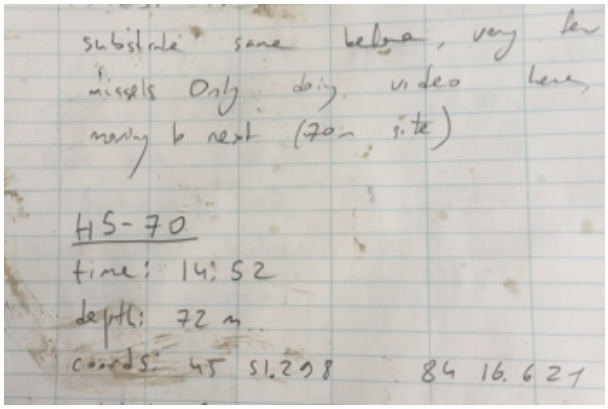
Lecture 08

Bill Perry

Lecture 7 Review

Covered

- What are the assumptions again and how do you assess them
- What to do when assumptions fail
  - Robust tests
  - Rank-based tests
  - Permutation tests



Lecture 8 Overview

Today we'll cover:

- Study design
- Causality in ecology
- Experimental design:
  - Replication, controls, randomization, independence
- Sampling in field studies
- Power analysis: *a priori* and *post hoc*
- Study design and analysis

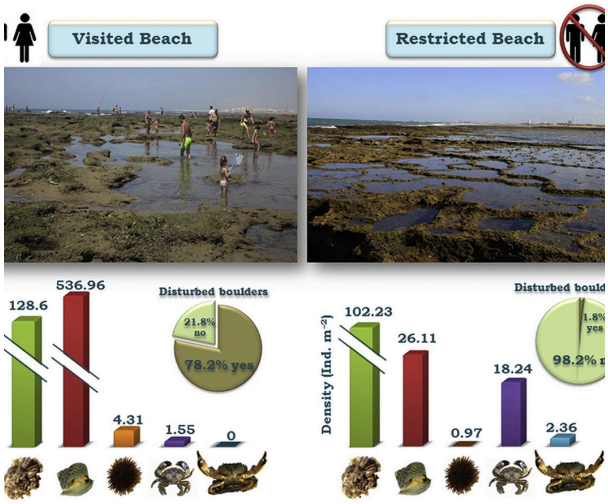
Lamberti and Resh 1983

Study Design Fundamentals

- Data analysis has close links to study design
- Statistics cannot save a poorly designed study!
- Key question: **what is your research question?**

Common scientific questions:

- Spatial/temporal patterns in variable Y?
- Effect of factor X on variable Y?
- Are values of variable Y consistent with hypothesis H?
- What is the best estimate of parameter θ?



https://ars.els-cdn.com/content/image/1-s2.0-S0272771416307958-fx1_lrg.jpg

Activity 1: Formulating Research Questions

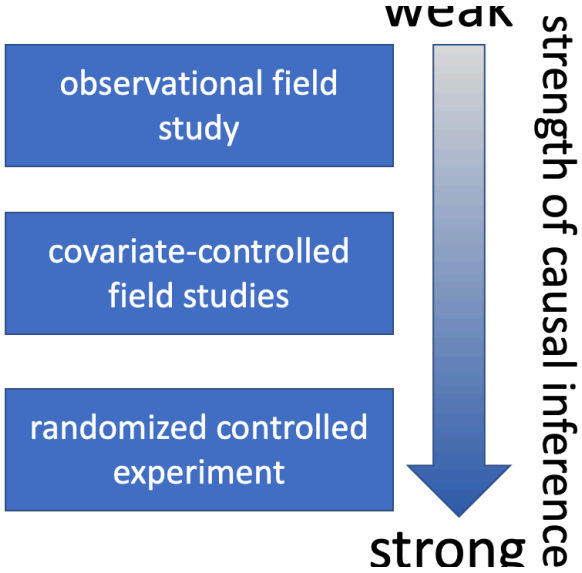
! Activity 1: Formulating Research Questions

- Take 5 minutes to write down 2-3 potential research questions about pine trees on our campus.
- Be as specific as possible about what you would measure.
- Share with a partner and discuss which questions would be easier to address experimentally.

