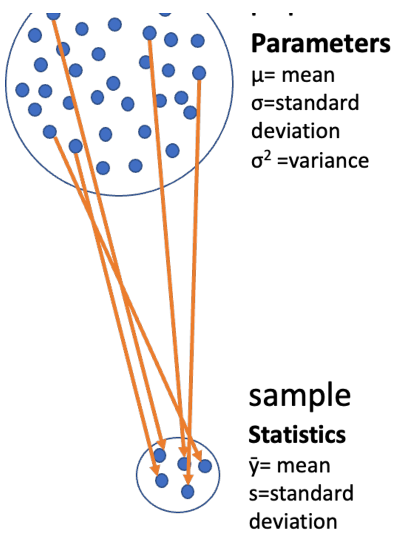
Lecture 06

Bill Perry

Lecture 5: Review

Covered

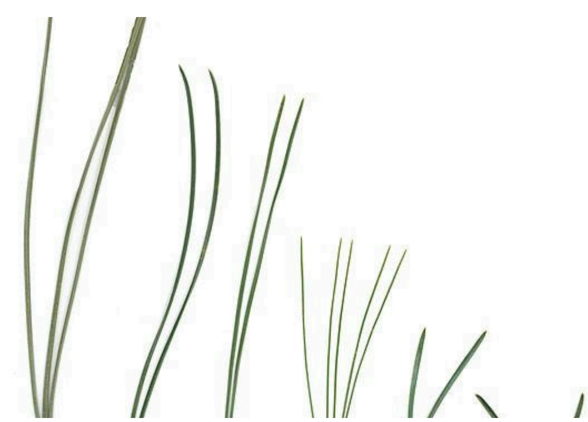
- Statistical inference fundamentals
- Hypothesis testing principles
- T Distributions
- One sample T Tests
- Two sample T



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: Hypothesis Components

Useful hypotheses:

- Rely on specifying
  - Ho is the hypothesis of "no effect"
    - two samples from population with same mean
    - sample is from population of mean = X
  - Ha (research or alternate hypothesis)
    - is the opposite of the Ho
    - or predicts that there is an effect of x on y
    - *but does NOT suggest a direction which is a prediction*

Lecture 6: Hypothesis Examples

Together Ho and Ha encompass all possible outcomes:

- Ho: $\mu=0$, Ha: $\mu \neq 0$
  - mean equals 0 or mean does not equal 0
- Ho: $\mu = 35$, Ha: $\mu \neq 35$
  - mean equals 35 or mean does not equal 35
- Ho: $\mu 1 = \mu 2$, Ha: $\mu 1 \neq \mu 2$
  - mean of population 1 equals mean of population 2 or it does not
- Ho: $\mu > 0$, Ha: $\mu \leq 0$
  - can be directional mean is greater than 0 or mean is not equal or less than 0
  - this becomes a one sided test as it predicts only one direction

Lecture 6: Statistical Testing Framework

Statistical 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: Understanding P-values

Hypothesis tests

- Expressed as p-value (0-never to 1-always)
- 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 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: P-value Interpretation

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?

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: Significance Levels and Interpretation

