Lecture 07 H testing and simple tests II

Bill Perry

# Lecture 7 - A Brief review

* H test for a single population
* 1- and 2-sided tests
* H test for two populations
* **Assumptions of parametric tests**



# Lecture 7 overview

What we will cover today:

* Assumptions of parametric tests and how to run them
* Statistical vs. biological significance - is there a difference
* What to do when assumptions fail
  + Robust tests
  + Rank-based tests
  + Permutation tests

Lets work with the Lake Trout data as the weights are pretty cool in this one and will bakc up the main points of this lecture.

This is easily translated into the mice weight data from Vancouver or the pine needle data and we could do those too on the fly if you want….

lake trout



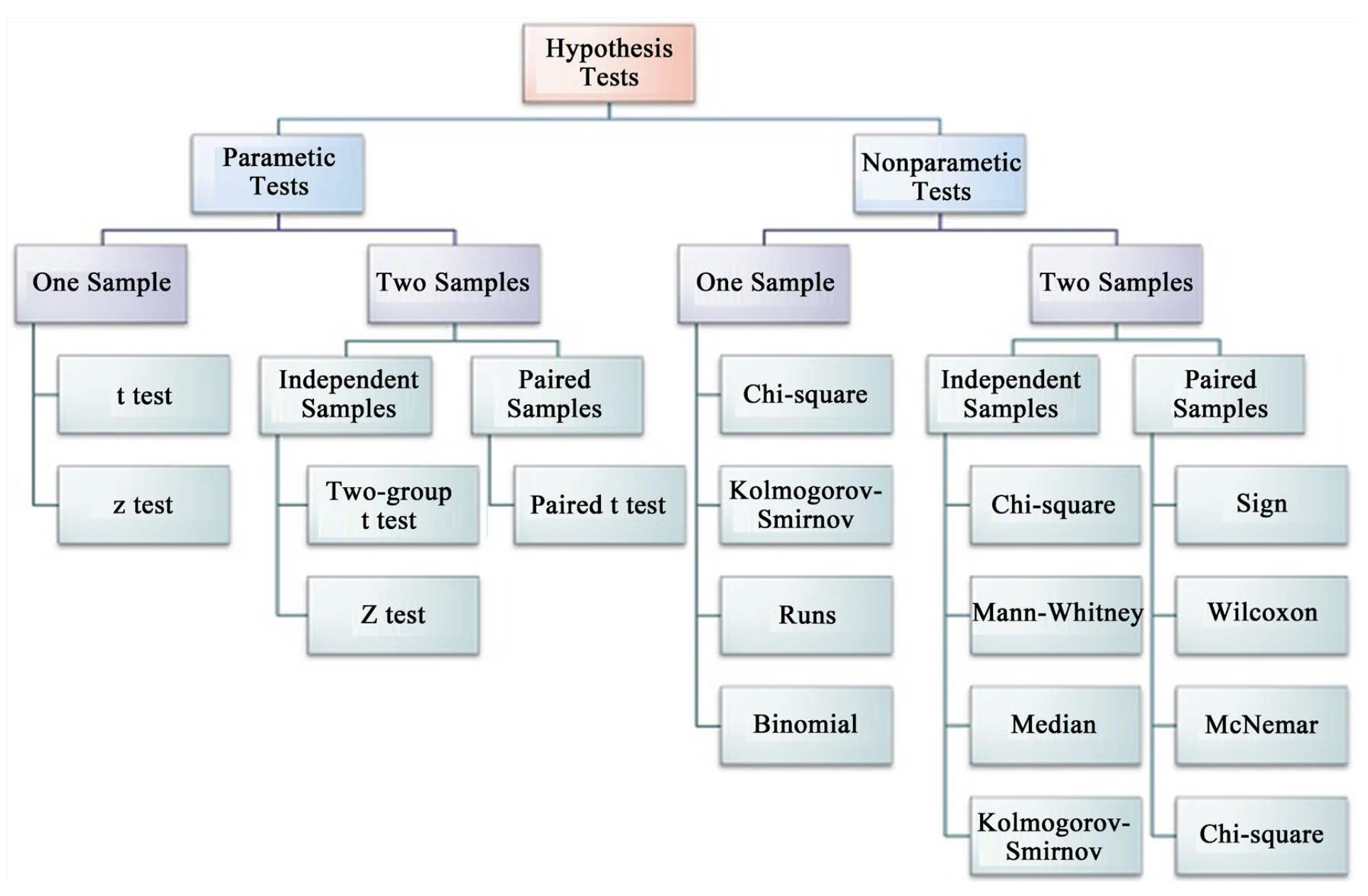
# Install packages if needed (uncomment if necessary)  
# install.packages("readr")  
# install.packages("tidyverse")  
# install.packages("car")  
# install.packages("here")  
  
