Lecture 05: Probability and Statistical Inference

Bill Perry

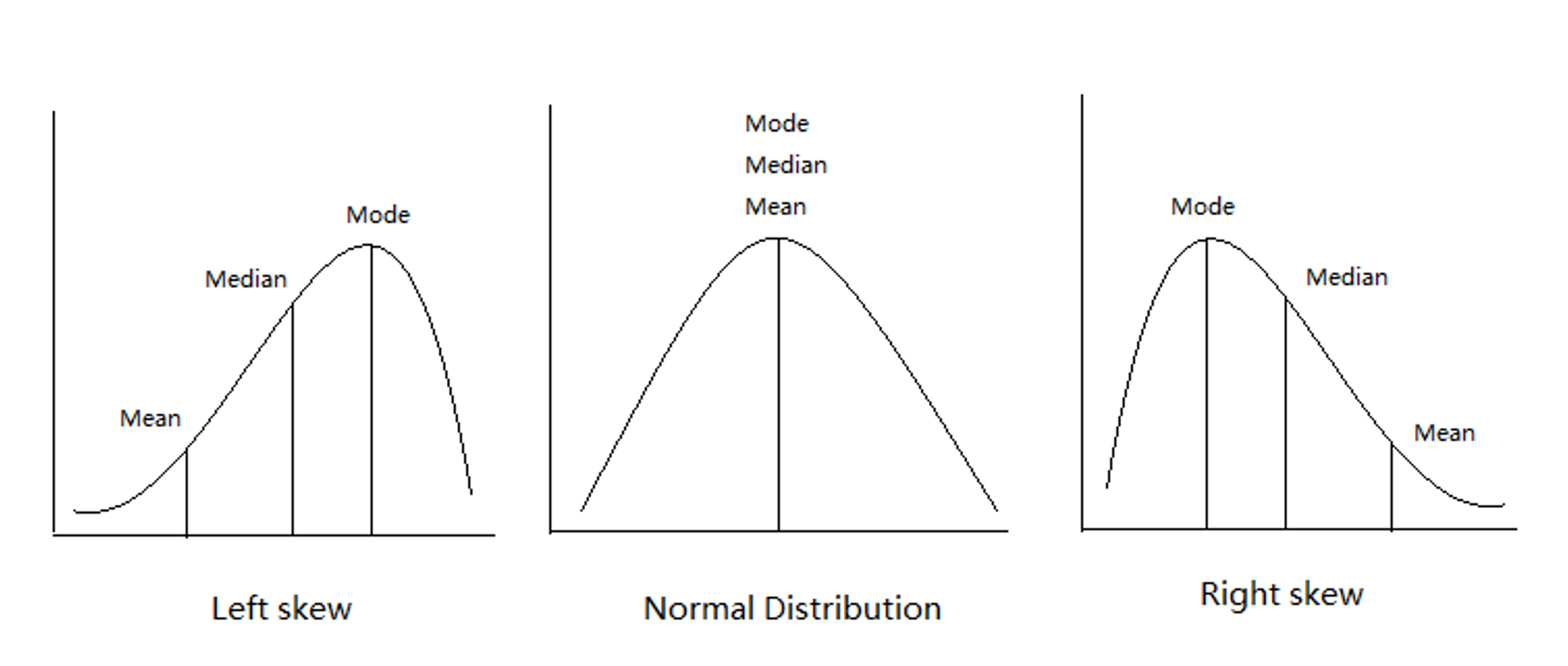
# **Lecture 4: Review**

* Introduction to histograms or frequency distributions
* Probability Distribution Functions (PDF)
* Descriptive Statistics
  + Center - mean, median, mode
  + Spread - range, variance, standard deviation



## **Lecture 4: Review**

* Introduction to histograms or frequency distributions
* Probability Distribution Functions (PDF)
* Descriptive Statistics
  + Center - mean, median, mode
  + Spread - range, variance, standard deviation



## **Lecture 4: Review**

* Introduction to histograms or frequency distributions
* Probability Distribution Functions (PDF)
* Descriptive Statistics
  + Center - mean, median, mode
  + Spread - range, variance, standard deviation

