribution = "approximate"  
)  
  
# Extract the p-value  
pvalue\_coin <- pvalue(coin\_wilcox)

## 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\_df)  
r <- abs(z\_score) / sqrt(N) # Rank-biserial correlation  
  
  
  
# Interpret effect size  
effect\_size <- r  
if(effect\_size < 0.1) {  
 effect\_interpretation <- "negligible effect"  
} else if(effect\_size < 0.3) {  
 effect\_interpretation <- "small effect"  
} else if(effect\_size < 0.5) {  
 effect\_interpretation <- "moderate effect"  
} else if(effect\_size < 0.7) {  
 effect\_interpretation <- "large effect"  
} else {  
 effect\_interpretation <- "very large effect"  
}  
  
cat("Effect size (rank-biserial correlation):")

Effect size (rank-biserial correlation):

round(r, 3)

[1] 0.292

cat("This represents a:")

This represents a:

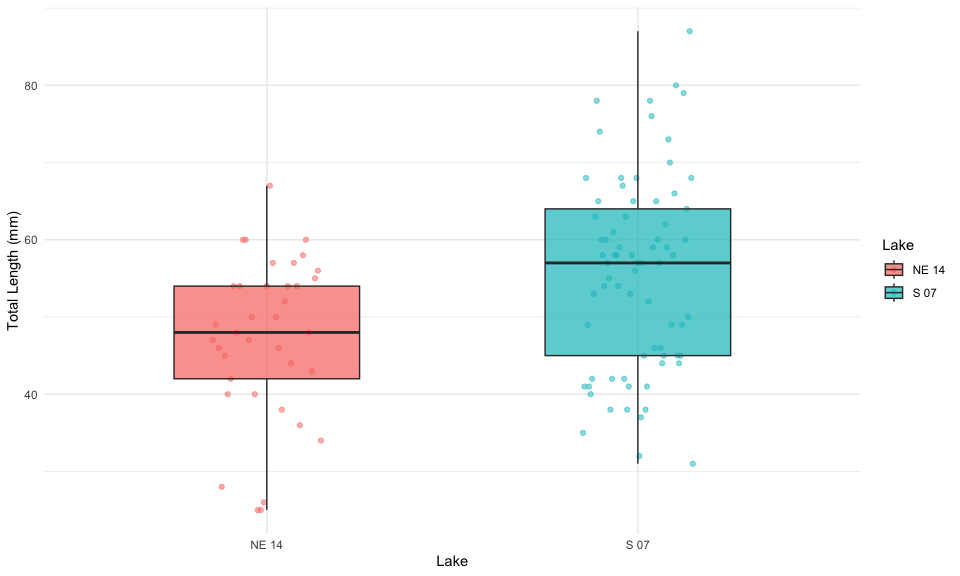
effect\_interpretation

[1] "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:

# Create median and IQR plot with data points  
ggplot() +  
 # Add individual data points in the background  
 geom\_point(data = sculpin\_df,   
 aes(x = lake, y = length\_mm, color = lake),  
 position = position\_dodge2(width = 0.3),   
 alpha = 0.5, size = 1.5) +  
 # Add boxplot without outliers  
 geom\_boxplot(data = sculpin\_df,  
 aes(x = lake, y = length\_mm, fill = lake),  
 alpha = 0.7, outlier.shape = NA, width = 0.5) +  
 labs(  
 x = "Lake",  
 y = "Total Length (mm)",  
 fill = "Lake",  
 color = "Lake"  
 ) +  
 theme\_minimal() +  
 theme(  
 plot.title = element\_text(hjust = 0.5, face = "bold"),  
 legend.position = "right"  
 )



## 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 median\_diff
2. **p-value**: The Mann-Whitney-Wilcoxon test yielded a p-value of p\_value ).
3. **Effect Size**: The rank-biserial correlation (r = r) indicates a effect\_interpretation effect size.
4. **Interpretation**: Since the p-value is p\_value < 0.05, we"fail OR reject") the null hypothesis. This indicates that the distributions of fish lengths between the two lakes are p\_value < 0.05 "significantly different", "not 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.

# 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: mm, respectively; Mann-Whitney-Wilcoxon test, W = wilcox\_test, p =(p\_value), r = r).”

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 =r.”

# 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 effect\_interpretation effect size (r = r )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 percent\_diff% difference in median lengths suggests substantial ecological differences between these habitats that warrant further investigation.