# Load libraries  
library(car) # For diagnostic tests  
library(patchwork)  
library(tidyverse) # For data manipulation and visualization

# the stuff above controls the output and is also set at the top so dont need here  
# Load the pine needle data  
# Use here() function to specify the path  
df <- read\_csv("data/lake\_trout.csv")  
  
# Examine the first few rows  
head(df)

# A tibble: 6 × 5  
 sampling\_site species length\_mm mass\_g lake   
 <chr> <chr> <dbl> <dbl> <chr>  
1 I8 lake trout 515 1400 I8   
2 I8 lake trout 468 1100 I8   
3 I8 lake trout 527 1550 I8   
4 I8 lake trout 525 1350 I8   
5 I8 lake trout 517 1300 I8   
6 I8 lake trout 607 2100 I8

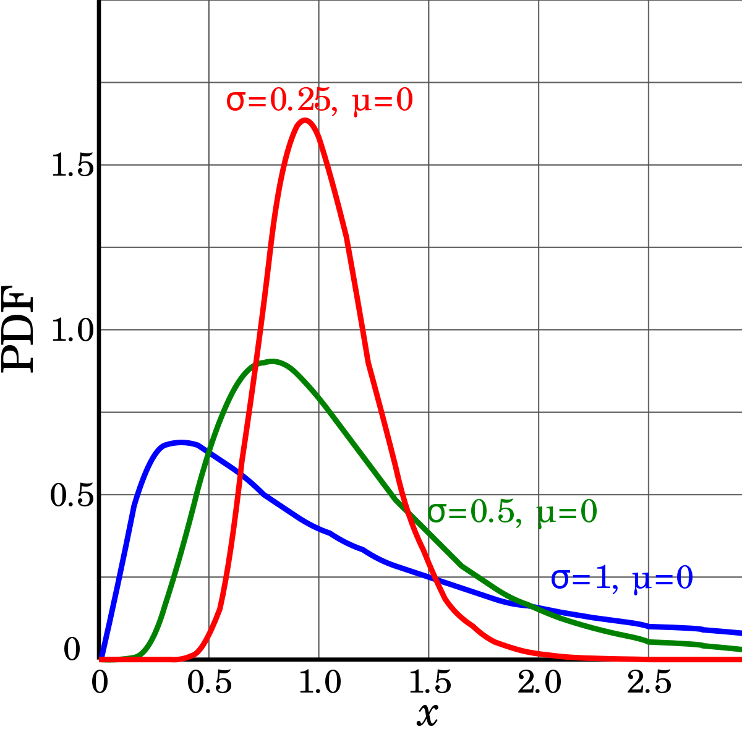
# Assumptions of parametric tests

* T-tests are **parametric** tests
  + Parametric tests: specify/assume probability distribution from which parameters came
* Non-parametric tests: no assumption about probability distribution
* Mukasa et al 2021 DOI: 10.4236/ojbm.2021.93081