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*

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

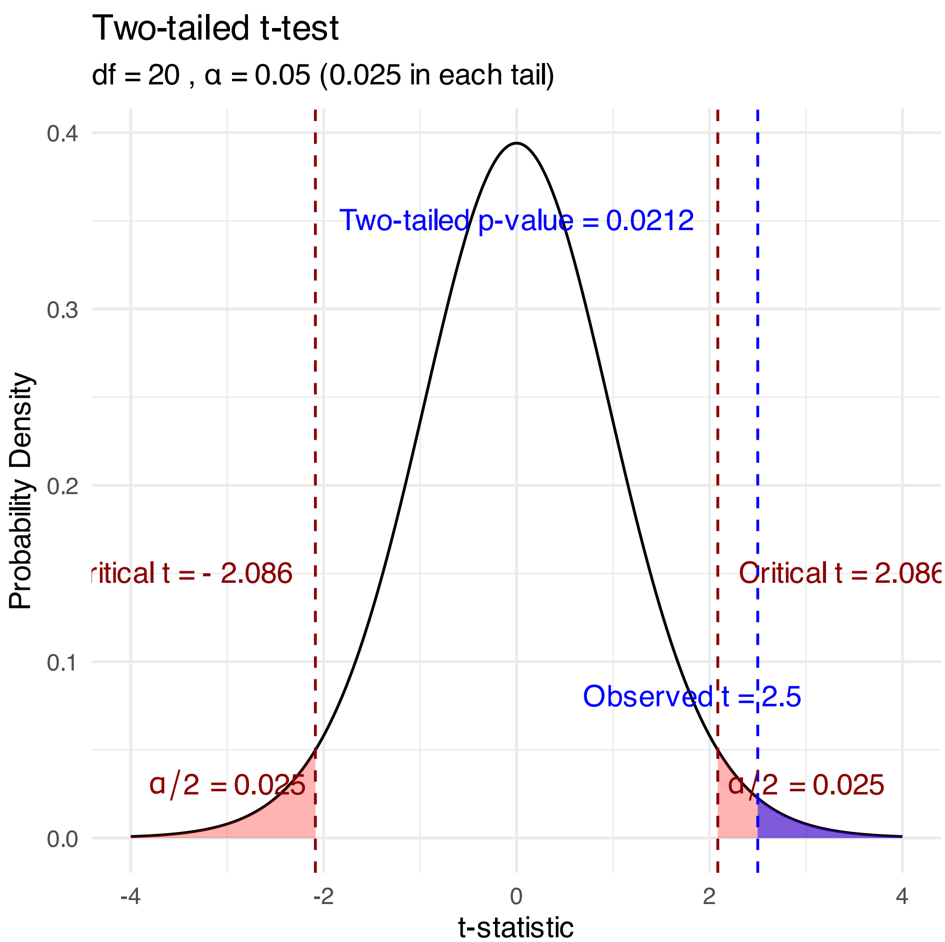
There is nothing magical about 0.05 - actual p-values need to be reported - also need to decide prior to study

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

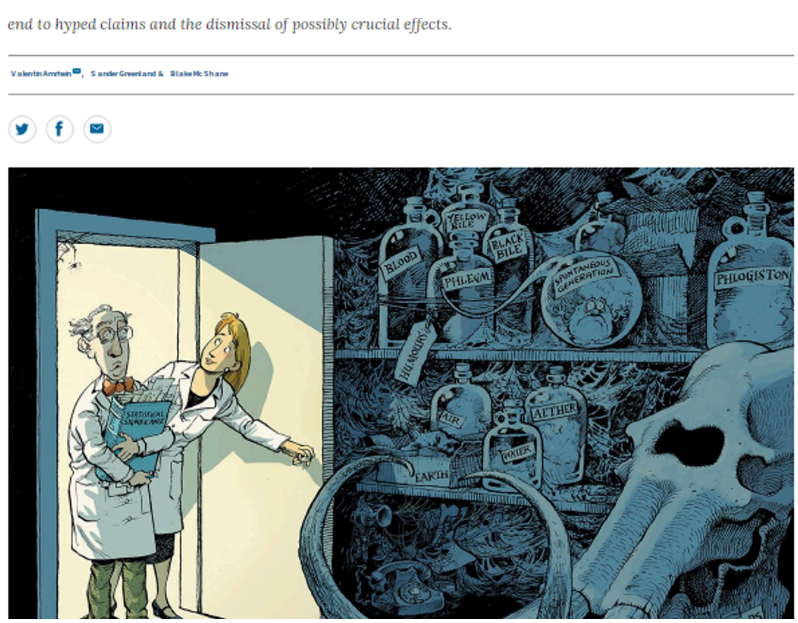
Lecture 6: Understanding P-values Visually

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 H_0
  - $0.05 \leq p < 0.10$: Moderate evidence against H_0
  - $p \geq 0.10$: Insufficient evidence against H_0
- **Common misinterpretations:**
  - p-value is NOT the probability that H_0 is true
  - p-value is NOT the probability that results occurred by chance
  - Statistical significance ≠ practical significance
- Note that there is a difference in how to state the hypotheses
  - one sample TTEST
  - two sample TTEST



