Lecture 06

Bill Perry

# Lecture 6: Review

Covered

* Introduction to hypothesis testing
* The standard normal distribution
* Standard error
* Confidence intervals
* Student’s t-distribution
* H testing
* **One and Two Sample T Test**
* **p-values**



# **Lecture 6:** Overview

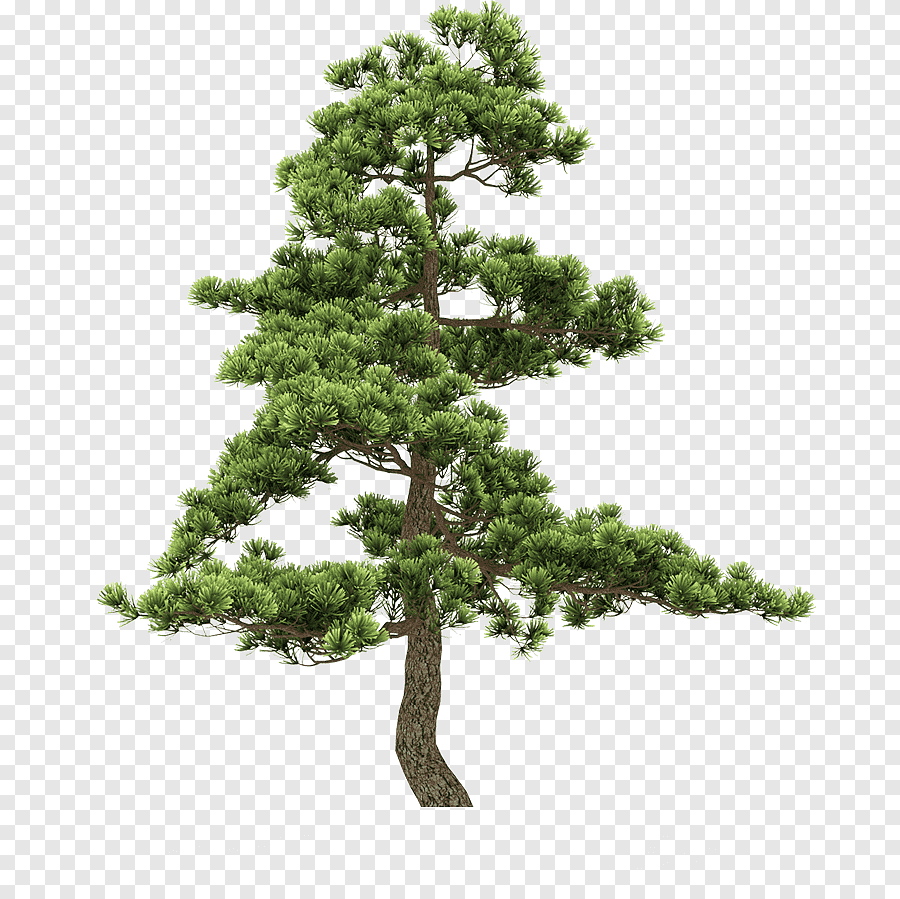
## The objectives:

* p-values
* Brief review
* H test for a single population
* 1- and 2-sided tests
* Hypothesis tests for two populations
* Assumptions of parametric tests



# **Lecture 6:** Statistical hypothesis testing

* Major goal of statistics:
  + inferences about populations from samples…
    - assign degree of confidence to inferences
  + Statistical hypothesis testing:
    - formalized approach to inference
  + Hypotheses ask whether samples come from populations with certain properties
  + Often interested in questions about population means
    - but other questions are of interest



# **Lecture 6:** Statistical hypothesis testing

Useful hypotheses: - Rely on specifying - null hypothesis (Ho) - alternate hypothesis (Ha)

* Ho is the hypothesis of “no effect”
  + two samples from population with same mean
  + sample is from population of mean = 0
* Ha (research hypothesis)
  + is the opposite of the Ho
  + or predicts that there is an effect of x on y
  + *but does NOT suggest a direction*

-

# **Lecture 6:** Statistical hypothesis testing

Together Ho and Ha encompass all possible outcomes:

* For Example:
* Ho: µ=0, Ha: µ ≠ 0
  + mean equals 0 or mean does not equal 0
* Ho: µ=35, Ha: µ ≠ 35
  + mean equals 35 or mean does not equal 35
  + Ho: µ1 = µ2, Ha: µ1 ≠ µ2
  + mean of population 1 equals mean of population 2 or it does not
  + Ho: µ > 0, Ha: µ ≤ 0
  + can be directional mean is greater than 0 or mean is not equal or less than 0



# **Lecture 6:** Statistical hypothesis testing

Tests assess likelihood of the null hypothesis being true

* If the Ho is likely false, then Ha assumed to be correct
* More precisely:
  + the long run probability of obtaining sample value (or more extreme one) if the null hypothesis is true
    - p(data|Ho) - the probability of observing the data given that the null hypothesis Ho is true

-

# **Lecture 6:** Statistical hypothesis testing

Hypothesis tests

* Expressed as p-value (0 to 1)
* Interpret p-value as:
  + probability of obtaining sample value of statistic (or more extreme one) if Ho is true
* High p-value:
  + high probability of obtaining sample statistic under Ho
    - if the null hypothesis (Ho) were true, you would frequently observe data similar to or more extreme than your sample statistic
    - your observed results are quite compatible with what the null hypothesis predicts
  + low p-value: low probability of obtaining sample statistic under Ho
    - if the null hypothesis (Ho) were true, you would rarely observe data similar to or more extreme than your sample statistic
    - Your results are unusual under the null hypothesis, suggesting that either you’ve witnessed a rare event or the null hypothesis may be incorrect

-

# **Lecture 6:** Statistical hypothesis testing

Statistical test results:

p = 0.3 means that if I repeated the study 100 times, I would get this (or more extreme) result due to chance 30 times

p = 0.03 means that if I repeated the study 100 times, I would get this (or more extreme) result due to chance 3 times

*Which p-value suggests Ho likely false?*



# **Lecture 6:** Statistical hypothesis testing

Statistical test results:

At what point reject Ho?