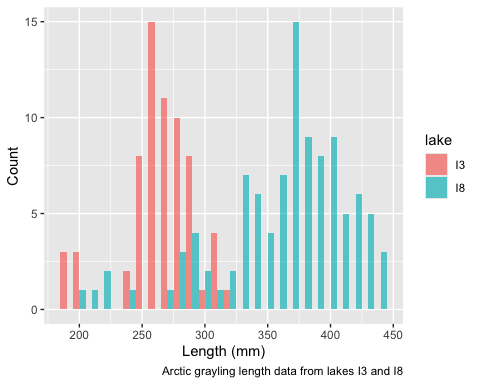
library(tidyverse)  
library(patchwork)  
library(car) # For diagnostic tests  
  
grayling\_df <- read\_csv("data/gray\_I3\_I8.csv")   
i3\_df <- grayling\_df %>% filter(lake=="I3")  
  
# Calculate summary statistics  
grayling\_summary <- grayling\_df %>%   
 group\_by(lake) %>%  
 summarize(  
 mean\_length = mean(length\_mm, na.rm = TRUE),  
 sd\_length = sd(length\_mm, na.rm = TRUE),  
 se\_length = sd\_length/sqrt(sum(!is.na(length\_mm))),  
 count = sum(!is.na(length\_mm)),  
 .groups = "drop")  
grayling\_summary

# A tibble: 2 × 5  
 lake mean\_length sd\_length se\_length count  
 <chr> <dbl> <dbl> <dbl> <int>  
1 I3 266. 28.3 3.48 66  
2 I8 363. 52.3 5.18 102

## **Lecture 5: Probability and Statistical Inference**

The goals for today

* Statistical inference fundamentals
* Hypothesis testing principles
* T Distributions
* One sample T Tests
* Two sample T Test

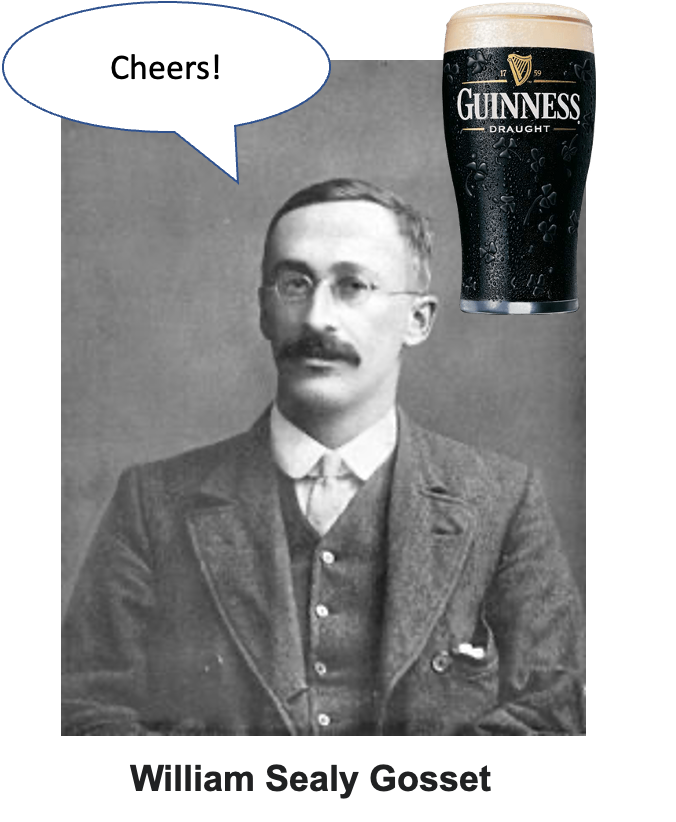


# **Lecture 5:** Confidence intervals

In the more typical case DON’T know the population σ or standard deviation

* estimate it from the samples
* and when sample size is <~30)
* can’t use the standard normal (z) distribution

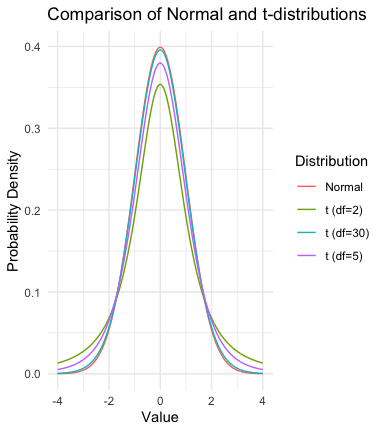
*Instead, we use Student’s t distribution*



# **Lecture 5:** Understanding t-distribution

When sample sizes are small, the **t-distribution** is more appropriate than the normal distribution.