Lecture 6: Historical Context

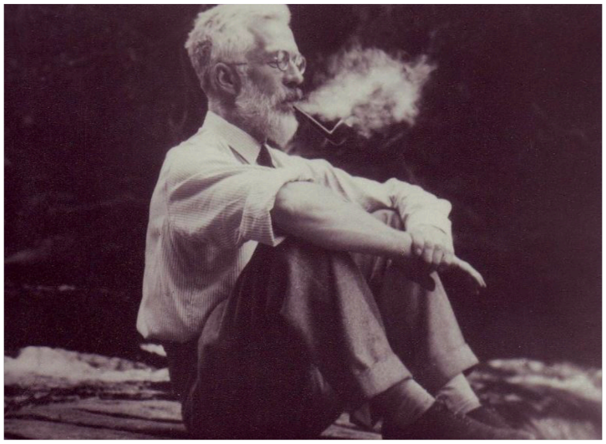


Lecture 6: Fisher's Perspective

Fisher:

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

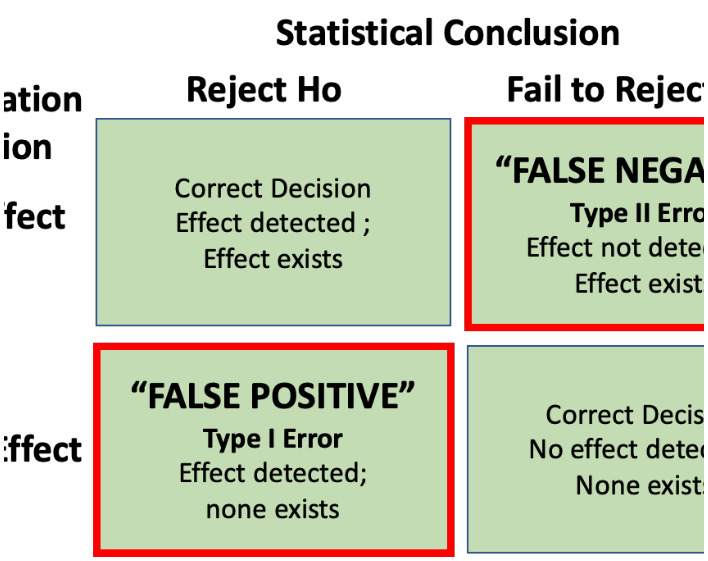
“If p is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested. If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05 …”



Ronald Fisher: 1890-1962

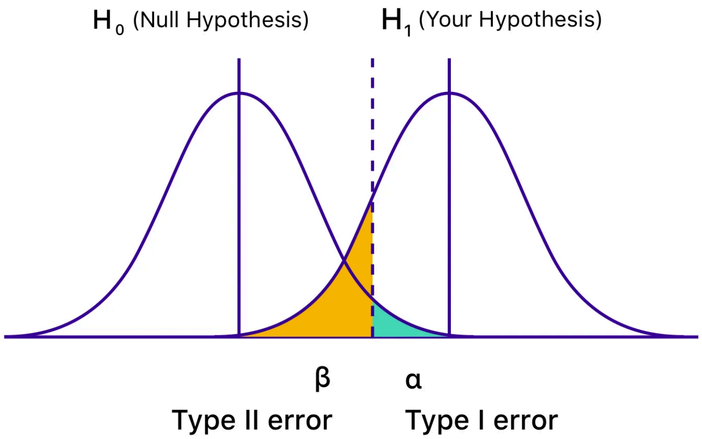
Decision errors

- Even good studies can reach incorrect conclusions
- “Decision errors”
- Two types of decision errors
- Want to know probability of making these errors



Type I and Type II Errors - Concept

- **Type I error rate**
  - **α**: wrongly reject H_0 when it’s true
  - $\alpha = 0.05$ means a type I error rate of 5%
- **Type II error rate, β**
  - wrongly fail to reject H_0 when it’s false
- **Power = 1-β**: probability of correctly rejecting H_0 when H_1 is true
- Inverse relationship between type I and type II error - but not straightforward
- Result of chance - sample not representative of population
- Which type of error is more dangerous?