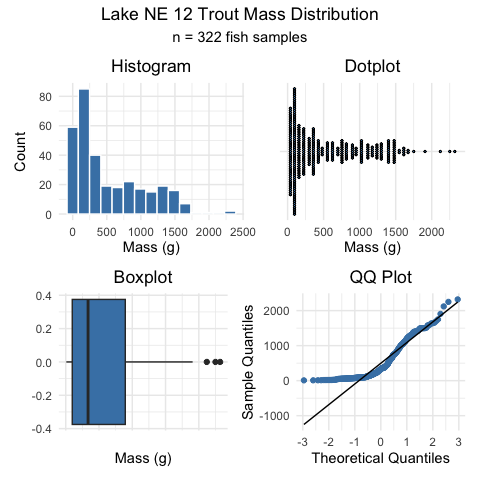
# Assumptions of parametric tests

* If assumptions of parametric test violated, test becomes unreliable
* This is because test statistic may no longer follow distribution
* Most parametric tests robust to mild/moderate violations of below assumptions



# Assumptions of parametric tests

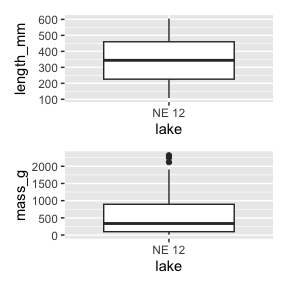
* Basic assumptions of parametric t-tests:
* Normality, equal variance, random sampling, no outliers
* Normality: Samples from normally distributed population
  + Graphical tests: histograms, dotplots, boxplots, qq-plots
  + “Formal” tests: Shapiro-Wilk test - sometimes not useful



# Assumptions of parametric tests

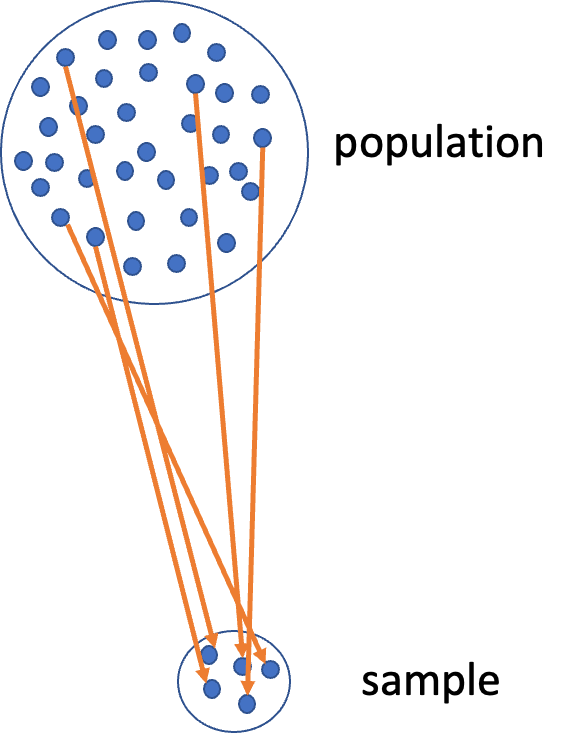
* Equal variance: samples are from populations with similar degree of variability
  + Graphical tests: boxplots
  + “Formal” tests: F-ratio test
* Parametric tests most robust to violations of normality and equal var. assumptions when samples sizes equal

length\_plot <- ne12\_data %>% ggplot(aes(x=lake, y = length\_mm)) +geom\_boxplot()   
mass\_plot <- ne12\_data %>% ggplot(aes(x=lake, y = mass\_g)) +geom\_boxplot()  
length\_plot + mass\_plot + plot\_layout(ncol=1)



# Assumptions of parametric tests

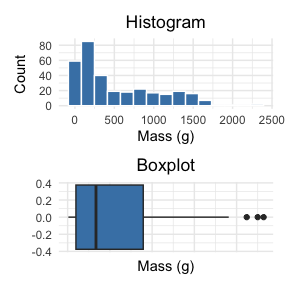
* Normality, equal variance, random sampling, no outliers
* Random sampling:
  + samples are randomly collected from populations; part of experimental design
* Necessary for sample -> population inference

