MARS 4910: Jan 23

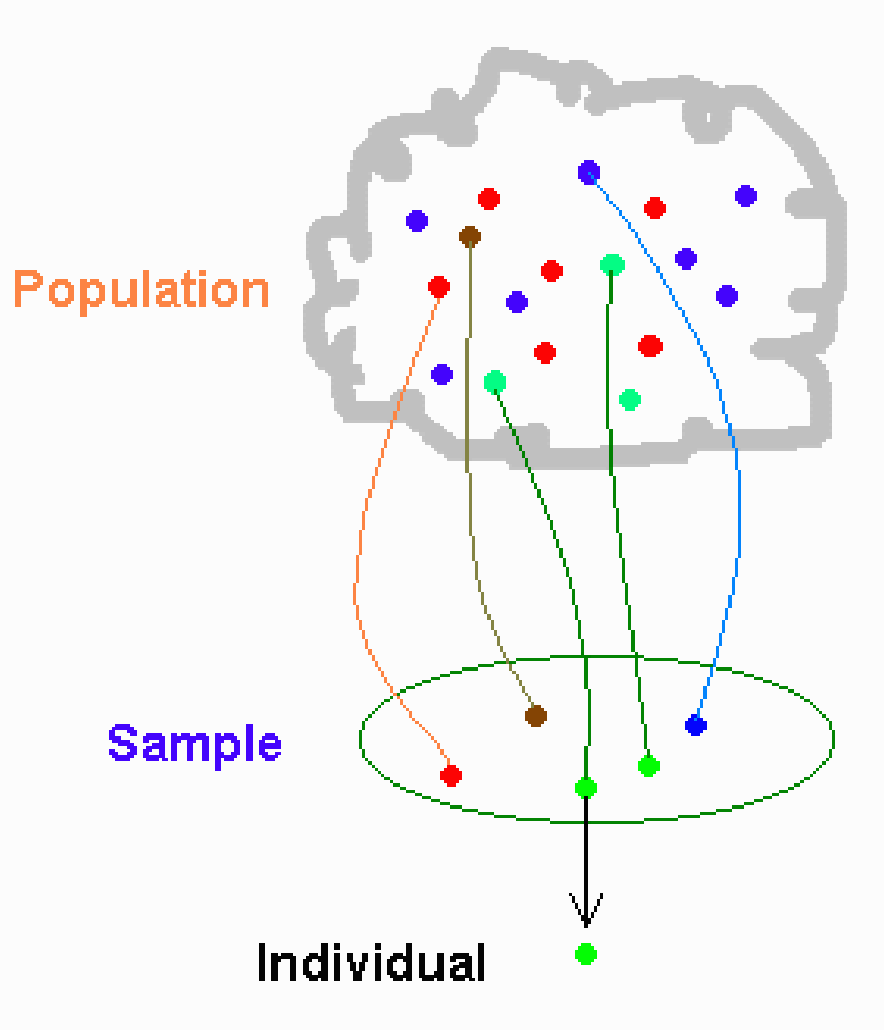
Plan for Today:

- Review Hypotheses and Research Plans
- Discuss experimental design:
  variables & samples
- Introduction to R and RStudio

Assignments for Today:

- Develop Research Plan: Hypotheses
  Field sites
  Sampling methods
- Put together equipment list (for 01/28 & 02/06)

Statistical Review I: Variables and Samples



Hypothesis Testing

Working Hypothesis:

Provides a verbal description of the patterns.

Lays out conceptual relationship(s) between the variables.

Describes sign (and strength) of these relationships.