the dotted line is the alpha = 0.05

Lecture 6: Type I and Type II Errors - Details

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

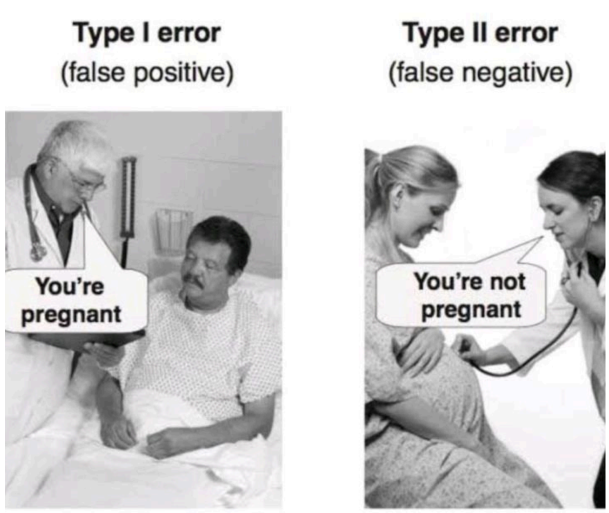
Type I Error (False Positive) - Rejecting H_0 when it's actually true - Probability = α (significance level) - "Finding an effect that isn't real"

Type II Error (False Negative) - Failing to reject H_0 when it's actually false - Probability = β - "Missing an effect that is real"

Statistical Power = 1 - β - Probability of correctly rejecting a false H_0 - Increases with: - Larger sample size - Larger effect size - Lower variability - Higher α level

The red area represents the power in the experiment

The farther apart the means the lower the beta error is… or you have higher power.



Lecture 6: Type I and Type II Errors - Visualization

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

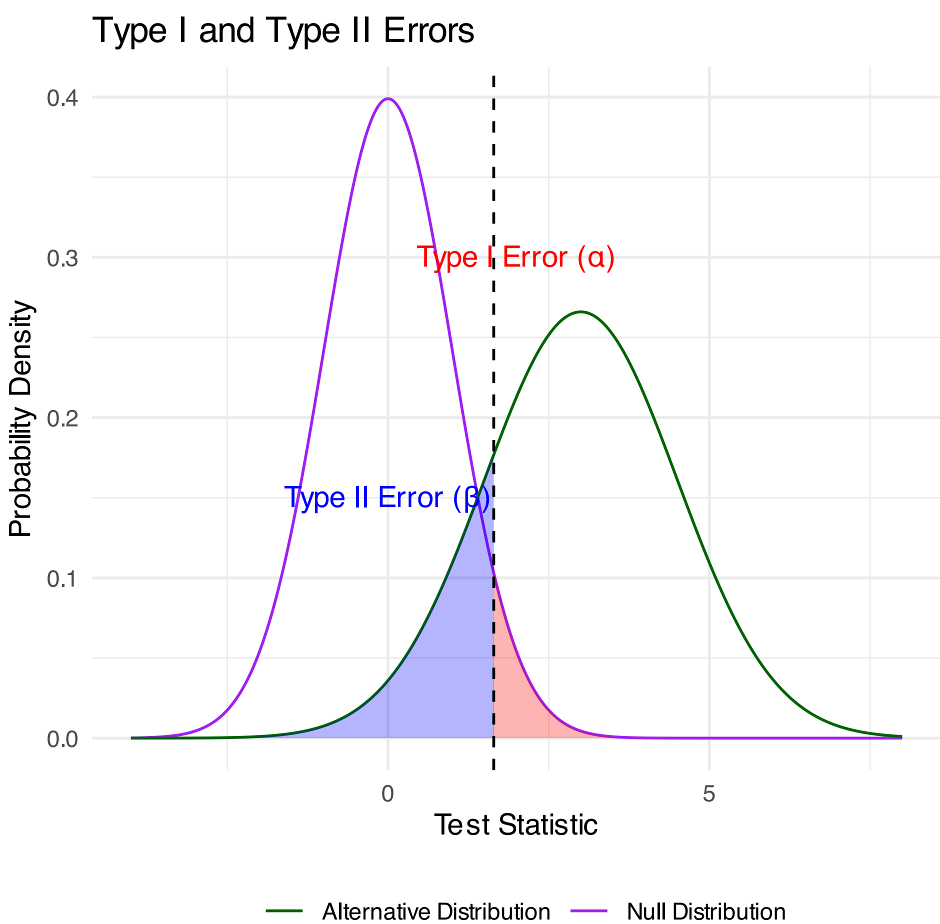
Type I Error (False Positive) - Rejecting H_0 when it's actually true - Probability = α (significance level) - "Finding an effect that isn't real"