<>

# Assumptions of parametric tests

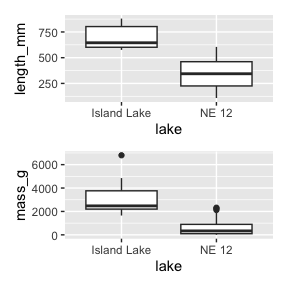
* Normality, equal variance, random sampling, no outliers
* No outliers: no “extreme” values that are very different from rest of sample
  + Graphical tests: boxplots, histograms
  + “Formal tests”: Grubb’s test - no one really does this
  + **Note: outliers also problem for non-parametric tests**

ne12\_histo\_plot+ne12\_box\_plot+ plot\_layout(ncol=1)



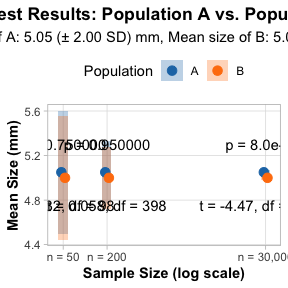
# Statistical vs. biological significance

* Statistical significance: difference unlikely due to chance
* Says nothing about biological significance of difference!
* With large sample size can detect very small differences between populations
* E.g.: consider 2 lake trout populations are the lengths the same
  + Island Lake and NE 12
    - Ho: µ~size A~ = µ~size B~
    - Ha: µ~size A~ ≠ µ~size B~



# Statistical vs. biological significance

* Size of A: 5.05 (± 2.00 SD)mm, size of B: 5.00 (± 2.00 SD)mm
* Sample 50, 200, 30,000 individuals from each pop:
  + n = 50: t = 0.32, df = 98, p-value = 0.75
  + n = 200: t = 0.058, df = 398, p-value = 0.95
  + n = 30,000: t = -4.47, df = 59998, p-value = 7.996\*10-6



# Statistical vs. biological significance

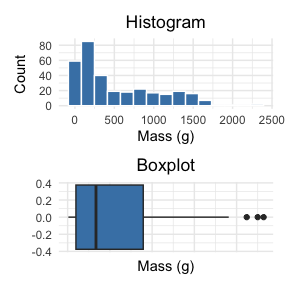
* Finally, statistically significant difference…
  + Meaningful?
  + Ecologically significant?
  + Statistics can’t answer this question
* **IMPORTANT** to report info that can assess biological significance
* “A two-tailed, two-sample independent t-test showed significant difference in size between pop. A (4.99 mm ± 1.99 SD) and pop. B (5.06 mm ± 1.99 SD) at á=0.05 (t = -4.47, df = 59998, p-value < 0.0001).”

<!– –>

# Assumptions of parametric tests

* Basic assumptions of parametric t-tests:
* Normality, equal variance, random sampling, no outliers
* What to do if assumptions are violated?

ne12\_histo\_plot+ne12\_box\_plot+ plot\_layout(ncol=1)

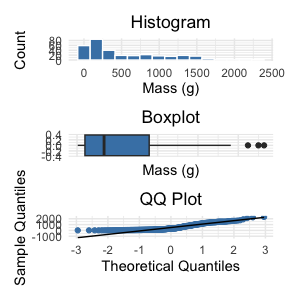


# Nonparametric test

* t-tests have several assumptions.
* Alternative tests, with more relaxed assumptions, are available to statisticians.
* In which case would you use the following tests?
  + Welch’s t-test: *when distribution normal but variance unequal*
  + Permutation test for two samples: *when distribution not normal (but both groups should still have similar distributions and ~equal variance)*
  + Mann-Whitney-Wilcoxon test: *when distribution not normal and/or outliers are present (but both groups should still have similar distributions and ~equal variance)*

< –

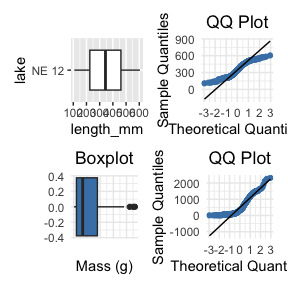
ne12\_histo\_plot+ne12\_box\_plot + ne12\_qq\_plot + plot\_layout(ncol=1)