p < 0.05 conventional “significance threshold” (α = alpha or p value)

p < 0.05 means: if Ho is true and we repeated the study 100 times - we would get this (or more extreme) result less than 5 times due to chance



# **Lecture 6:** Statistical hypothesis testing

Statistical test results:

α is the rate at which we will reject a true null hypothesis (Type I error rate)

Lowering α will lower likelihood of incorrectly rejecting a true null hypothesis (e.g., 0.01, 0.001)

*Both Hs and α are specified* *BEFORE collection of data and analysis*



# **Lecture 6:** Statistical hypothesis testing

Traditionally α=0.05 is used as a cut off for rejecting null hypothesis

There is nothing magical about 0.05 - actual p-values need to be reported - also need to decide prior to study

| p-value range | Interpretation |
| --- | --- |
| P > 0.10 | No evidence against Ho - data appear consistent with Ho |
| 0.05 < P < 0.10 | Weak evidence against the Ho in favor of Ha |
| 0.01 < P < 0.05 | Moderate evidence against Ho in favor of Ha |
| 0.001 < P < 0.01 | Strong evidence against Ho in favor of Ha |
| P < 0.001 | Very strong evidence against Ho in favor of Ha |

# **Lecture 6:** Statistical hypothesis testing

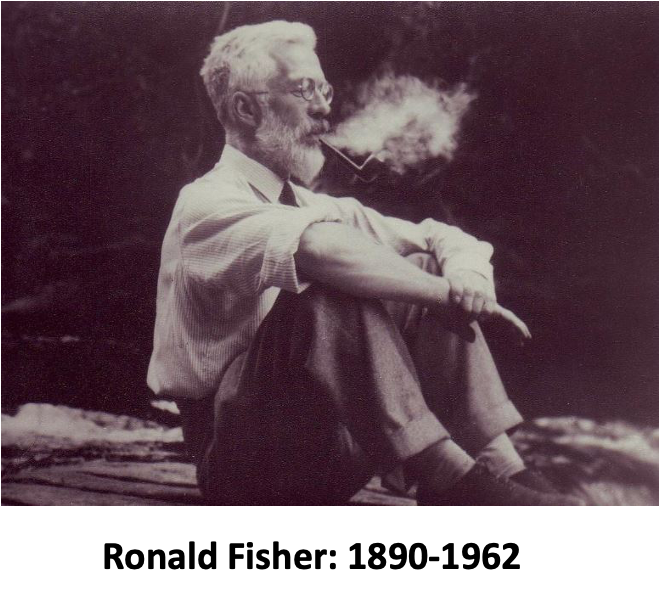


# **Lecture 6:** Statistical hypothesis testing

Fisher:

p-value as informal measure of discrepancy between data and Ho

“If p is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested. If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05 …”



# **Lecture 6:** Statistical hypothesis testing

General procedure for H testing:

* Specify Null (Ho) and alternate (Ha)
* Determine test (and test statistic) to be used
* Test statistic is used to compare your data to expectation under Ho (null hypothesis)
* Specify significance (α or p value) level below which Ho will be rejected



# **Lecture 6:** Statistical hypothesis testing

* General procedure for H testing:
* Collect data
* Perform test
  + If p-value < α, conclude Ho is likely false and reject it
  + If p-value > α, conclude no evidence Ho is false and retain it



# **Lecture 6:** Brief review

Recall…

* Major goal of statistics: inferences about populations from samples… and assign degree of confidence to inferences
* Statistical H-testing: formalized approach to inference
* Relies on specifying null hypothesis (Ho) and alternate hypothesis (Ha
* Tests assess likelihood of the null hypothesis being true
* Expressed as p-value: probability of obtaining sample value of statistic (or more extreme one) if Ho is true

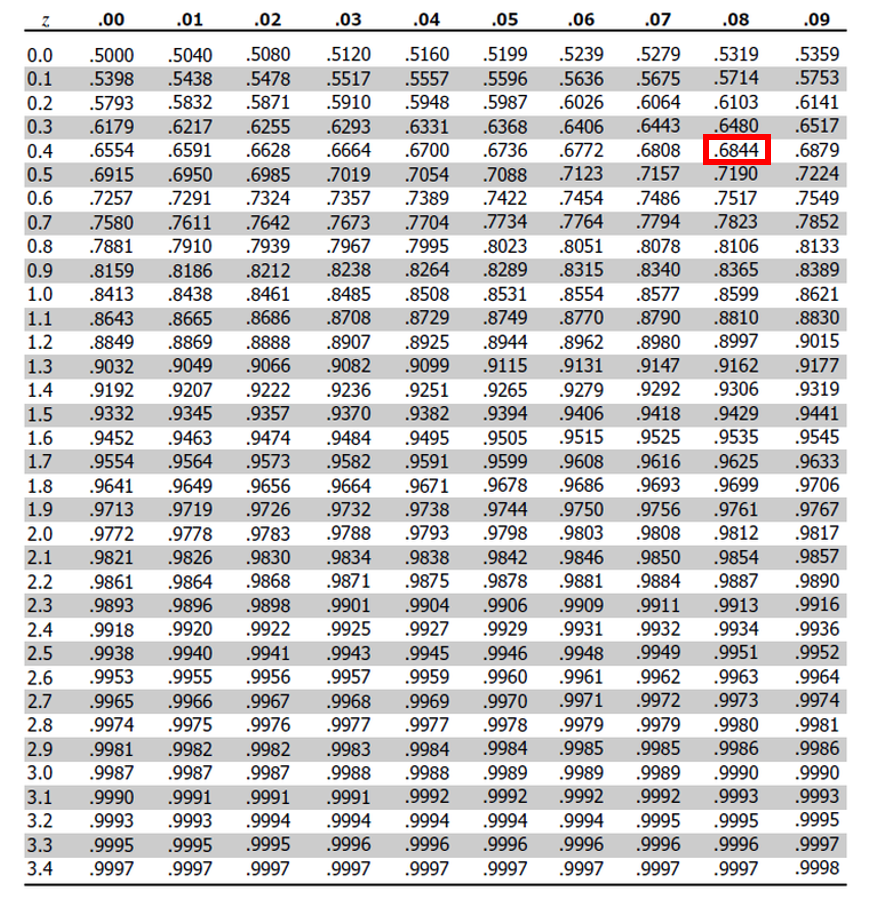
# **Lecture 6:** Brief review

Recall pine needle example

* Probability of getting sample
* with ȳ at least as far away from 21 as 35)? - p(ȳ ≤ 3500 or ȳ ≥ 3900)
  + What about - 1-tailed or 2-tailed test?
  + Can solve using SND and z-scores