* Similar to normal distribution but with heavier tails
* Shape depends on **degrees of freedom** (df = n-1)
* With large df (>30), approaches the normal distribution
* Used for:
  + Small sample sizes
  + When population standard deviation is unknown
  + Calculating confidence intervals
  + Conducting t-tests



# **Lecture 5:** Understanding t-distribution

When sample sizes are small, the **t-distribution** is more appropriate than the normal distribution.

* Similar to normal distribution (1.96 = 2.5% tails) but with heavier tails
* Shape depends on **degrees of freedom** (df = n-1)
* With large df (>30), approaches the normal distribution
* Used for:
  + Small sample sizes
  + When population standard deviation is unknown
  + Calculating confidence intervals
  + Conducting t-tests



|  |
| --- |
| Practice Exercise 4: Using the t-distribution |
| Let’s compare confidence intervals using the normal approximation (z) versus the t-distribution for our fish data.  # Calculate CI using both z and t distributions for a smaller subset small\_sample <- grayling\_df %>%   filter(lake == "I3") %>%   slice\_sample(n = 10)  # Calculate statistics sample\_mean <- mean(small\_sample$length\_mm) sample\_sd <- sd(small\_sample$length\_mm) sample\_n <- nrow(small\_sample) sample\_se <- sample\_sd / sqrt(sample\_n)  # Calculate confidence intervals z\_ci\_lower <- sample\_mean - 1.96 \* sample\_se z\_ci\_upper <- sample\_mean + 1.96 \* sample\_se  # For t-distribution, get critical value for 95% CI with df = n-1 t\_crit <- qt(0.975, df = sample\_n - 1) t\_ci\_lower <- sample\_mean - t\_crit \* sample\_se t\_ci\_upper <- sample\_mean + t\_crit \* sample\_se  # Display results cat("Mean:", round(sample\_mean, 1), "mm\n")  Mean: 267.5 mm  cat("Standard deviation:", round(sample\_sd, 2), "mm\n")  Standard deviation: 30.87 mm  cat("Standard error:", round(sample\_se, 2), "mm\n")  Standard error: 9.76 mm  cat("95% CI using z:", round(z\_ci\_lower, 1), "to", round(z\_ci\_upper, 1), "mm\n")  95% CI using z: 248.4 to 286.6 mm  cat("95% CI using t:", round(t\_ci\_lower, 1), "to", round(t\_ci\_upper, 1), "mm\n")  95% CI using t: 245.4 to 289.6 mm  cat("t critical value:", round(t\_crit, 3), "vs z critical value: 1.96\n")  t critical value: 2.262 vs z critical value: 1.96 |

# Student’s t-distribution

To calculate CI for sample from “unknown” population:

Where:

* ȳ is sample mean
* 𝑛 is sample size
* s is sample standard deviation
* t t-value corresponding the probability of the CI
* t in t-table for different degrees of freedom (n-1)



# **Lecture 5:** Student’s t-distribution

Here is a t-table

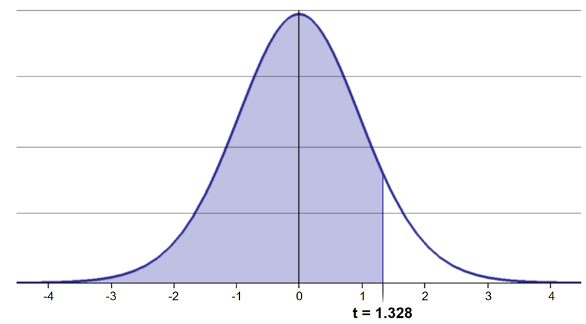
* Values of t that correspond to probabilities
* Probabilities listed along top
* Sample dfs are listed in the left-most column
* Probabilities are given for one-tailed and two-tailed “questions”