Type II Error (False Negative) - Failing to reject H_0 when it's actually false - Probability = β - "Missing an effect that is real"

Statistical Power = 1 - β - Probability of correctly rejecting a false H_0 - Increases with: - Larger sample size - Larger effect size - Lower variability - Higher α level

The red area represents the power in the experiment

The farther apart the means the lower the beta error is… or you have higher power.



Practice Exercise: Interpreting Errors and Power

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:

- pooled standard deviation
  - This is the combined standard deviation of both groups weighted by respective degrees of freedom.
- Cohen's d
  - standardized difference between means - here assuming a difference of 30 units (mm)
  - `delta = 0.6741298`: The standardized effect size (Cohen's d)

```
library(car)
library(patchwork)
library(tidyverse)

grayling_df <- read_csv("data/gray_I3_I8.csv")
i3_df <- grayling_df %>% filter(lake=="I3")
i8_df <- grayling_df %>% filter(lake=="I8")
# Calculate power for detecting a 30 mm difference

n1 <- nrow(i3_df)
n2 <- nrow(i8_df)
sd_pooled <- sqrt((var(i3_df$length_mm) * (n1-1) +
                  var(i8_df$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
```

What is Power

Statistical power represents the probability of detecting a true effect (rejecting the null hypothesis when it is false). In this case, with a power of 97%, there's a 97% chance of detecting a true difference of 30 units between the means of the two groups if such a difference actually exists.

A power analysis like this is typically done for one of these purposes:

1. Before data collection to determine required sample size
2. After a study to evaluate if the sample size was adequate
3. To determine the minimum detectable effect size with the given sample

With 97% power, this test has excellent ability to detect the specified effect size. Generally, **80% power is considered acceptable**, so 97% indicates a very well-powered study for detecting a difference of 30mm between the groups.

Lecture 6: Error Bars and Their Interpretation

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

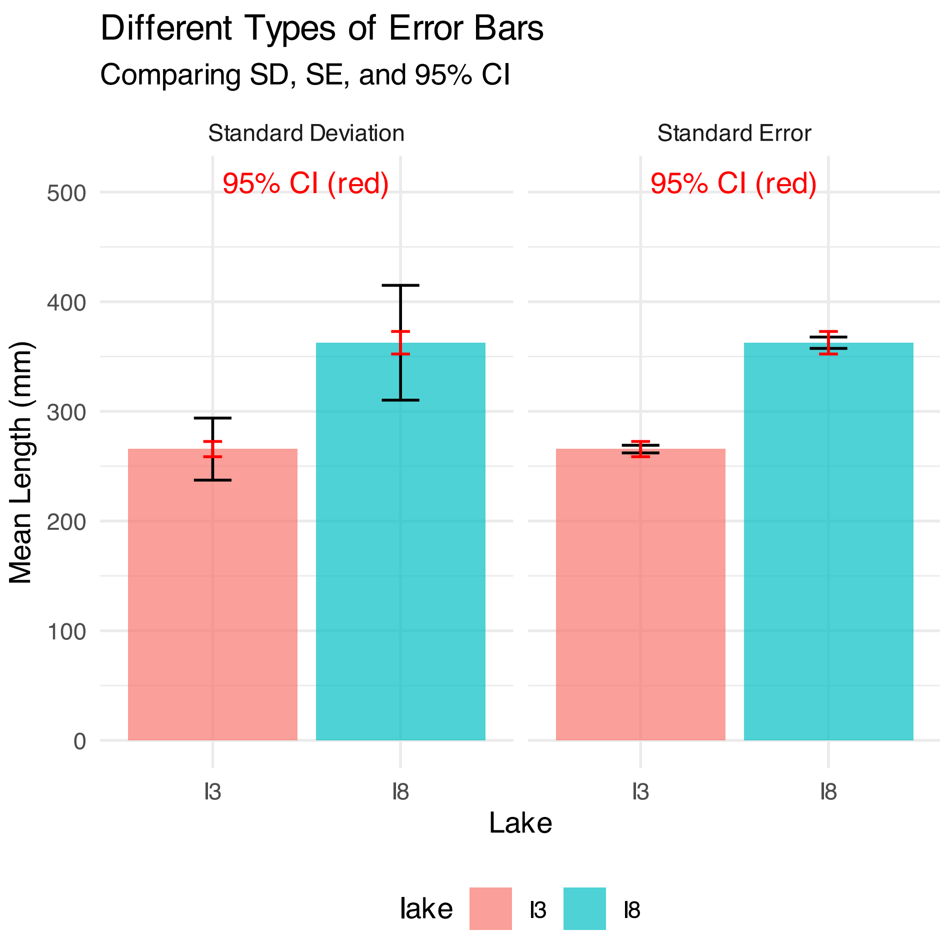
- The **precision** of a measurement
- The **uncertainty** around an estimate
- A **confidence interval** for a parameter