# **Lecture 6:** Brief review

* z= (21-35)/40 = -0.48
  + From z table: p= 0.6368 X 2
  + p of getting sample as far away from µ as A is = 0.6368 (63.6%)
* But - usually can’t use z!
* Can use t-distribution instead…



# **Pine Needle Length: Hypothesis Testing Activity**

This activity will guide you through the process of conducting single-sample and two-sample t-tests on pine needle data. We’ll explore how environmental factors like wind exposure might affect pine needle length.

You’ll learn to:

* Formulate hypotheses
* Test assumptions
* Perform t-tests
* Visualize data
* Report results accurately



Pine needles from trees

# **Part 1:** Single Sample T-test

A single sample t-test asks whether a population parameter (like ) differs from some expected value.

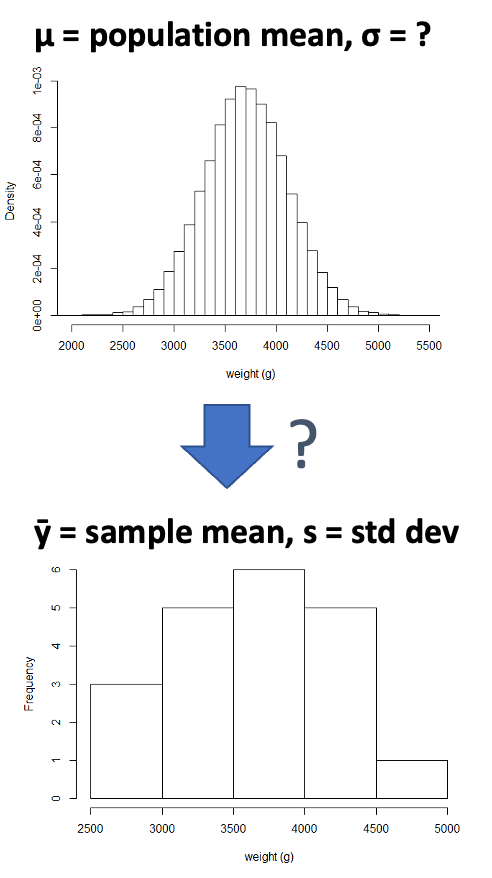
The question: **Is the average pine needle length from our windward sample different from 55mm?**

# One-sample t-test

Used when we want to compare a sample mean to a known or hypothesized population value.

where:

* is the sample mean
* is the hypothesized population mean
* is the sample standard deviation
* is the sample size



# How to do this in R

# 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(tidyverse) # For data manipulation and visualization

# Load the pine needle data  
# Use here() function to specify the path  
pine\_data <- read\_csv("data/pine\_needles.csv")  
  
# Examine the first few rows  
head(pine\_data)

# A tibble: 6 × 6  
 date group n\_s wind tree\_no len\_mm  
 <chr> <chr> <chr> <chr> <dbl> <dbl>  
1 3/20/25 cephalopods n lee 1 20  
2 3/20/25 cephalopods n lee 1 21  
3 3/20/25 cephalopods n lee 1 23  
4 3/20/25 cephalopods n lee 1 25  
5 3/20/25 cephalopods n lee 1 21  
6 3/20/25 cephalopods n lee 1 16

grayling\_df <- read\_csv("data/gray\_I3\_I8.csv")

# **Part 1:** Exploratory Data Analysis

Before conducting hypothesis tests, we should always explore our data to understand its characteristics.

Let’s calculate summary statistics and create visualizations.

**Activity: Calculate basic summary statistics for pine needle length**

# YOUR TASK: Calculate summary statistics for pine needle length  
# Hint: Use summarize() function to calculate mean, sd, n, etc.  
  
# Create a summary table for all pine needles  
pine\_summary <- pine\_data %>%  
 summarize(  
 mean\_length = mean(len\_mm),  
 sd\_length = sd(len\_mm),  
 n = n(),  
 se\_length = sd\_length / sqrt(n)  
 )  
  
print(pine\_summary)

# A tibble: 1 × 4  
 mean\_length sd\_length n se\_length  
 <dbl> <dbl> <int> <dbl>  
1 17.7 3.53 48 0.509

# Now calculate summary statistics by wind exposure  
# YOUR CODE HERE

# **Part 1:** Visualizing the Data

**Activity: Create visualizations of pine needle length**

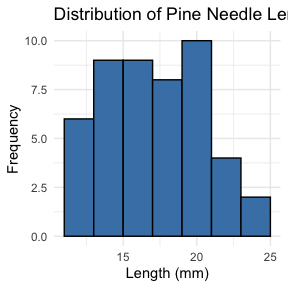
Create a histogram and a boxplot to visualize the distribution of pine needle length values.

Effective data visualization helps us understand:

* The central tendency
* The spread of the data
* Potential outliers
* Shape of distribution

# Your Task

# YOUR TASK: Create a histogram of pine needle length  
# Hint: Use ggplot() and geom\_histogram()  
  
# Histogram of all pine needle lengths  
ggplot(pine\_data, aes(x = len\_mm)) +  
 geom\_histogram(binwidth = 2, fill = "steelblue", color = "black") +  
 labs(title = "Distribution of Pine Needle Length",  
 x = "Length (mm)",  
 y = "Frequency") +  
 theme\_minimal()



# Boxplot of pine needle length by wind exposure  
# YOUR CODE HERE

# **Part 1:** Single Sample T-Test

We want to test if the mean pine needle length on the windward side differs from 55mm.