# Assumptions of parametric tests

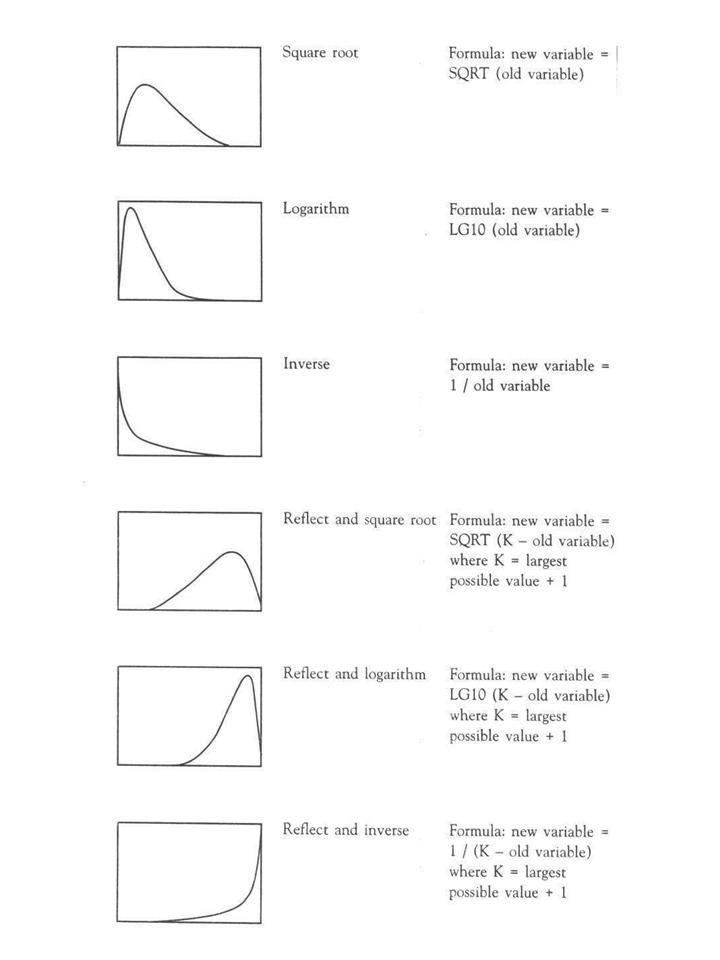
* QQ-plots: tool for assessing normality
  + On x- theoretical quantiles from SND
  + On y- ordered sample values
  + Deviation from normal can be detected as deviation from straight line

length\_ne12\_box\_plot <- isl\_ne12\_df %>% filter(lake =="NE 12") %>% ggplot(aes(x=lake, y = length\_mm)) +geom\_boxplot() + coord\_flip()  
length\_ne12\_qq\_plot <- isl\_ne12\_df %>% filter(lake =="NE 12") %>%ggplot(aes(sample = length\_mm)) +  
 stat\_qq(color = "steelblue") +  
 stat\_qq\_line() +  
 labs(title = "QQ Plot", x = "Theoretical Quantiles", y = "Sample Quantiles") +  
 theme\_minimal() +  
 theme(plot.title = element\_text(hjust = 0.5))  
(length\_ne12\_box\_plot+ length\_ne12\_qq\_plot) / (ne12\_box\_plot + ne12\_qq\_plot)



# Assumptions of parametric tests

* In some cases, data can be mathematically “transformed” to meet assumptions of parametric tests
* this can be done in r and usually involves
  + log10 transformations
  + square root transformations
  + and many others… I will have a description soon

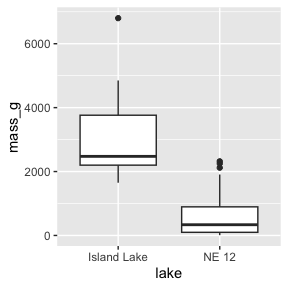
<>

[source](https://www.elsblog.org/the_empirical_legal_studi/2006/08/variable_transf.html)