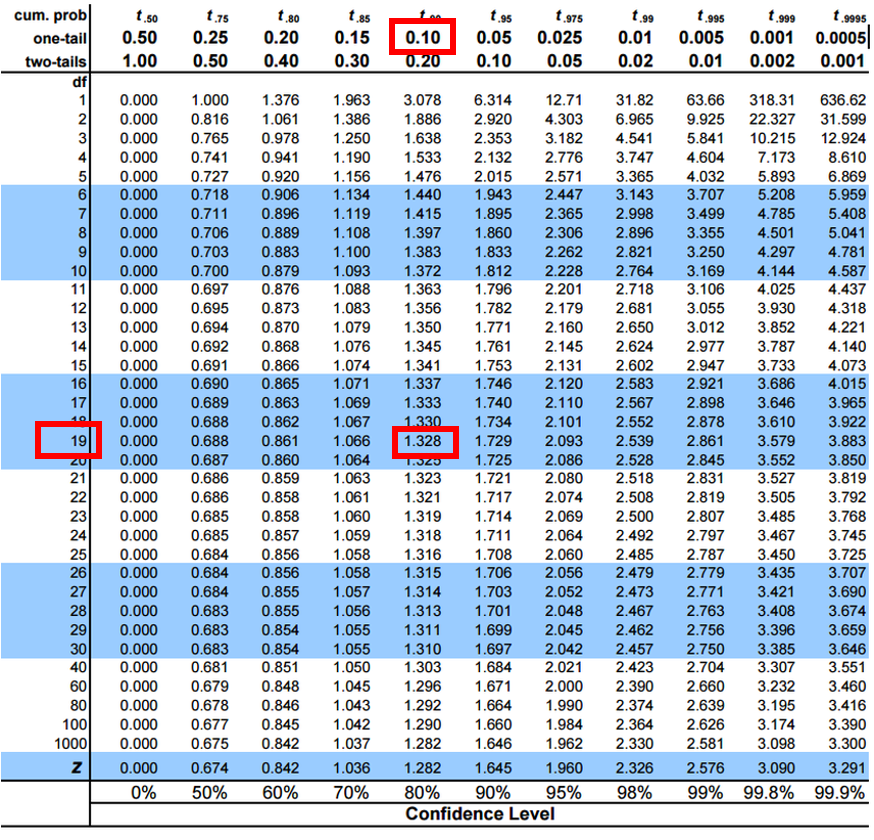


# **Lecture 5:** Student’s t-distribution

One-tailed questions: area of distribution left or (right) of a certain value

* n=20 (df=19) - 90% of the observations found left
* t= 1.328 (10% are outside)