**Activity: Define hypotheses and identify assumptions**

H₀: μ = 55 (The mean pine needle length on windward side is 55mm) H₁: μ ≠ 55 (The mean pine needle length on windward side is not 55mm)

## Assumptions for t-test:

1. Data is normally distributed
2. Observations are independent
3. No significant outliers

# **Part 1:** Testing Assumptions

Before conducting our t-test, we need to verify that our data meets the necessary assumptions.

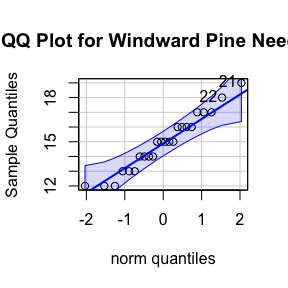
**Activity: Test the normality assumption**

Methods to test normality:

* Visual methods:
  + QQ plots or histograms
  + Statistical tests: Shapiro
  + Wilk test

# Assumptions in R - qqplots

# Filter for just windward side needles  
windward\_data <- pine\_data %>%   
 filter(wind == "wind")  
  
# YOUR TASK: Test normality of windward pine needle lengths  
# QQ Plot  
qqPlot(windward\_data$len\_mm,   
 main = "QQ Plot for Windward Pine Needles",  
 ylab = "Sample Quantiles")



[1] 21 22

## Shapiro Wilk

# Shapiro-Wilk test  
shapiro\_test <- shapiro.test(windward\_data$len\_mm)  
print(shapiro\_test)

Shapiro-Wilk normality test  
  
data: windward\_data$len\_mm  
W = 0.96062, p-value = 0.451

# Check for outliers using boxplot  
# YOUR CODE HERE

# **Part 1:** Conducting the Single Sample T-Test

Now that we’ve checked our assumptions, we can perform the single sample t-test.

**Activity: Conduct a single sample t-test to compare windward needle length to 55mm** **What is probability of getting sample at least as far from 55mm as our sample mean?**

This is our p-value, which helps us decide whether to reject the null hypothesis.

# Calculate summary statistics for windward needles  
windward\_summary <- windward\_data %>%  
 summarize(  
 mean\_length = mean(len\_mm),  
 sd\_length = sd(len\_mm),  
 n = n(),  
 se\_length = sd\_length / sqrt(n)  
 )  
  
print(windward\_summary)

# A tibble: 1 × 4  
 mean\_length sd\_length n se\_length  
 <dbl> <dbl> <int> <dbl>  
1 14.9 1.91 24 0.390

# Your Task

# YOUR TASK: Conduct a single sample t-test  
t\_test\_result <- t.test(windward\_data$len\_mm, mu = 55, var.equal = TRUE )  
print(t\_test\_result)

One Sample t-test  
  
data: windward\_data$len\_mm  
t = -102.85, df = 23, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 55  
95 percent confidence interval:  
 14.11050 15.72284  
sample estimates:  
mean of x   
 14.91667

# Calculate t-statistic manually   
# YOUR CODE HERE: t = (sample\_mean - hypothesized\_mean) / (sample\_sd / sqrt(n))  
  
# can you do this manually or manually with R?

# **Part 1:** Interpreting and Reporting Results

**Activity: Interpret the t-test results**

* What does the p-value tell us?
* Should we reject or fail to reject the null hypothesis?

**How to report this result in a scientific paper:**

“A two-tailed, one-sample t-test at α=0.05 showed that the mean pine needle length on the windward side (… mm, SD = …) [was/was not] significantly different from the expected 55 mm, t(…) = …, p = …”

# **Part 2:** Two Sample T-Test

Now, let’s compare pine needle lengths between windward and leeward sides of trees.

Question: **Is there a significant difference in needle length between the windward and leeward sides?**

This requires a two-sample t-test.

Two-sample t-test compares means from two independent groups.

where:

* x̄₁ and x̄₂: These represent the sample means of the two groups you’re comparing
* s²ₚ: This is the pooled variance, calculated as: s²ₚ = [(n₁ - 1)s₁² + (n₂ - 1)s₂²] / (n₁ + n₂ - 2), where s₁² and s₂² are the sample variances of the two groups.
* **n₁ and n₂:** These are the sample sizes of the two groups.
* **√(1/n₁ + 1/n₂):** This represents the pooled standard error.

# **Part 2:** Exploratory Data Analysis by Group

**Activity: Calculate summary statistics grouped by wind exposure** Before conducting the test, we need to understand the data for each group.

# YOUR TASK: Calculate summary statistics by wind exposure  
# Hint: Use group\_by() and summarize()  
  
group\_summary <- pine\_data %>%  
 group\_by(wind) %>%  
 summarize(  
 mean\_length = mean(len\_mm),  
 sd\_length = sd(len\_mm),  
 n = n(),  
 se\_length = sd\_length / sqrt(n)  
 )  
  
print(group\_summary)

# A tibble: 2 × 5  
 wind mean\_length sd\_length n se\_length  
 <chr> <dbl> <dbl> <int> <dbl>  
1 lee 20.4 2.45 24 0.500  
2 wind 14.9 1.91 24 0.390

# Alternative 1

# Calculate the difference in means  
# YOUR CODE HERE  
  
# Assuming your dataframe is called df  
group\_summary %>%  
 summarize(difference = mean\_length[wind == "wind"] - mean\_length[wind == "lee"])

# A tibble: 1 × 1  
 difference  
 <dbl>  
1 -5.5

# Alternative 2

# Or alternatively using filter and pull:  
lee\_mean <- group\_summary %>% filter(wind == "lee") %>% pull(mean\_length)  
wind\_mean <- group\_summary %>% filter(wind == "wind") %>% pull(mean\_length)  
difference <- wind\_mean - lee\_mean  
difference

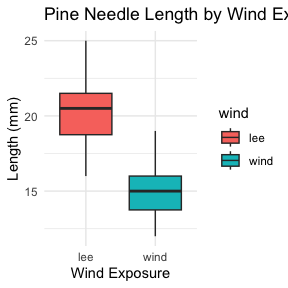
[1] -5.5

# **Part 2:** Visualizing Group Differences

**Activity: Create visualizations to compare the groups** Effective visualizations for group comparisons:

* Side-by-side boxplots
* Violin plots
* Error bar plots

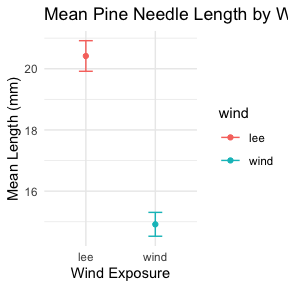
# YOUR TASK: Create boxplots to compare groups  
ggplot(pine\_data, aes(x = wind, y = len\_mm, fill = wind)) +  
 geom\_boxplot() +  
 labs(title = "Pine Needle Length by Wind Exposure",  
 x = "Wind Exposure",  
 y = "Length (mm)") +  
 theme\_minimal()