Causality in Ecology - Introduction

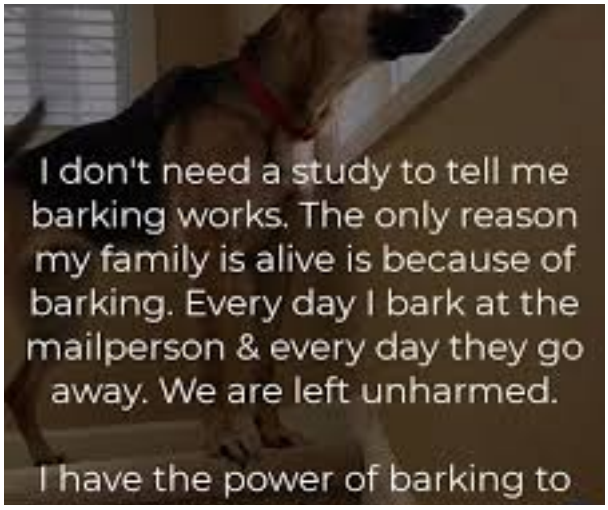
- Common question: what is the **cause** of Y?

- Causality is challenging; modern statistics lacks clear language for causality
- Strength of causal inference varies with study design
- Key factor: control of confounding variables



Causality in Ecology - Framework

- Common question: what is the **cause** of Y?
- Causality is challenging; modern statistics lacks clear language for causality
- Strength of causal inference varies with study design
- Key factor: control of confounding variables

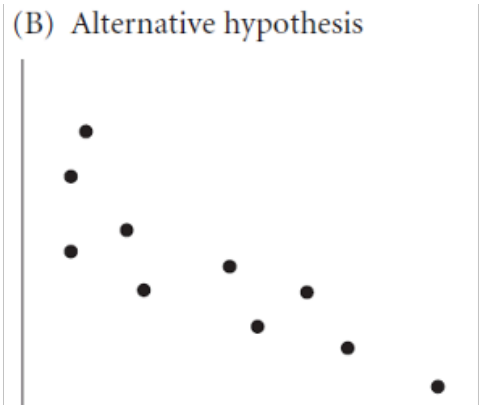


Causality Example

Example: Spider and lizard populations on small islands

Hypothesis: On small islands, lizard predation controls spider density

We're interested in causality. How do we get there?

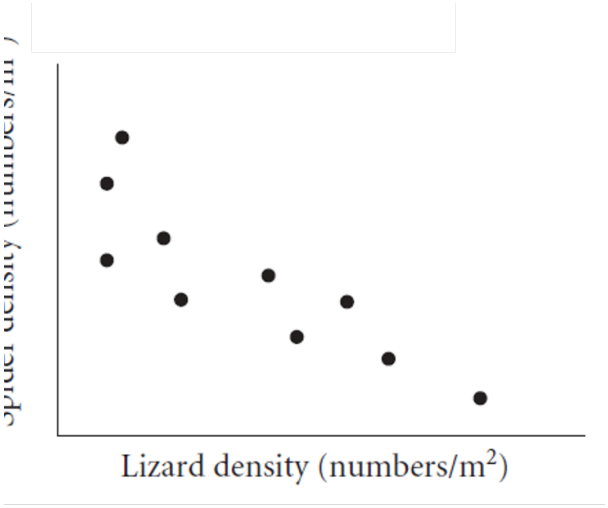


Natural Experiments

- Not really experiments at all!
- Utilizes natural variation in predictor variable
- E.g., survey plots across natural gradient of lizard density

Potential Problems:

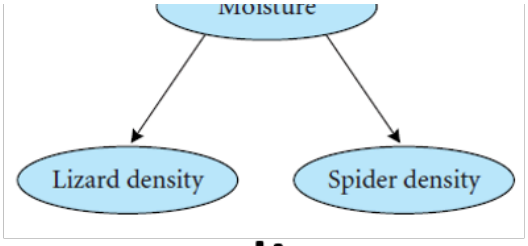
- Cannot determine direction of cause ↔ effect relationship
- Uncontrolled variables may affect results



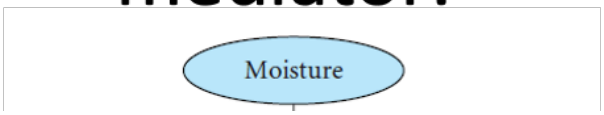
Strengthening Natural Experiments

Good design: Stronger inference from natural experiments

- Reduce confounding (select plots similar in relevant ways)
- Adjust for confounding (measure relevant covariates)
- Identify and measure potential confounding variables



mediator:



Activity 2: Pine Needle Natural Experiment Design

i Activity 2: Pine Needle Natural Experiment Design

Suppose we want to investigate whether wind exposure affects pine needle length across our campus. We decide to pick exposed sheltered locations. What sort of design would we conduct?

```
# Let's simulate some pine needle data from exposed and sheltered locations
exposed_locations <- data.frame(
  location = paste0("E", 1:5),
  wind = "exposed",
  length_mm = rnorm(5*10, mean = 75, sd = 10),
  tree_id = rep(1:5, each = 10)
)

sheltered_locations <- data.frame(
  location = paste0("S", 1:5),
  wind = "sheltered",
  length_mm = rnorm(5*10, mean = 90, sd = 12),
  tree_id = rep(6:10, each = 10)
)

fake_pine_data <- rbind(exposed_locations, sheltered_locations)

# View the data
fake_pine_data %>% ggplot(aes(wind, length_mm, color = wind)) + geom_boxplot()
```

In small groups, discuss:

1. What confounding variables might affect this natural experiment?
2. How would you design a study to reduce these confounding effects?
3. What data would you collect besides needle length?

Manipulative Experiments

Experimenter directly manipulates predictor variable and measures response

Randomized, controlled trials: gold standard

Challenges:

- Often restricted to small “plots”; scale-replication trade-off
- Often restricted to small, short-lived organisms
- Often limited to small number of treatments; treatment-replication trade-off

- Still requires **careful control of confounding!**

Cedar Creek Ecosystem Science Reserve

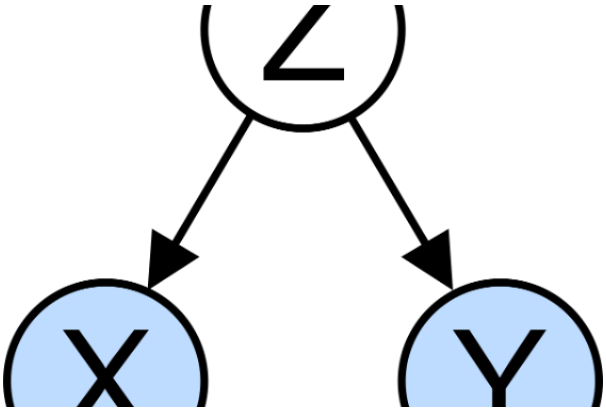
Experimental Design Principles

Main problem of study design & interpretation: **confounding**

- Is the result due to X or other factors?

Good study design seeks to eliminate confounding through:

- Replication
- Randomization
- Controls
- Independence



Replication

Replication is important because:

- Ecological systems are variable
- Need estimate of variability for many statistical methods

Without appropriate replication: Is the difference due to manipulation or something else?