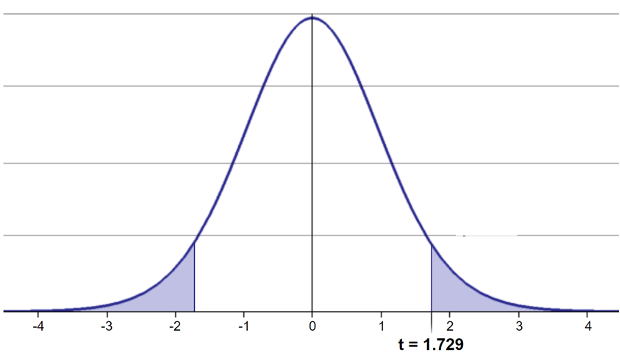


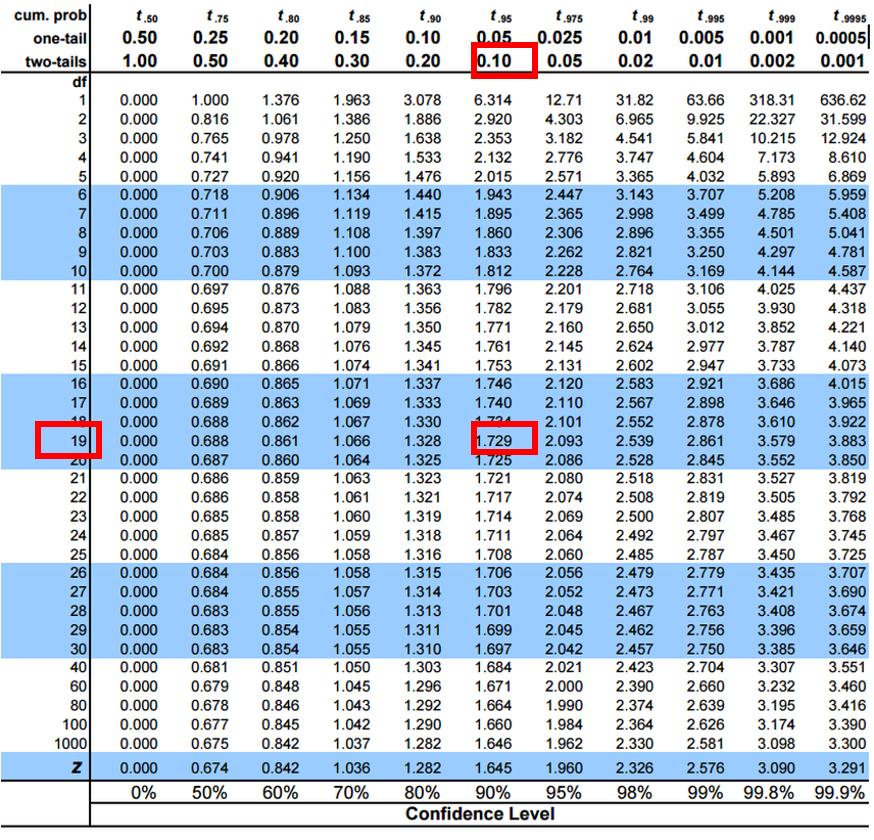


# **Lecture 5:** Student’s t-distribution

Two-tailed questions refer to area between certain values

* n= 20 (df=19), 90% of the observations are between
* t=-1.729 and t=1.729 (10% are outside)





# **Lecture 5:** Student’s t-distribution

Let’s calculate CIs again:

Use two-sided test

* 95% CI Sample A: = 272.8 ± 2.262 \* (37.81/(9^0.5)) = 1.650788
* The 95% CI is between 244.3 and 301.3
* “The 95% CI for the population mean from sample A is 272.8 ± 28.5”



# **Lecture 5:** Student’s t-distribution

So:

* Can assess confidence that population mean is within a certain range
* Can use t distribution to ask questions like:
  + “What is probability of getting sample with mean = ȳ from population with mean = µ?” (1 sample t-test)
  + “What is the probability that two samples came from same population?” (2 sample t-test)

# **Lecture 5:** Single Sample T-Test

We want to test if the mean fish length in I3 differs from 240mm.

**Activity: Define hypotheses and identify assumptions**

H₀: μ = 240 (The mean fish length in I3 is 240mm)

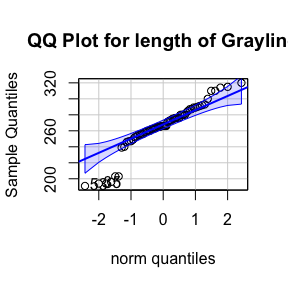
H₁: μ ≠ 240 (The mean fish length in I3 is not 55mm)

## Assumptions for t-test:

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

# Assumptions in R - qqplots from car

# Filter for just windward side needles  
  
# YOUR TASK: Test normality of windward pine needle lengths  
# QQ Plot  
qqPlot(i3\_df$length\_mm,   
 main = "QQ Plot for length of Grayling",  
 ylab = "Sample Quantiles")



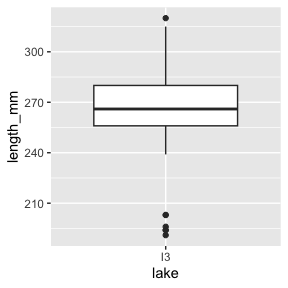
[1] 53 35

## Statistical Test of Normality - Shapiro-Wilk test

# Shapiro-Wilk test  
shapiro\_test <- shapiro.test(i3\_df$length\_mm)  
print(shapiro\_test)

Shapiro-Wilk normality test  
  
data: i3\_df$length\_mm  
W = 0.91051, p-value = 0.0001623

# Check for outliers using boxplot  
# YOUR CODE HERE  
i3\_df %>% ggplot(aes(lake, length\_mm))+geom\_boxplot()



# Now for Practice

|  |
| --- |
| Practice Exercise 1: One-Sample t-Test |
| Let’s perform a one-sample t-test to determine if the mean fish length in I3 Lake differs from 240 mm:  # what is the mean i3\_mean <- mean(i3\_df$length\_mm, na.rm=TRUE) cat("Mean:", round(i3\_mean, 1), "mm\n")  Mean: 265.6 mm  # Perform a one-sample t-test t\_test\_result <- t.test(i3\_df$length\_mm, mu = 240)  # View the test results t\_test\_result  One Sample t-test  data: i3\_df$length\_mm t = 7.3497, df = 65, p-value = 4.17e-10 alternative hypothesis: true mean is not equal to 240 95 percent confidence interval:  258.6481 272.5640 sample estimates: mean of x   265.6061  Interpret this test result by answering these questions:   1. What was the null hypothesis? 2. What was the alternative hypothesis? 3. What does the p-value tell us? 4. Should we reject or fail to reject the null hypothesis at α = 0.05? 5. What is the practical interpretation of this result for fish biologists? |