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 6: Sampling and Pseudoreplication

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

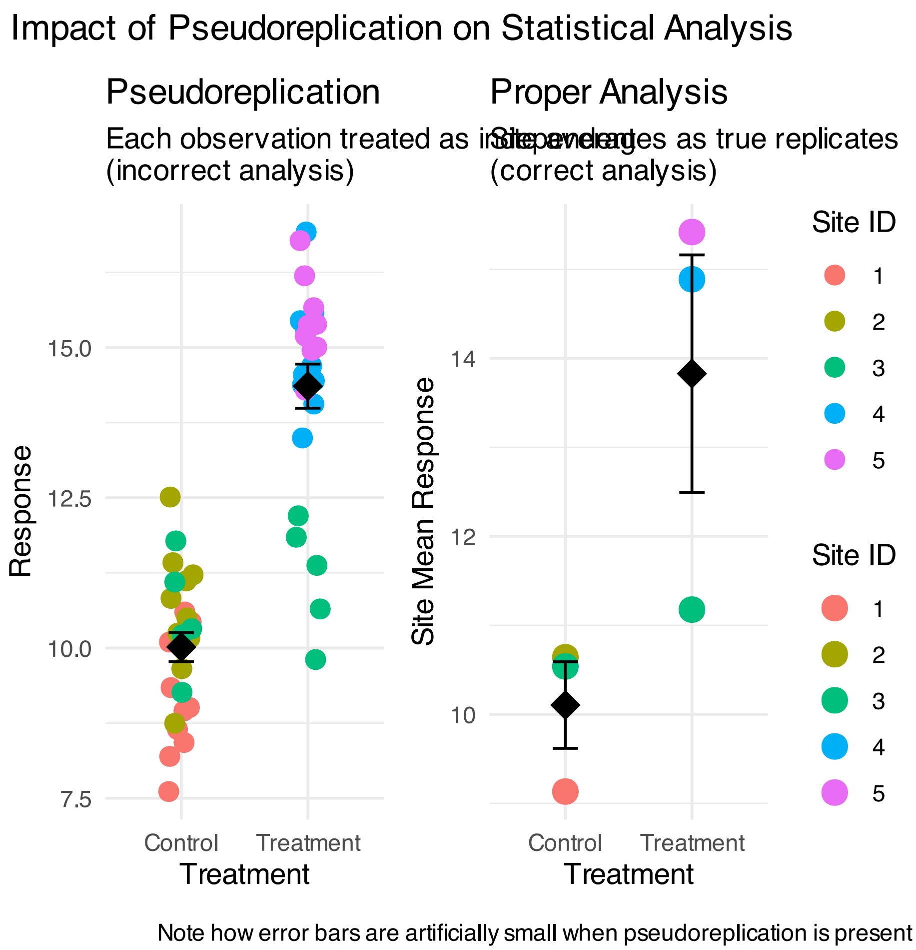
- A critical consideration in experimental design
- Results in underestimated standard errors and confidence intervals
- Leads to inflated Type I error rates (false positives)