Statistical Hypotheses:

Null Hypothesis

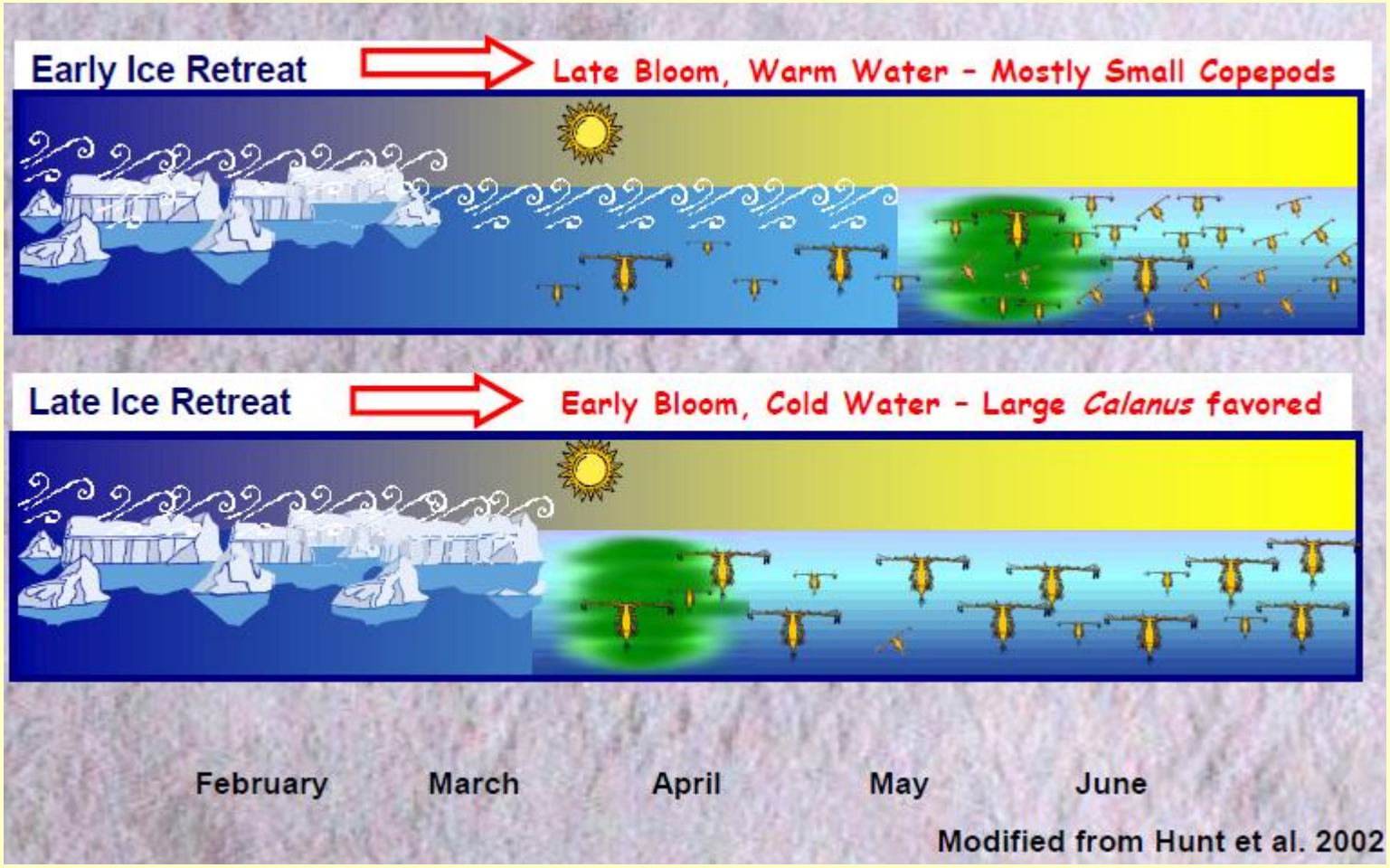
There is no relationship between the variables

Alternative Hypothesis

There is a relationship between the variables

Example: Bering Sea Ice

Timing of sea ice retreat influences timing of the spring bloom and structure of zooplankton grazers



Identify your Working Hypothesis

- Write down overarching conceptual idea(s) you will test
- Organize these ideas from more general to more specific

  (or logically, so answering the first question informs the next question you ask?)