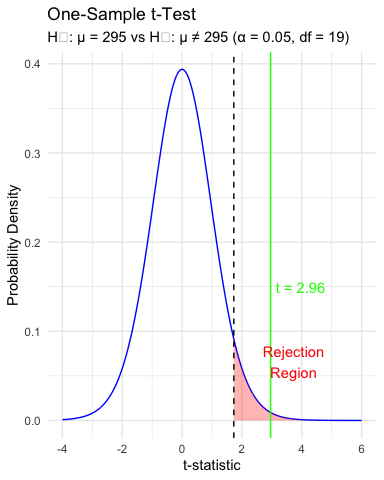
# **Lecture 5:** Intro to Hypothesis Testing

Hypothesis testing is a systematic way to evaluate research questions using data.

**Key components:**

1. **Null hypothesis (Ho)**: Typically assumes “no effect” or “no difference”
2. **Alternative hypothesis (Ha)**: The claim we’re trying to support
3. **Statistical test**: Method for evaluating evidence against H₀
4. **P-value**: Probability of observing our results (or more extreme) if H₀ is true
5. **Significance level (α)**: Threshold for rejecting H₀, typically 0.05

**Decision rule**: Reject Ho if p-value < α == p < 0.05



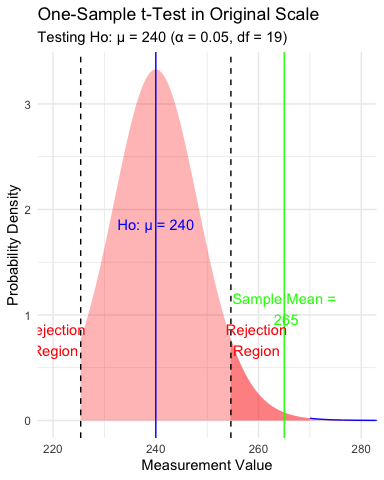
# **Lecture 5:** Intro to Hypothesis Testing

Hypothesis testing is a systematic way to evaluate research questions using data.

**Key components:**

1. **Null hypothesis (H₀)**: Typically assumes “no effect” or “no difference”
2. **Alternative hypothesis (Hₐ)**: The claim we’re trying to support
3. **Statistical test**: Method for evaluating evidence against H₀
4. **P-value**: Probability of observing our results (or more extreme) if H₀ is true
5. **Significance level (α)**: Threshold for rejecting H₀, typically 0.05

**Decision rule**: Reject H₀ if p-value < α



# Lecture 5: Interpreting and Reporting Results of a 1 sample T Test

**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 = …”

# **Lecture 5:** Next steps - two sample T Tests

For example

* what is probability that population X is the same as population Y?

How would you assess this question using what we learned?

This is what we will do with the pine needles…



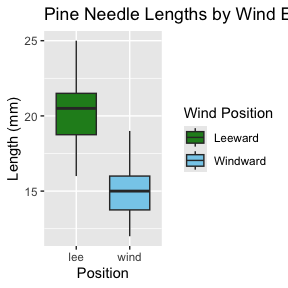
# **Lecture 5:** Next steps - two samples

For example

* what is probability that population X is the same as population Y?

How would you assess this question using what we learned?