# Robust tests

* **Welch’s t-test**
  + common “robust” test for means of two populations
  + Robust to violation of equal variance assumption, deals better with unequal sample size
  + Parametric test (assumes normal distribution)
  + Calculates a t statistic but recalculates df based on samples sizes and s

mass\_ne12\_plot



# Robust tests

* **Welch’s t-test**
  + **t.test(y1, y2, var.equal = FALSE, paired = FALSE)**
  + will use the Welch approach
* T-test

# T test for lenght  
# Perform standard t-test  
t\_test\_length\_result <- t.test(  
 length\_mm ~ lake,   
 data = isl\_ne12\_df,  
 var.equal = TRUE # Standard t-test with equal variance assumption  
)  
  
# Perform Welch's t-test (unequal variances)  
welch\_test\_length\_result <- t.test(  
 length\_mm ~ lake,   
 data = isl\_ne12\_df,  
 var.equal = FALSE # Welch's t-test  
)

[1] "Standard t-test results for lenght\_mm:"

Two Sample t-test  
  
data: length\_mm by lake  
t = 8.616, df = 331, p-value = 2.888e-16  
alternative hypothesis: true difference in means between group Island Lake and group NE 12 is not equal to 0  
95 percent confidence interval:  
 270.1939 430.0761  
sample estimates:  
mean in group Island Lake mean in group NE 12   
 698.200 348.065

[1] "Welch's t-test results for lenght\_mm:"

Welch Two Sample t-test  
  
data: length\_mm by lake  
t = 9.0183, df = 9.6241, p-value = 5.309e-06  
alternative hypothesis: true difference in means between group Island Lake and group NE 12 is not equal to 0  
95 percent confidence interval:  
 263.1673 437.1026  
sample estimates:  
mean in group Island Lake mean in group NE 12   
 698.200 348.065

# T test for lenght  
# Perform standard t-test  
t\_test\_mass\_result <- t.test(  
 mass\_g ~ lake,   
 data = isl\_ne12\_df,  
 var.equal = TRUE # Standard t-test with equal variance assumption  
)  
  
# Perform Welch's t-test (unequal variances)  
welch\_test\_mass\_result <- t.test(  
 mass\_g ~ lake,   
 data = isl\_ne12\_df,  
 var.equal = FALSE # Welch's t-test  
)

[1] "Standard t-test results for mass\_g:"

Two Sample t-test  
  
data: mass\_g by lake  
t = 14.181, df = 330, p-value < 2.2e-16  
alternative hypothesis: true difference in means between group Island Lake and group NE 12 is not equal to 0  
95 percent confidence interval:  
 2266.304 2996.360  
sample estimates:  
mean in group Island Lake mean in group NE 12   
 3165.0000 533.6677

[1] "Welch's t-test results for mass\_g:"

Welch Two Sample t-test  
  
data: mass\_g by lake  
t = 5.1368, df = 9.0578, p-value = 0.0006016  
alternative hypothesis: true difference in means between group Island Lake and group NE 12 is not equal to 0  
95 percent confidence interval:  
 1473.676 3788.989  
sample estimates:  
mean in group Island Lake mean in group NE 12   
 3165.0000 533.6677

# Rank based tests

* Rank-based tests: no assumptions about distribution (non-parametric)
* Ranks of data: observations assigned ranks, sums (and signs for paired tests) of ranks for groups compared
* **Mann-Whitney U test** common alternative to independent samples t-test
* **Wilcoxon signed-rank** test is alternative to paired t-test

# Perform Mann-Whitney U test (Wilcoxon rank-sum test)  
mann\_whitney\_length\_test <- wilcox.test(  
 length\_mm ~ lake,   
 data = isl\_ne12\_df,  
 exact = FALSE, # Use approximation with ties  
 conf.int = TRUE # Calculate confidence interval  
)

<

[1] "Mann-Whitney U test results length:"