# how can you do this by wind to see both plots

# your task

# YOUR TASK: Create a plot using stat\_summary to show means and standard errors  
ggplot(pine\_data, aes(x = wind, y = len\_mm, color = wind)) +  
 stat\_summary(fun = mean, geom = "point") +  
 stat\_summary(fun.data = mean\_se, geom = "errorbar", width = 0.2) +  
 labs(title = "Mean Pine Needle Length by Wind Exposure",  
 x = "Wind Exposure",  
 y = "Mean Length (mm)") +  
 theme\_minimal()



# **Part 2:** Testing Assumptions for Two-Sample T-Test

**Activity: Test assumptions for two-sample t-test**

For a two-sample t-test, we need to check:

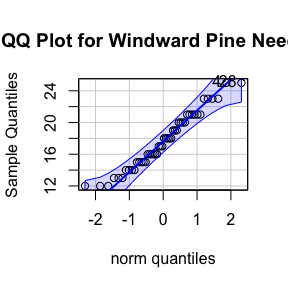
1. Normality within each group
2. Equal variances between groups (for standard t-test)
3. Independent observations

If assumptions are violated:

* Welch’s t-test (unequal variances)
* Non-parametric alternatives (Mann-Whitney U test)

# your task

# YOUR TASK: Test normality of windward pine needle lengths  
# QQ Plot  
qqPlot(pine\_data$len\_mm,   
 main = "QQ Plot for Windward Pine Needles",  
 ylab = "Sample Quantiles")



[1] 4 28

# Testing normality for each group  
# Leeward group  
lee\_data <- pine\_data %>% filter(wind == "lee")  
shapiro\_lee <- shapiro.test(lee\_data$len\_mm)  
print("Shapiro-Wilk test for leeward data:")

[1] "Shapiro-Wilk test for leeward data:"

print(shapiro\_lee)

Shapiro-Wilk normality test  
  
data: lee\_data$len\_mm  
W = 0.95477, p-value = 0.3425

# windward group

# Windward group  
# YOUR CODE HERE for windward group normality test

# Remember you can always do it in one go

# there are always two ways  
# Test for normality using Shapiro-Wilk test for each wind group  
# All in one pipeline using tidyverse approach  
normality\_results <- pine\_data %>%  
 group\_by(wind) %>%  
 summarize(  
 shapiro\_stat = shapiro.test(len\_mm)$statistic,  
 shapiro\_p\_value = shapiro.test(len\_mm)$p.value,  
 normal\_distribution = if\_else(shapiro\_p\_value > 0.05, "Normal", "Non-normal")  
 )  
  
# Print the results  
print(normality\_results)

# A tibble: 2 × 4  
 wind shapiro\_stat shapiro\_p\_value normal\_distribution  
 <chr> <dbl> <dbl> <chr>   
1 lee 0.955 0.343 Normal   
2 wind 0.961 0.451 Normal

# Conduct a Levenes Test

# Test for equal variances  
# YOUR TASK: Conduct Levene's test for equality of variances  
levene\_test <- leveneTest(len\_mm ~ wind, data = pine\_data)  
print(levene\_test)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 1.2004 0.2789  
 46

# Visual check for normality with QQ plots  
# YOUR CODE HERE

# **Part 2:** Conducting the Two-Sample T-Test

**Activity: Conduct a two-sample t-test**

Now we can compare the mean pine needle lengths between windward and leeward sides.

H₀: μ₁ = μ₂ (The mean needle lengths are equal) H₁: μ₁ ≠ μ₂ (The mean needle lengths are different)

Deciding between:

* Standard t-test (equal variances)
* Welch’s t-test (unequal variances)

# Based on our Levene’s test result.

# YOUR TASK: Conduct a two-sample t-test  
# Use var.equal=TRUE for standard t-test or var.equal=FALSE for Welch's t-test  
  
# Standard t-test (if variances are equal)  
t\_test\_result <- t.test(len\_mm ~ wind, data = pine\_data, var.equal = TRUE)  
print("Standard two-sample t-test:")

[1] "Standard two-sample t-test:"

print(t\_test\_result)

Two Sample t-test  
  
data: len\_mm by wind  
t = 8.6792, df = 46, p-value = 3.01e-11  
alternative hypothesis: true difference in means between group lee and group wind is not equal to 0  
95 percent confidence interval:  
 4.224437 6.775563  
sample estimates:  
 mean in group lee mean in group wind   
 20.41667 14.91667

# Welch's t-test (if variances are unequal)  
# YOUR CODE HERE  
  
# Calculate t-statistic manually (optional)  
# YOUR CODE HERE: t = (mean1 - mean2) / sqrt((s1^2/n1) + (s2^2/n2))

# **Part 2:** Interpreting and Reporting Two-Sample T-Test Results

**Activity: Interpret the results of the two-sample t-test**

What can we conclude about the needle lengths on windward vs. leeward sides?

**How to report this result in a scientific paper:**

“A two-tailed, two-sample t-test at α=0.05 showed [a significant/no significant] difference in needle length between windward (M = …, SD = …) and leeward (M = …, SD = …) sides of pine trees, t(…) = …, p = ….”



# **Part 3:** Paired T-Test (Extended Activity)

If we collected data in pairs (same tree, different sides), we would use a paired t-test. **How would the analysis differ?**

1. We’d calculate the difference for each pair
2. Test if the mean difference equals zero
3. The paired approach often has more statistical power

Paired t-test formula:

where:

* is the mean difference
* is the standard deviation of differences
* is the number of pairs

# **Final Activity:** Assumptions of Parametric Tests

**Common assumptions for t-tests:**

1. Normality: Data comes from normally distributed populations
2. Equal variances (for two-sample tests)
3. Independence: Observations are independent
4. No outliers: Extreme values can influence results

What can we do if our data violates these assumptions?