# Now create a boxplot to visualize the difference in fish lengths between these lakes:  
pine\_df <- read\_csv("data/pine\_needles.csv")  
  
# Create a boxplot comparing the two lakes  
pine\_wind\_plot <- pine\_df %>%  
 ggplot(aes(x = wind, y = length\_mm, fill = wind)) +  
 geom\_boxplot() +  
 labs(title = "Pine Needle Lengths by Wind Exposure",  
 x = "Position",  
 y = "Length (mm)",  
 fill = "Wind Position") +  
 scale\_fill\_manual(values = c("lee" = "forestgreen", "wind" = "skyblue"),  
 labels = c("lee" = "Leeward", "wind" = "Windward"))  
pine\_wind\_plot



# Based on the t-test results and the boxplot  
#   
# what can you conclude about the fish populations in these two lakes?

|  |
| --- |
| Practice Exercise 2: Formulating Hypotheses |
| For the following research questions about pine needles write the null and alternative hypotheses:   1. Are needles on the lee side longer than the needles on the windy side?   What are the hypotheses?  Ho =  Ha = |

# **Lecture 5:** Next steps - two samples

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.

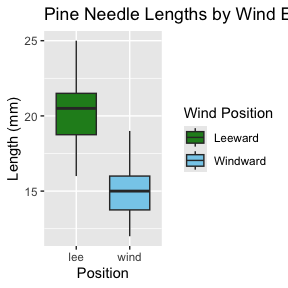
Two-sample t-test compares means from two independent groups.

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.

|  |
| --- |
| Practice Exercise 3: **Calculate summary statistics grouped by wind exposure** |
| Before conducting the test, we need to understand the data for each group.   1. You need this and the graph to see what is goin on ….  * group\_summary <- pine\_df %>%  group\_by(wind) %>%  summarize(  mean\_length = mean(length\_mm),  sd\_length = sd(length\_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 |

# Create a boxplot comparing the two sides  
pine\_wind\_plot



|  |
| --- |
| Practice Exercise 4: Effect size |
| We could also look at the difference in means… some cool 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 |

|  |
| --- |
| Practice Exercise 5: Using GGPLOT to get summary stats |
| GGplot also has code to make the mean and standard error plots we are interested in along whit a lot of others  # Assuming your dataframe is called df pine\_mean\_se\_plot <- ggplot(pine\_df, aes(x = wind, y = length\_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)") +  coord\_cartesian(ylim = c(0,25))+  scale\_color\_manual(values = c("lee" = "forestgreen", "wind" = "skyblue"),  labels = c("lee" = "Leeward", "wind" = "Windward"))+  theme\_classic() pine\_mean\_se\_plot |

# **Lecture 5:** Testing 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)

|  |
| --- |
| Practice Exercise 6: Test normality of windward pine needle lengths |
| qqplots  Note you need to test each groups separately…  # Assuming your dataframe is called df pine\_mean\_se\_plot |

|  |
| --- |
| Practice Exercise 7: Test normality of windward pine needle lengths |
| qqplots  Note you need to test each groups separately…  # how do you make separate dataframes to do this on? # Separate data by groups windward\_data <- pine\_df %>% filter(wind == "wind") leeward\_data <- pine\_df %>% filter(wind == "lee") head(leeward\_data)  # A tibble: 6 × 6  date group n\_s wind tree\_no length\_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 |

|  |
| --- |
| Practice Exercise 8: Test normality of windward pine needle lengths |
| qqplots  Note you need to test each groups separately…  # QQ Plot for windward group qqPlot(windward\_data$length\_mm,   main = "QQ Plot for Windward Pine Needles",  ylab = "Sample Quantiles")    [1] 21 22 |

|  |
| --- |
| Practice Exercise 9: Test normality of windward pine needle lengths |
| Shapiro-Wilk test  Note you need to test each groups separately…  # Shapiro-Wilk test for windward group shapiro\_windward <- shapiro.test(windward\_data$length\_mm) print("Shapiro-Wilk test for windward data:")  [1] "Shapiro-Wilk test for windward data:"  print(shapiro\_windward)  Shapiro-Wilk normality test  data: windward\_data$length\_mm W = 0.96062, p-value = 0.451 |

|  |
| --- |
| Practice Exercise 10: Test normality of windward pine needle lengths |
| qqplots  Note you need to test each groups separately…  # You can also test the leeward group # QQ Plot for leeward group qqPlot(leeward\_data$length\_mm,   main = "QQ Plot for Leeward Pine Needles",  ylab = "Sample Quantiles")    [1] 4 16 |