- For each working hypothesis, identify the statistical hypotheses (null hypothesis / alternative hypothesis).
- Note: these hypotheses relate to the relationships between specific variables

Variables and Hypothesis Testing

Testing hypotheses requires making predictions and taking measurements of different variables

Often, the goal is to measure the response of one variable to an experimental change in other variables

Independent	Dependent
Variable denotes the cause (the driver of the pattern)	Variable denotes the effect (responds to the driver)
Termed: predictor variable	Termed: outcome variable

Classes of Variables

Numerical Variables:

Variable takes on numerical values (e.g., ant nests, length)

Can be ordered and ranked

Subclasses:
-Discrete: few possible values (integers)
- Continuous: Measurements take any value within range

Categorical Variables:

Variable takes on different entities or categories (e.g., color)

Can be ordered and ranked (ordinal variables)

Subclasses:
-Binary: Two Options
-Nominal: > 2 Options

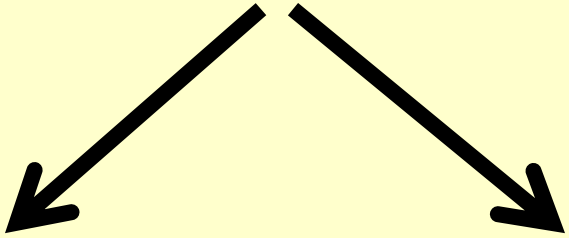
Analysis Overview

- Identify Variables:

  Driver (Independent) OR Response (Dependent)

  Categorical OR Continuous

- Determine Analysis Framework: Test Data Distributions



Parametric statistics
(require normal distributions)
(test other assumptions)

Nonparametric statistics
(do not make assumptions)
(compare ranked data)

Characterizing your Variables

NOTE: Consider one response variable at a time

Hypothesis Name (Order) and Verbal Description	# Continuous Predictors	# Categorical Predictors	Paired OR Unpaired Data *	Do Data Meet Parametric Assumptions ?

* Paired / Unpaired Categorical Predictors:
Are the same objects (individuals / samples) used in each category ? (YES / NO / sometimes)

Paired or Independent Data

Paired Data: Two datasets are "paired" when a one-to-one relationship exists between the values in the two datasets, such that:

- Each data set has the same number of data points.
- Each data point in one data set is related to one, and only one, datapoint in the other data set.

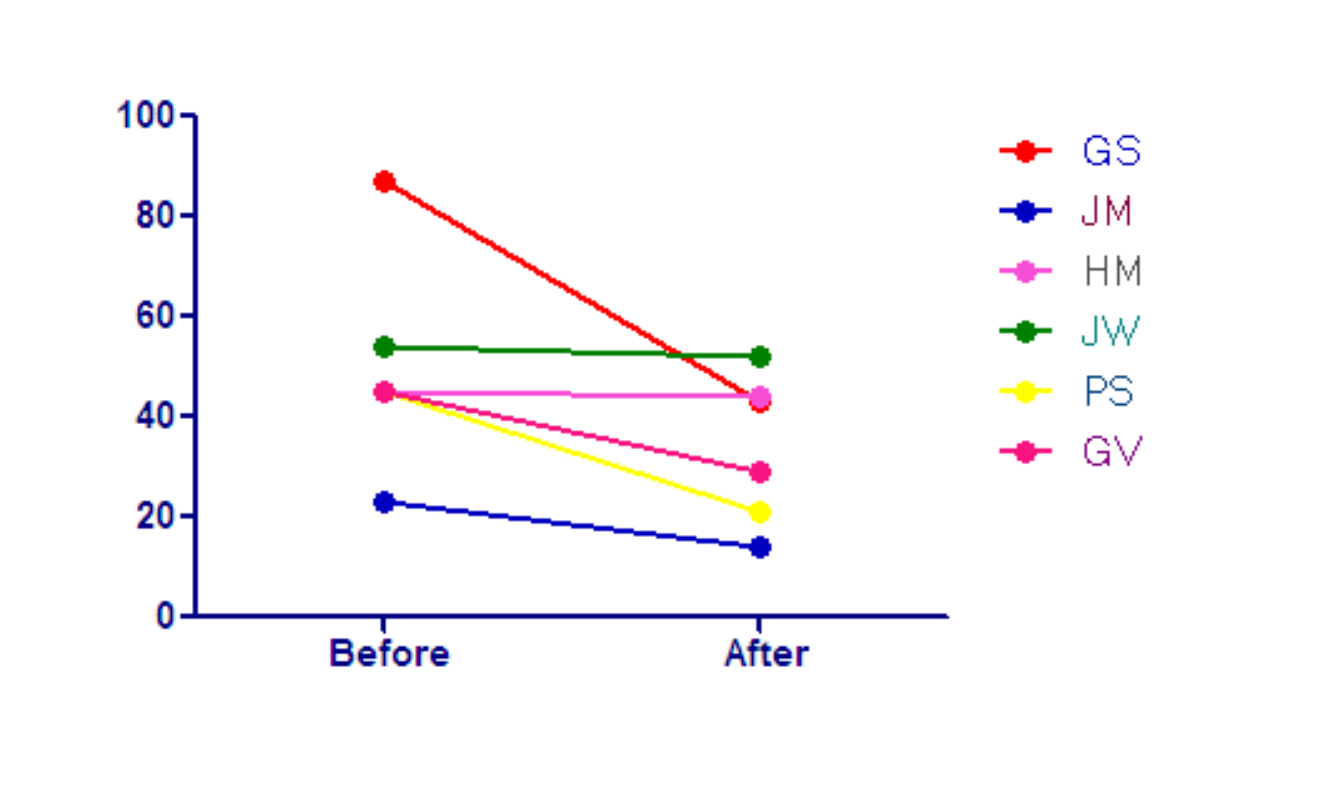
Pragmatically:

Unpaired or independent data when the two sets of observations arise from separate individuals.

Paired data arise from the same individual under different circumstances (e.g., at different points in time).

Paired Data

Examples: Did you know more statistics after listening to this lecture?



Sampling

Question: Estimating organism distributions

- Presence / Absence vs Abundance

- Relative Abundance vs Total Abundance

Examples:

- Corals, Algae
- Fish, Urchins, Turtles
- Scavengers

Standardized Samples

Question: How to ensure sample comparability ?

- Presence / Absence vs Abundance

- Relative Abundance vs Total Abundance

Examples:

- Corals, Algae
- Fish, Urchins, Turtles
- Scavengers

Standardized Samples

Tasked with comparing nest density of ground-foraging ant species in two adjacent habitats: agricultural field / forest.

Delineate Two Study Areas

Sample Randomly (lat / long)

Standardize Survey Effort (1 m x 1 m)

Random Sampling

A random sample is a subset of individuals (a sample) chosen from a larger set (a population) such that:

Each individual is randomly chosen (by chance):

- each individual has an equal
- and independent probability

of being chosen during the sampling process,

A random sample is an unbiased surveying technique.

Controls

Question: How do we avoid the influence of "outside" factors and variability ?

Table 2.1. Sources of variability in experiments, and procedures for reducing confusion caused by such variability

Source of Variability	Reduction by Experimental Design
1. Variability among experimental units	Replication, interspersion, and simultaneous measurement
2. Random error in measurement of response variables	Replication
3. Change in conditions through time	Controls
4. Unsuspected side effects of treatment procedures	Controls
5. Bias of investigator	Randomized assignment of treatments to experimental units
6. Chance influences on experiment in progress	Replication and interspersion

Controls

- In manipulative experiments: controls are experimental units that are identical to those receiving manipulative treatments except in the critical treatment factor.

- For example, if fertilizer is applied to a plot of grassland by spraying a solution of fertilizer in water, control plots should …

Controls

- In field research: controls can reveal whether some change is occurring in the plots over time, because of factors the experimenter cannot hold constant, such as seasonal changes in day length / rain.