Alternatives when assumptions are violated:

* Data transformation (log, square root, etc.)
* Non-parametric tests
* Bootstrapping approaches

# **Summary and Conclusions**

In this activity, we’ve:

1. Formulated hypotheses about pine needle length
2. Tested assumptions for parametric tests
3. Conducted one-sample and two-sample t-tests
4. Visualized data using appropriate methods
5. Learned how to interpret and report t-test results

**Key takeaways:**

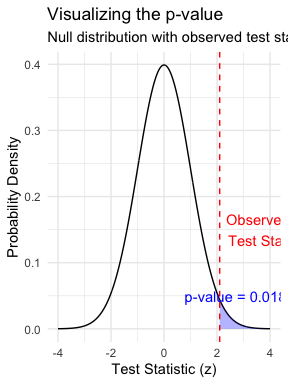
* Always check assumptions before conducting tests
* Visualize your data to understand patterns
* Report results comprehensively
* Consider alternatives when assumptions are violated

# **Lecture 5:** Understanding P-values

A **p-value** is the probability of observing the sample result (or something more extreme) if the null hypothesis is true.

**Common interpretations:** - p < 0.05: Strong evidence against H₀ - 0.05 ≤ p < 0.10: Moderate evidence against H₀ - p ≥ 0.10: Insufficient evidence against H₀

**Common misinterpretations:** - p-value is NOT the probability that H₀ is true - p-value is NOT the probability that results occurred by chance - Statistical significance ≠ practical significance



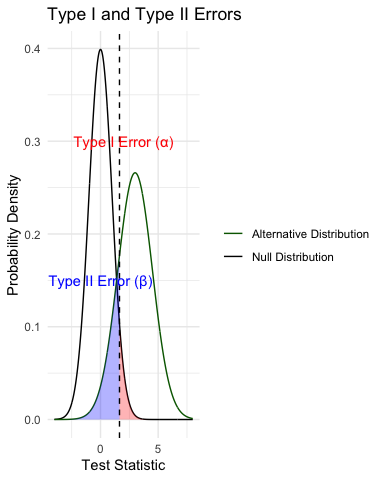
# **Lecture 5:** Type I and Type II Errors

When making decisions based on hypothesis tests, two types of errors can occur:

**Type I Error (False Positive)** - Rejecting H₀ when it’s actually true - Probability = α (significance level) - “Finding an effect that isn’t real”

**Type II Error (False Negative)** - Failing to reject H₀ when it’s actually false - Probability = β - “Missing an effect that is real”

**Statistical Power = 1 - β** - Probability of correctly rejecting a false H₀ - Increases with: - Larger sample size - Larger effect size - Lower variability - Higher α level

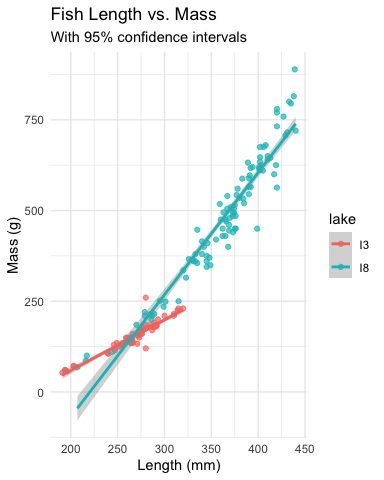


|  |
| --- |
| Practice Exercise 6: Interpreting P-values and Errors |
| Given the following scenarios, identify whether a Type I or Type II error might have occurred:   1. A researcher concludes that a new fishing regulation increased grayling size, when in fact it had no effect. 2. A study fails to detect a real decline in grayling population due to warming water, concluding there was no effect. 3. Let’s calculate the power of our t-test to detect a 30 mm difference in length between lakes:   # Calculate power for detecting a 30 mm difference # First determine parameters lake\_I3 <- grayling\_df %>% filter(lake == "I3") lake\_I8 <- grayling\_df %>% filter(lake == "I8")   n1 <- nrow(lake\_I3) n2 <- nrow(lake\_I8) sd\_pooled <- sqrt((var(lake\_I3$total\_length\_mm) \* (n1-1) +   var(lake\_I8$total\_length\_mm) \* (n2-1)) /   (n1 + n2 - 2))  # Calculate power effect\_size <- 30 / sd\_pooled # Cohen's d df <- n1 + n2 - 2 alpha <- 0.05 power <- power.t.test(n = min(n1, n2),   delta = effect\_size,  sd = 1, # Using standardized effect size  sig.level = alpha,  type = "two.sample",  alternative = "two.sided")  # Display results power  Two-sample t test power calculation    n = 66  delta = 0.6741298  sd = 1  sig.level = 0.05  power = 0.9702076  alternative = two.sided  NOTE: n is number in \*each\* group |

# **Lecture 5:** Summary

**Key concepts covered:**

1. **Probability distributions** model random phenomena
   * Normal distribution is especially important
   * Z-scores standardize measurements
2. **Standard error** measures precision of estimates
   * Decreases with larger sample sizes
   * Used to construct confidence intervals
3. **Confidence intervals** express uncertainty
   * Provide plausible range for parameters
   * 95% CI: mean ± 1.96 × SE
4. **Hypothesis testing** evaluates claims
   * Null vs. alternative hypotheses
   * P-values quantify evidence against H₀
   * Consider both statistical and practical significance