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



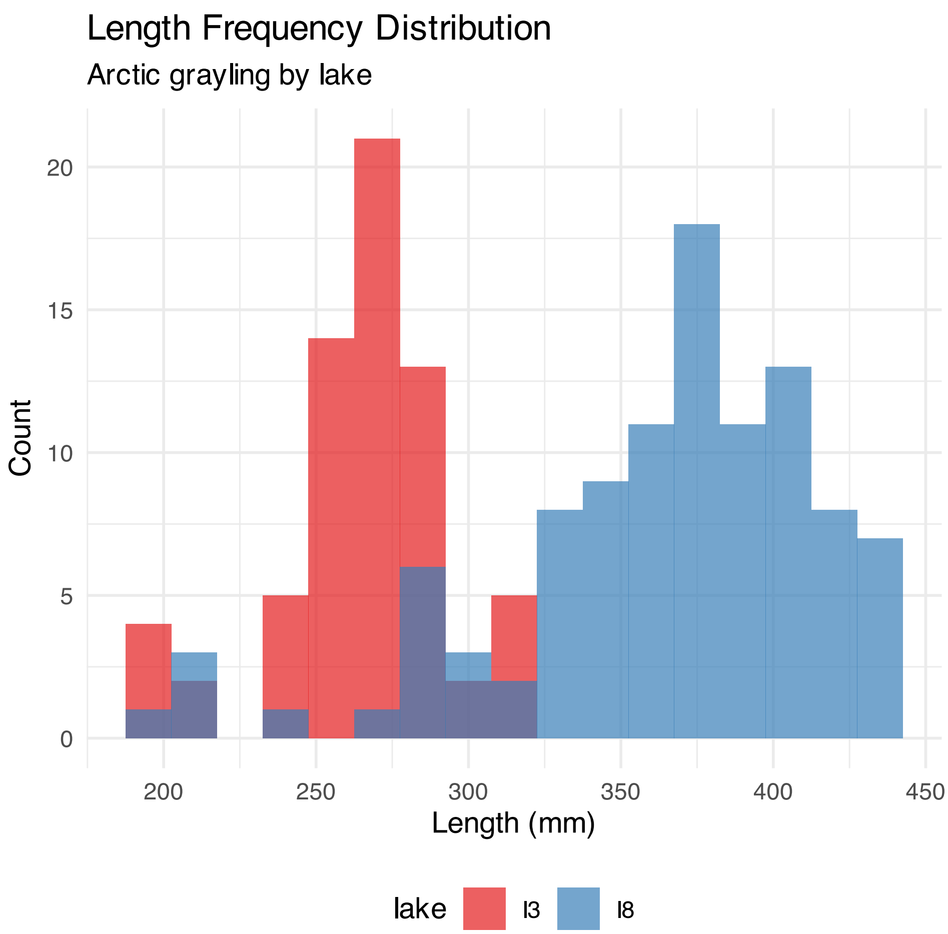
Lecture 6: Practical Applications in Fish Biology

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

- **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 6: Summary and Key Takeaways

Key concepts covered:

1. **P-values** measure evidence against the null hypothesis
   - Not the probability that H_0 is true
   - Should be interpreted in context with effect size
2. **Hypothesis testing** provides a framework for making decisions
   - Null and alternative hypotheses must be specified beforehand
   - α level determines Type I error rate
3. **Type I and Type II errors** represent different kinds of mistakes
   - Type I (α): False positive - rejecting true H_0
   - Type II (β): False negative - failing to reject false H_0
   - Statistical power = 1 - β
4. **Error bars** communicate uncertainty in different ways
   - Always check what type of error bar is shown
   - CI bars help assess statistical significance
5. **Pseudoreplication** inflates significance
   - Identify true experimental units
   - Account for non-independence in analysis