Replication must be on the appropriate scale: match scale of replication to population of interest, otherwise run into **pseudoreplication** (Hurlbert)

Replication Examples

Example 1: Effects of forest fire on soil invertebrate diversity

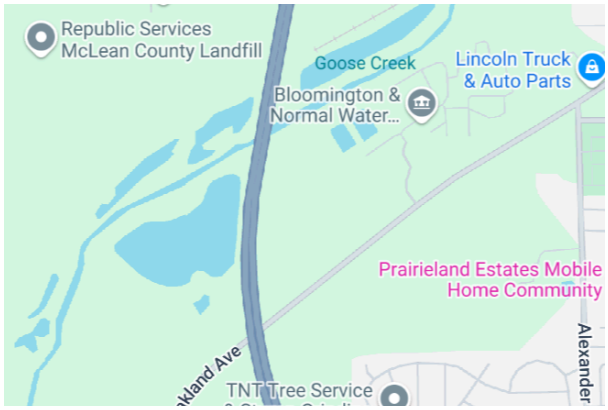
- Replicate samples from burnt and unburnt parts of *a single* forest
- What hypothesis is this design addressing?

Example 2: Effects of copper on barnacle settling

- 2 aquaria (+Cu, control), 5 settling plates in each
- Are settling plates replicates?

Example 3: Effects of sewage discharge on water quality

- 10 water samples above discharge, 10 below
- Are samples replicates?



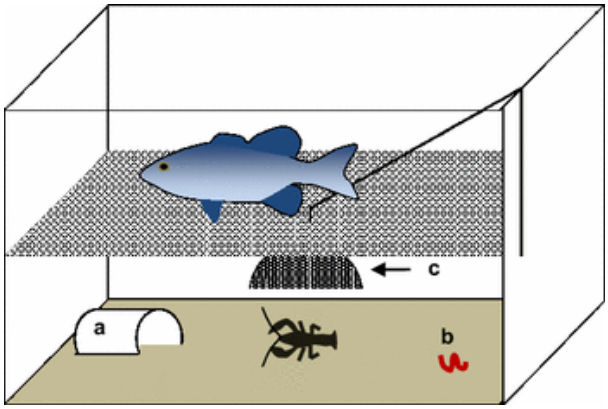
Consequences of Pseudoreplication

When you pseudoreplicate, you:

- Underestimate variability
- Increase type I error rate

Replicates must be on scale appropriate to population (& hypothesis!) of interest:

- Different burnt/unburnt forest areas
- Different aquaria
- Different plants and streams



Activity 3: Identifying True Replication

Activity 3: Identifying True Replication

For each scenario below, identify whether there is true replication or pseudoreplication:

1. Testing soil pH effects on pine seedling growth by using one large pot with acidic soil and one with basic soil, with 10 seedlings in each pot
2. Testing the effect of a fertilizer on pine growth by applying it to 5 randomly selected trees in a stand, with 5 other trees as controls
3. Measuring air pollution effects by sampling needle damage in 3 pine stands near a factory and 3 stands 50km away

Discuss how you would redesign any pseudoreplicated studies.

When Replication is Difficult

What if replication is impossible/difficult/expensive?

Example: Effect of temperature on phytoplankton growth

- 4 chambers (5, 10, 15, 20℃), 10 beakers in each
- Are beakers true replicates?

Possible solutions:

- Rerun the experiment a few times, changing temperature of chambers
- Try to account for all possible differences between chambers (light levels, humidity, contamination)

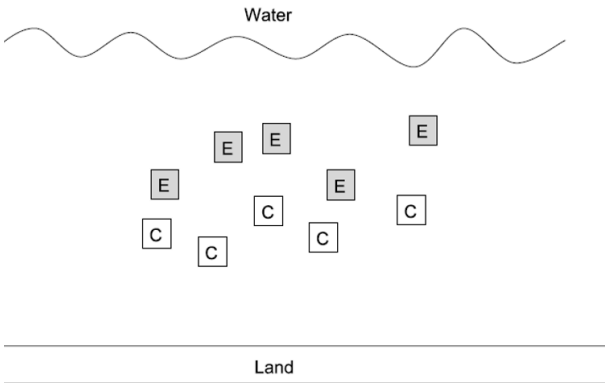
Randomization

Randomization helps deconfound “lurking” variables:

- Attempts to equalize effects of confounders

Random sampling from population:

- Experimental units should represent random sample from population of interest
- Ensures unbiased population estimates and inference
- E.g., animals in experiment are random subset of all animals that could have been used



Randomization in Practice

Allocation of experimental units to treatment/control:

- Experimental units must have equal chance of being allocated to control or experimental group
- Properly done by random number generation

Randomization is essential at two levels:

- Random selection from population
- Random assignment to treatments

```
# Example of randomization in R
# Select 10 trees randomly from 100 possible trees
all_trees <- 1:100
selected_trees <- sample(all_trees, 10)

# Randomly assign 5 trees to treatment and 5 to control
treatment_trees <- sample(selected_trees, 5)
control_trees <- selected_trees[!selected_trees %in% treatment_trees]

# Display results
data.frame(
  Treatment = treatment_trees,
  Control = control_trees
)
```

```
  Treatment Control
1        18      49
2        74     100
3        65      47
4        24      71
5        25      89
```

Controls

Key question: Is response due to manipulation/hypothesized mechanism or external factor?

Controls help address this question:

- Experimental units treated exactly as the manipulated units, except no manipulation under investigation
- Can be tricky to implement; requires careful thought

Examples:

- In toxicology, controls and treatment groups must both be injected, but control does not receive the substance under study
- Predator exclosures often produce “cage effects”
- need two controls: a grazer/predator control and a “cage control”

Activity 4: Designing Controls

⚠ Activity 4: Designing Controls for Pine Experiments

Work in small groups to design appropriate controls for each experiment:

1. Testing whether pine needle length is affected by a particular fertilizer
2. Testing whether pine needle density affects water retention during drought using enclosed branches
3. Testing whether sunlight exposure affects pine seedling growth using shade cloth

For each experiment, identify:

- What would be appropriate controls?
- What factors need to be controlled besides the main variable?
- Could there be “cage effects” or similar issues to consider?

Independence

Independence of observations: assumption of many statistical methods

Events are independent if occurrence of one has no effect on occurrence of another

- E.g., offspring of one mother for treatment, offspring of another for control

Temporal/spatial autocorrelation: violation of independence

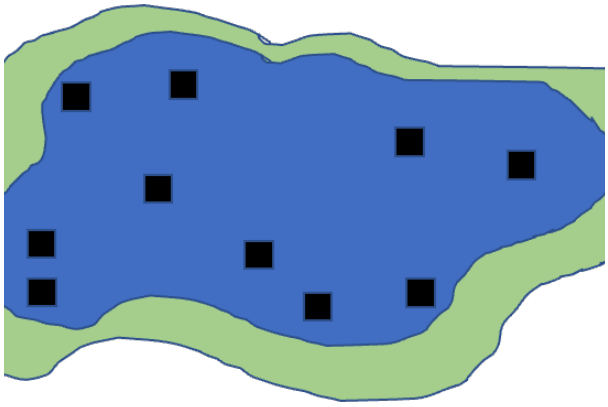
- Values of variables at certain place/time correlated with values at another place/time
- “Everything is related to everything, but near things are more related than distant things”
- Special methods to adjust for autocorrelation

CONTROL

Sampling Design in Field Studies - Simple Random

Simple random design:

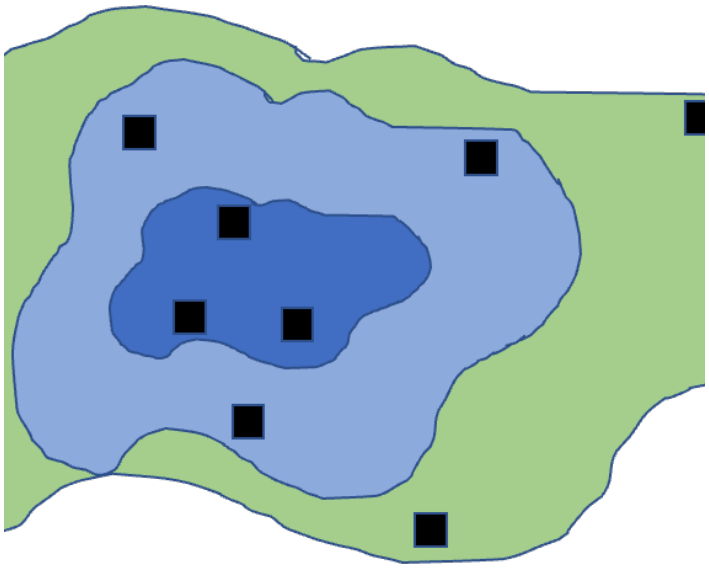
- all individuals/sampling units have equal chance of being selected
- Assign number to all possible units, select units using random number generator
- Often tricky in ecology; haphazard is common alternative
- Most population estimates and tests assume random sampling



Sampling Design - Stratified

Stratified designs: if there are distinct strata (groups) in population, may want to sample each independently

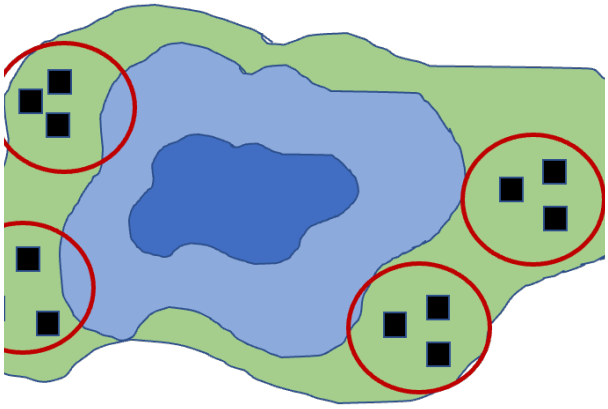
- Samples collected from each stratum randomly, n proportional to “size” of stratum
- Means and variances need to be estimated using different procedure; strata included in model



Sampling Design - Cluster

Cluster designs:

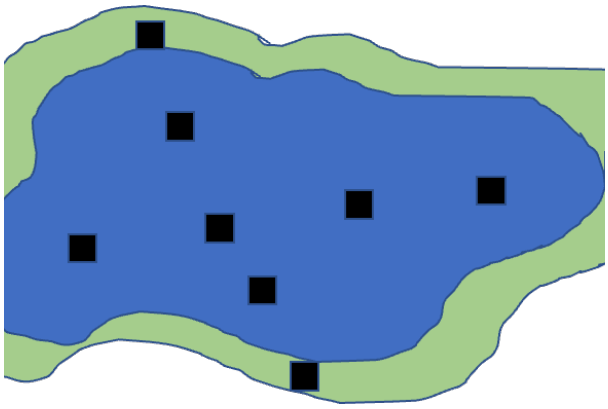
- focuses on sampling subunits nested in larger units
- Used when other designs impractical (e.g., due to cost)
- Mean calculation easy, modified procedure for variance
- Nested ANOVA is often appropriate analytical method



Sampling Design - Systematic

Systematic designs:

- sampling units evenly dispersed: “transect” sampling common in ecology
- Used to determine changes along gradient
- Risk: might coincide with some natural pattern



Activity 5: Field Sampling Pine Trees

Activity 5: Field Sampling Pine Trees

Let's consider sampling pine needles across campus:

```
# Let's create a campus map grid (simplified)
campus_grid <- expand.grid(x = 1:10, y = 1:10)

# Place "pine trees" clustered toward the north side (higher y values)
set.seed(46)
pine_locations <- data.frame(
  x = sample(1:10, 30, replace = TRUE),
  # Using rbeta to skew distribution toward higher y values
  # Alpha=1, Beta=3 creates right-skewed distribution, then scale to 1-10 range
  y = round(rbeta(30, 3, 1) * 9 + 1)
)

# Plot the campus and trees
ggplot() +
  geom_point(data = campus_grid, aes(x, y), color = "lightgrey", size = 0.5) +
  geom_point(data = pine_locations, aes(x, y), color = "darkgreen", size = 3) +
  theme_minimal() +
  labs(title = "Pine Tree Locations on Campus Grid (North Clustered)")
```