|  |
| --- |
| Practice Exercise 11: Test normality of windward pine needle lengths |
| Shapiro-Wilk test  Note you need to test each groups separately…  # Shapiro-Wilk test for leeward group shapiro\_lee <- shapiro.test(leeward\_data$length\_mm) print("Shapiro-Wilk test for leeward data:")  [1] "Shapiro-Wilk test for leeward data:"  print(shapiro\_lee)  Shapiro-Wilk normality test  data: leeward\_data$length\_mm W = 0.95477, p-value = 0.3425 |

|  |
| --- |
| Practice Exercise 12: Test Normality at one time |
| There are always a lot of ways to do this in R  # 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\_df %>%  group\_by(wind) %>%  summarize(  shapiro\_stat = shapiro.test(length\_mm)$statistic,  shapiro\_p\_value = shapiro.test(length\_mm)$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 |

|  |
| --- |
| Practice Exercise 13: Test equal variances |
| Levenes test can be done on the original dataframe  # Method 1: Using car package's leveneTest # This is often preferred as it's more robust to departures from normality levene\_result <- leveneTest(length\_mm ~ wind, data = pine\_df) print("Levene's Test for Homogeneity of Variance:")  [1] "Levene's Test for Homogeneity of Variance:"  print(levene\_result)  Levene's Test for Homogeneity of Variance (center = median)  Df F value Pr(>F) group 1 1.2004 0.2789  46 |

# **Lecture 5:** Conducting the Two-Sample T-Test

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

Ho: μ₁ = μ₂ (The mean needle lengths are equal)

Ha: μ₁ ≠ μ₂ (The mean needle lengths are different)

Deciding between:

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

**Note the Levenes Test should be NOT SIGNIFICANT - What is the null hypothesis**

levene\_result

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

# **Lecture 5:** Conducting the Two-Sample T-Test

Now we can do a two sample TTEST

Calculate t-statistic manually (optional)

YOUR CODE HERE:

t = (mean1 - mean2) / sqrt((s1^2/n1) + (s2^2/n2))

# 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(length\_mm ~ wind, data = pine\_df, var.equal = TRUE)  
print("Standard two-sample t-test:")

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

print(t\_test\_result)

Two Sample t-test  
  
data: length\_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

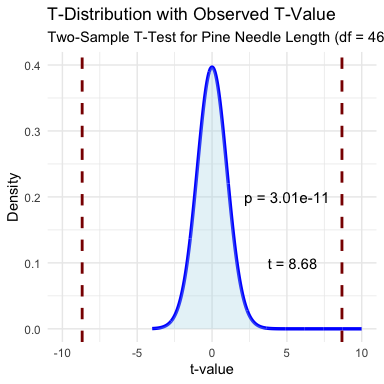
# **Lecture 5:** Interpreting and Reporting Two-Sample T-Test Results

**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 = ….”



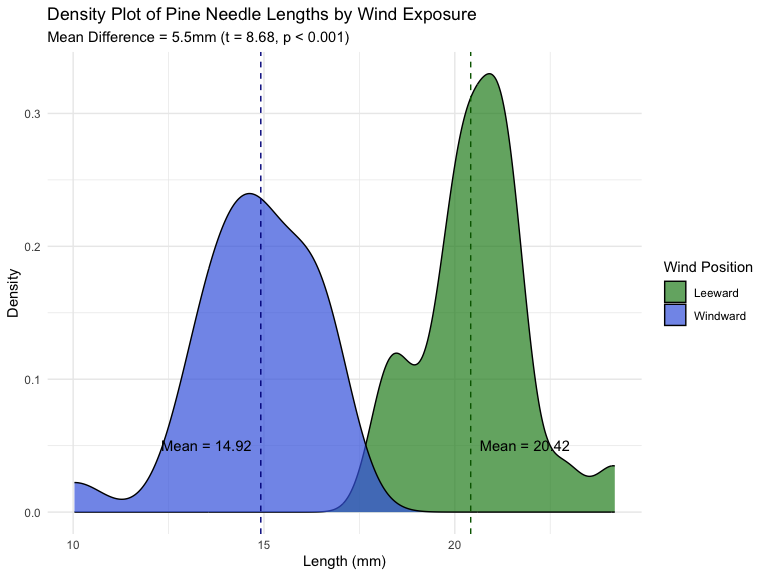
# **Lecture 5:** Interpreting and Reporting Two-Sample T-Test Results

**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 = ….”



# **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

# **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