Wilcoxon rank sum test with continuity correction  
  
data: length\_mm by lake  
W = 3226, p-value = 7.814e-08  
alternative hypothesis: true location shift is not equal to 0  
95 percent confidence interval:  
 262.0000 426.9999  
sample estimates:  
difference in location   
 357

# Rank based tests

* Assumptions: similar distributions for groups, equal variance
* Less power than parametric tests
* Best when normality assumption can not be met by transformation (weird distribution) or large outliers

A: n= 15, y= 8, s= 4 B : n= 15, y= 10, s= 5

Approach A vs. B

T-test df= 28 t= -3.53 p= 0.0014 M-W U (Wilcoxon’s) W= 41 p= 0.002

<

[1] "Standard t-test results for lenght\_mm:"

Two Sample t-test  
  
data: length\_mm by lake  
t = 8.616, df = 331, p-value = 2.888e-16  
alternative hypothesis: true difference in means between group Island Lake and group NE 12 is not equal to 0  
95 percent confidence interval:  
 270.1939 430.0761  
sample estimates:  
mean in group Island Lake mean in group NE 12   
 698.200 348.065

[1] "Mann-Whitney U test results length:"

Wilcoxon rank sum test with continuity correction  
  
data: length\_mm by lake  
W = 3226, p-value = 7.814e-08  
alternative hypothesis: true location shift is not equal to 0  
95 percent confidence interval:  
 262.0000 426.9999  
sample estimates:  
difference in location   
 357

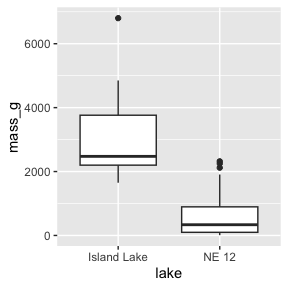
# Permutation tests

* Permutation tests based on resampling: reshuffling of original data
* Resampling allows parameter estimation when distribution unknown, including SEs and CIs of statistics (means, medians)
* Common approach is bootstrap: resample sample with replacement many times, recalculate sample stats
* Use the perm package

# Permutation tests

* Ho: µA = µB, Ha: µA ≠µB
* Calculates the differnce ∆ in means between two groups

<!



# Permutation tests

* Randomly reshuffle observations between groups (keeping nNE 12=323 and nIsland=10), calculate ∆
* Repeat >1,000 times
* Record proportion of the different means i
* This is equivalent to p-value and can be used in “traditional” H test framework
* For a graphical explanation:
  + [Graphical Explanation](https://www.jwilber.me/permutationtest/)

# Permutation tests

* In R (using ‘perm’ package):
* Assumptions: both groups have similar distribution; equal variance

library(perm)   
  
# Prepare data for permutation test  
ne12\_perm\_data <- isl\_ne12\_df %>%   
 filter(lake == "NE 12") %>%   
 pull(length\_mm)  
  
# Randomly sample exactly 25 observations from NE 12 (set seed for reproducibility)  
set.seed(123)  
ne12\_perm\_data <- sample(ne12\_perm\_data, size = 25, replace = FALSE)  
  
island\_perm\_data <- isl\_ne12\_df %>%   
 filter(lake == "Island Lake") %>%   
 pull(length\_mm)  
  
# Calculate the observed difference in means  
observed\_diff <- mean(ne12\_perm\_data, na.rm = TRUE) - mean(island\_perm\_data, na.rm = TRUE)  
  
# Perform permutation test for difference in means using perm package  
permTS(ne12\_perm\_data, island\_perm\_data,   
 alternative = "two.sided",   
 method = "exact.mc",   
 control = permControl(nmc = 10000))

Exact Permutation Test Estimated by Monte Carlo  
  
data: GROUP 1 and GROUP 2  
p-value = 2e-04  
alternative hypothesis: true mean GROUP 1 - mean GROUP 2 is not equal to 0  
sample estimates:  
mean GROUP 1 - mean GROUP 2   
 -333.08   
  
p-value estimated from 10000 Monte Carlo replications  
99 percent confidence interval on p-value:  
 0.000000000 0.001059383