In groups of 3-4, design a sampling strategy to:

1. Estimate average needle length across campus (simple random sampling)
2. Compare needle lengths between north and south campus areas (stratified sampling)
3. Study how needle length changes with distance from the main road (systematic sampling)

For each strategy, describe:

- How many samples you would take
- Where you would take them
- What additional variables you might measure

Power Analysis Introduction

Power is an important aspect of experimental design:

- Low power $\longrightarrow$ higher likelihood of type II error (1-β)
- A study's power tells us how likely we are to see an effect if one really exists

Can use power analysis:

- Before experiment (*a priori*): how many samples do we need?
  - what effect size can we detect?
- After experiment (*post hoc*): was finding of no effect due to lack of effect or poor design?

Power is a function of:

- ES - Effect size
- n - Sample size
- sigma - standard deviation
- α (significance level) - 0.05

$$\text{Power} \propto \frac{ES\alpha\sqrt{n}}{\sigma}$$

A Priori Power Analysis

Using power analysis to plan experiments:

Sample size calculation: how many samples will be needed?

Need to know: desired power, variability, significance level, effect size

Effect size calculation: what kind of effect can we find, given particular design?

Need to know: desired power, variability, significance level, n

Cohen's d - standardized measure of effect size used in statistical analysis, particularly when comparing two means

- 0.2 = small effect
- 0.5 = medium effect
- 0.8 = large effect

Helps determine the practical significance of research findings, as opposed to just statistical significance (p-values). A Cohen's d of 0.8 means that the difference between groups is large enough to be substantial in practical terms - specifically, it indicates that the means differ by 0.8 standard deviations.

A Priori Power Analysis Example

```r
# A priori power analysis for t-test
# How many samples needed per group?

# Parameters
effect_size <- 0.8  # Cohen's d
significance <- 0.05
desired_power <- 0.8

# Calculate sample size needed
pwr.t.test(d = effect_size,
           sig.level = significance,
           power = desired_power,
           type = "two.sample")
```

```
     Two-sample t test power calculation

              n = 25.52458
```

```
              d = 0.8
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Post Hoc Power Analysis

Imagine you did not reject null hypothesis - still worth publishing result?

Is non-significant result due to low power (poor design) or actual no-effect situation?

- Have n and estimate of σ
- Need to define effect size that wanted to detect
- In return get estimate of experiment's power

Cohen's d is calculated as: d = (Mean1 - Mean2) / SD_pooled Where SD_pooled is the pooled standard deviation of both groups.

Can help convince reviewers that you are a good experimenter, but there really is no effect... please publish my non-significant finding!

Post Hoc Power Analysis Example

```r
# Post hoc power analysis
# If we had n = 20 per group

# Parameters
effect_size <- 0.5  # Medium effect size
significance <- 0.05
sample_size <- 20  # per group

# Calculate achieved power
pwr.t.test(n = sample_size,
           d = effect_size,
           sig.level = 0.05,
           type = "two.sample")
```

```
     Two-sample t test power calculation

              n = 20
              d = 0.5
      sig.level = 0.05
          power = 0.337939
    alternative = two.sided

NOTE: n is number in *each* group
```

Activity 6: Power Analysis for Pine Needle Experiment

! Activity 6: Power Analysis for Pine Needle Experiment

Let's design a study to compare needle lengths between exposed and sheltered pine trees:

```
# Based on pilot data, we have these estimates:
exposed_mean <- 75    # mm
sheltered_mean <- 85  # mm
pooled_sd <- 12       # mm

# Calculate Cohen's d effect size
effect_size <- abs(exposed_mean - sheltered_mean) / pooled_sd
effect_size

# A priori power analysis
pwr.t.test(d = effect_size,
           sig.level = 0.05,
           power = 0.8,
           type = "two.sample")
```

Activity 6: Power Curve Visualization

! Activity 6: Power Analysis for Pine Needle Experiment

Let's design a study to compare needle lengths between exposed and sheltered pine trees:

```
# Visualize the power curve
sample_sizes <- seq(5, 30, by = 1)
power_values <- sapply(sample_sizes, function(n) {
  power <- pwr.t.test(n = n,
                      d = effect_size,
                      sig.level = 0.05,
                      type = "two.sample")$power
  return(power)
})

power_df <- data.frame(
  sample_size = sample_sizes,
  power = power_values
)

ggplot(power_df, aes(x = sample_size, y = power)) +
  geom_line(color = "blue", size = 1) +
  geom_hline(yintercept = 0.8, linetype = "dashed", color = "red") +
  theme_minimal() +
  labs(title = "Power Analysis for Pine Needle Study",
       x = "Sample Size (per group)",
       y = "Statistical Power")
```

Questions:

1. How many trees should we sample to achieve 80% power?
2. If we can only sample 5 trees per group, what is our power?
3. How would increasing variability (SD) affect our sample size requirements?

Interactive Power Analysis

> **Note**
>
> Power vs. Effect Size Interactive Demonstration
>
> Try adjusting these parameters to see how they affect required sample size:

```
# Adjust these parameters
mean_difference <- 10   # Difference between groups (mm)
std_deviation <- 12     # Standard deviation (mm)
target_power <- 0.8     # Desired statistical power

# Calculate effect size
effect_size <- mean_difference / std_deviation

# Calculate required sample size
power_result <- pwr.t.test(d = effect_size,
                           sig.level = 0.05,
                           power = target_power,
                           type = "two.sample")

# Display results
cat("Effect size (Cohen's d):", round(effect_size, 2), "\n")
```

```
Effect size (Cohen's d): 0.83
```

```
cat("Required sample size per group:", ceiling(power_result$n), "trees\n")
```

```
Required sample size per group: 24 trees
```

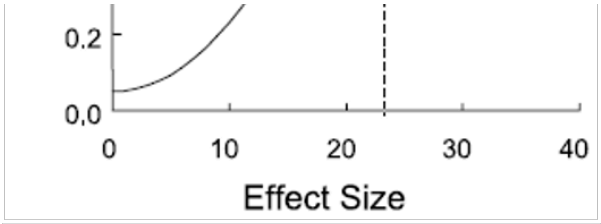
Study Design and Analysis

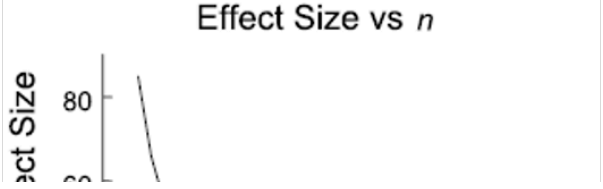
Study design is closely linked to statistical analysis

Recall: - Categorical vs. continuous variables - Dependent vs. independent variables

Nature of variables dictates analytical approach:

- Match your analysis to your design
- Cannot “fix” poor design with fancy statistics





Summary and Take-Home Messages

Key concepts we covered today:

1. **Study design is critical** - statistics cannot save poor design
2. **Natural vs. manipulative experiments** - different approaches to causality
3. **Principles of good design:**
   - Replication at the right scale
   - Proper randomization
   - Appropriate controls
   - Independence
4. **Power analysis** - planning for sufficient sample size
5. **Match analysis to design** - your statistical approach should follow from your experimental design

Remember:

- Correlation ≠ causation
- Beware of pseudoreplication
- Design before you collect data
- Consider practical constraints
- Report everything transparently

References and Additional Resources

- Gotelli, N. J., & Ellison, A. M. (2012). A primer of ecological statistics (2nd ed.). Sinauer Associates.
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. Ecological Monographs, 54(2), 187-211.
- Quinn, G. P., & Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge University Press.
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution, 1(1), 3-14.