|  |
| --- |
| Final Exercise: Comprehensive Analysis |
| Now that we’ve covered the key concepts, let’s perform a complete analysis of the Arctic grayling data:  # Comprehensive analysis of Arctic grayling data # 1. Data visualization length\_boxplot <- grayling\_df %>%  ggplot(aes(x = lake, y = total\_length\_mm, fill = lake)) +  geom\_boxplot() +  labs(title = "Fish Length by Lake",  x = "Lake",  y = "Length (mm)") +  theme\_minimal()  # 2. Compare means with t-test length\_ttest <- t.test(total\_length\_mm ~ lake, data = grayling\_df)  # 3. Length-mass relationship length\_mass\_model <- lm(mass\_g ~ total\_length\_mm \* lake, data = grayling\_df) model\_summary <- summary(length\_mass\_model)  # 4. Display results length\_boxplot    length\_ttest  Welch Two Sample t-test  data: total\_length\_mm by lake t = -15.532, df = 161.63, p-value < 2.2e-16 alternative hypothesis: true difference in means between group I3 and group I8 is not equal to 0 95 percent confidence interval:  -109.32342 -84.66053 sample estimates: mean in group I3 mean in group I8   265.6061 362.5980  model\_summary  Call: lm(formula = mass\_g ~ total\_length\_mm \* lake, data = grayling\_df)  Residuals:  Min 1Q Median 3Q Max  -151.223 -14.839 -0.764 10.670 153.130   Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -219.3313 47.9087 -4.578 9.30e-06 \*\*\* total\_length\_mm 1.3924 0.1794 7.763 8.88e-13 \*\*\* lakeI8 -522.5506 56.5882 -9.234 < 2e-16 \*\*\* total\_length\_mm:lakeI8 1.9738 0.1972 10.009 < 2e-16 \*\*\* --- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  Residual standard error: 40.93 on 162 degrees of freedom  (2 observations deleted due to missingness) Multiple R-squared: 0.9644, Adjusted R-squared: 0.9637  F-statistic: 1461 on 3 and 162 DF, p-value: < 2.2e-16  # 5. Calculate 95% confidence intervals for each lake ci\_results <- grayling\_df %>%  group\_by(lake) %>%  summarize(  mean\_length = mean(total\_length\_mm, na.rm = TRUE),  sd\_length = sd(total\_length\_mm, na.rm = TRUE),  n = sum(!is.na(total\_length\_mm)),  se\_length = sd\_length / sqrt(n),  t\_crit = qt(0.975, df = n - 1),  margin\_error = t\_crit \* se\_length,  ci\_lower = mean\_length - margin\_error,  ci\_upper = mean\_length + margin\_error,  .groups = "drop"  )  # Display confidence intervals ci\_results  # A tibble: 2 × 9  lake mean\_length sd\_length n se\_length t\_crit margin\_error ci\_lower  <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> 1 I3 266. 28.3 66 3.48 2.00 6.96 259. 2 I8 363. 52.3 102 5.18 1.98 10.3 352. # ℹ 1 more variable: ci\_upper <dbl>  # 6. Visualize regression with confidence intervals regression\_plot <- grayling\_df %>%  ggplot(aes(x = total\_length\_mm, y = mass\_g, color = lake)) +  geom\_point(alpha = 0.7) +  geom\_smooth(method = "lm", se = TRUE) +  labs(title = "Length-Mass Relationship by Lake",  x = "Length (mm)",  y = "Mass (g)") +  theme\_minimal()  regression\_plot    Based on this analysis: 1. Are there significant differences in fish length between the two lakes? 2. How does the length-mass relationship differ between lakes? 3. What conclusions can you draw about Arctic grayling in these two lakes? |

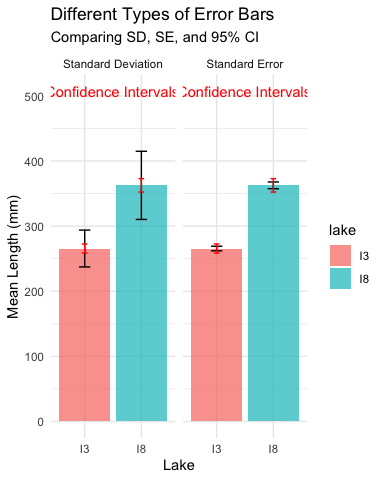
# **Lecture 5:** Error Bars and Their Interpretation

Error bars are graphical representations of the variability of data that show:

* The **precision** of a measurement
* The **uncertainty** around an estimate
* A **confidence interval** for a parameter

Common types of error bars: 1. **Standard Error (SE)**: Shows precision of the mean 2. **Standard Deviation (SD)**: Shows variability in the data 3. **Confidence Interval (CI)**: Shows plausible range for parameter

When interpreting graphs: - Always check what the error bars represent - Non-overlapping 95% CI bars suggest statistically significant differences - Error bars help assess both statistical and practical significance



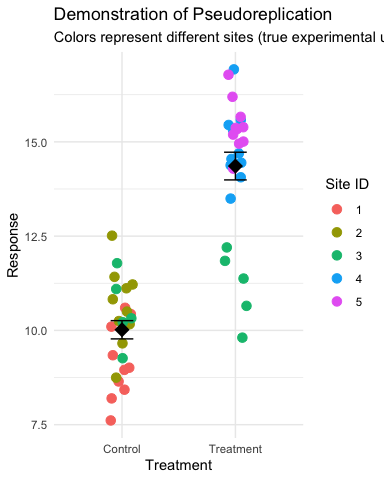
# **Lecture 5:** Sampling and Pseudoreplication

**Pseudoreplication** occurs when measurements that are not independent are analyzed as if they were independent.

* A critical consideration in experimental design
* Results in underestimated standard errors and confidence intervals
* Leads to inflated Type I error rates (false positives)

**Examples of pseudoreplication:** - Measuring the same individual multiple times - Treating multiple fish from the same tank as independent - Using multiple data points from a single site

**How to avoid pseudoreplication:** - Identify the true experimental unit - Use appropriate statistical techniques (e.g., mixed models) - Be clear about the level of replication

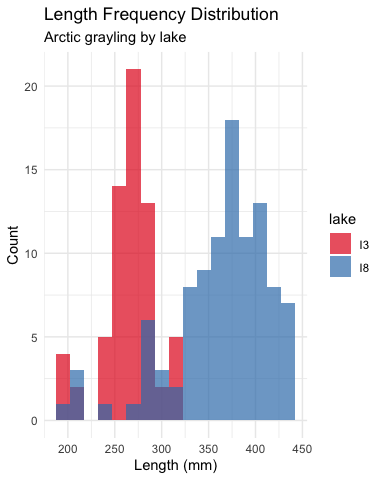


# **Lecture 5:** Practical Applications in Fish Biology

The statistical concepts we’ve covered today are essential for fisheries biologists and ecologists:

* **Z-scores** help identify unusual fish sizes in a population
* **Standard error** quantifies uncertainty in growth rate estimates
* **Confidence intervals** provide plausible ranges for population parameters
* **Hypothesis testing** evaluates effects of management practices
* **P-values** determine significance of environmental impacts

**Real-world applications:** - Assessing population health and structure - Evaluating effectiveness of fishing regulations - Quantifying relationships between fish size and habitat variables - Predicting impacts of climate change on fish populations - Designing effective conservation strategies



# **Lecture 5:** Statistical hypothesis testing

Major goal of statistics:

inferences about populations from samples assign degree of confidence to inferences

Statistical H-testing:

formalized approach to inference

* hypotheses ask whether samples come from populations with certain properties
* often interested in questions about population means (but not only)

# **Lecture 5:** Statistical hypothesis testing

Relies on specifying null hypothesis (Ho) and alternate hypothesis (Ha)

* Ho is the hypothesis of “no effect”
  + (two samples from population with same mean, sample is from population of mean=0)
* Ha (research hypothesis) the opposite of the Ho

|  |
| --- |
| Practice Exercise 10: Formulating Hypotheses |
| For the following scenarios, write out the null and alternative hypotheses:   1. Testing if the mean fish length in Lake S 06 is greater than 50 mm. 2. Testing if there is a difference in mean fish lengths between lakes Toolik and S 06. 3. Testing if lake E 01 has a higher variance in fish lengths compared to Lake Toolik.   For each scenario, remember that the null hypothesis typically represents “no effect” or “no difference”, while the alternative hypothesis represents what you are trying to demonstrate. |

# **Lecture 5:** Statistical hypothesis testing

* p = 0.3 means that if study repeated 100 times
  + would get this (or more extreme) result due to chance 30 times
* p = 0.03 means that if study repeated 100 times
  + would get this (or more extreme) result due to chance 3 times

Which p-value suggests Ho likely false?

# **Lecture 5:** Statistical hypothesis testing

At what point reject Ho?

* p < 0.05 conventional “significance threshold” (α)
* p < 0.05 means:
  + if Ho is true - if study repeated 100 times
    - would get this (or more extreme) result less than 5 times due to chance

# **Lecture 5:** Statistical hypothesis testing

* α is the rate at which we will reject a true null hypothesis (Type I error rate)
* Lowering α will lower likelihood of incorrectly rejecting a true null hypothesis (e.g., 0.01, 0.001)
* Both hypotheses and α are specified *BEFORE* collection of data and analysis

# **Lecture 5:** Statistical hypothesis testing

Traditionally α=0.05 is used as a cut off for rejecting null hypothesis

Nothing magical about 0.0 - actual p-values need to be reported.

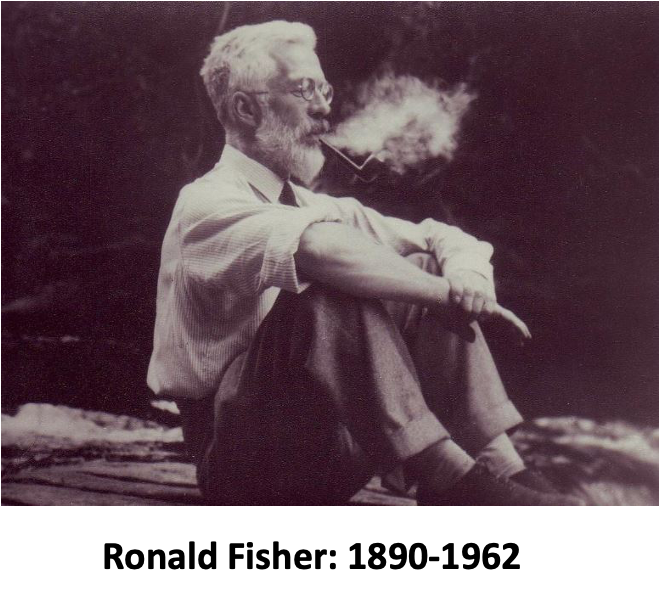
| p-value range | Interpretation |
| --- | --- |
| P > 0.10 | No evidence against Ho - data appear consistent with Ho |
| 0.05 < P < 0.10 | Weak evidence against the Ho in favor of Ha |
| 0.01 < P < 0.05 | Moderate evidence against Ho in favor of Ha |
| 0.001 < P < 0.01 | Strong evidence against Ho in favor of Ha |
| P < 0.001 | Very strong evidence against Ho in favor of Ha |

# **Lecture 5:** Statistical hypothesis testing

Fisher:

p-value as informal measure of discrepancy betwen data and Ho

“If p is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested. If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05 …”

s

# **Lecture 5:** Statistical hypothesis testing

General procedure for H testing:

* Specify Null (Ho) and alternate (Ha)
* Determine test (and test statistic) to be used
* Test statistic is used to compare your data to expectation under Ho (null hypothesis)
* Specify significance (α or p value) level below which Ho will be rejected

# **Lecture 5:** Statistical hypothesis testing

General procedure for H testing:

* Collect data - Perform test
* If p-value < α, conclude Ho is likely false and reject it
* If p-value > α, conclude no evidence Ho is false and retain it

|  |
| --- |
| Tip |
|  |

# **Lecture 5:** Next Steps in Statistical Analysis

In future lectures, we’ll explore:

* One-sample and two-sample t-tests
* Analysis of variance (ANOVA)
* Linear regression and correlation
* Chi-square tests
* Non-parametric methods
* Multiple regression and model selection
* Mixed effects models

Each method builds on the statistical foundation we’ve established today, applying probability concepts to make inferences from data.

|  |
| --- |
| Learning Resources |
| * Practice problems in the textbook (Chapter 4 & 5) * Online resources:   + Khan Academy: Probability and Statistics   + StatQuest with Josh Starmer (YouTube channel)   + R for Data Science (r4ds.had.co.nz) * Office hours: Wednesdays 2-4pm |