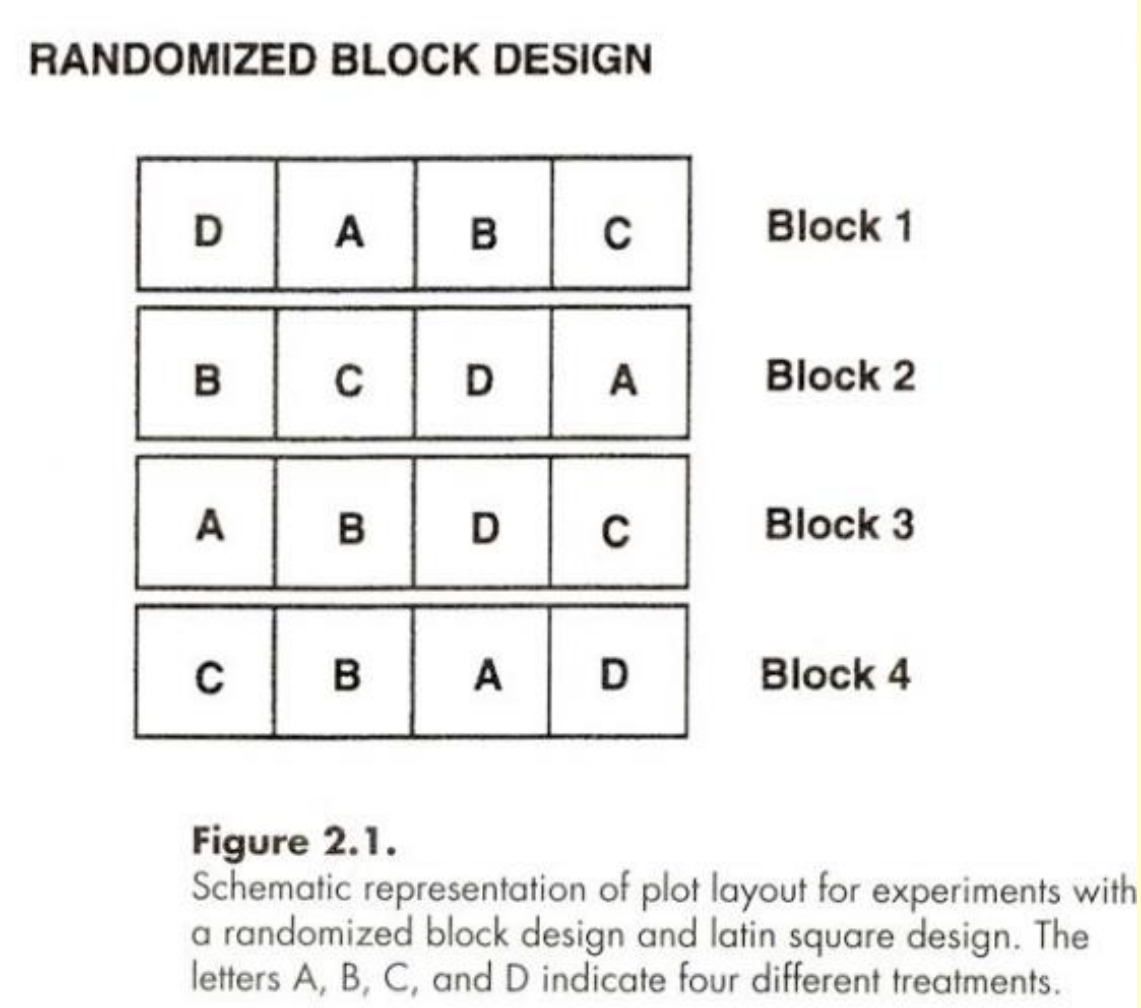


Figure 2.1.
Schematic representation of plot layout for experiments with a randomized block design and latin square design. The letters A, B, C, and D indicate four different treatments.

Controls

- Controls essential for field experiments, because:
  - It can rarely be assumed that conditions will remain constant for any substantial time
  - Almost any measurement or manipulation involves incidental impacts of the investigator.

Controls

Question: What "outside" factors and sources of variability may influence your measurements / samples ?

Question: What steps can be taken to remove those "outside" factors / sources ?

Characterizing your Variables

NOTE: Consider one response variable at a time

Hypothesis Number and Verbal Description	# Continuous Predictors	# Categorical Predictors	Paired OR Unpaired Data ? *	Do Data Meet Parametric Assumptions

* Paired / Unpaired Categorical Predictors:
Are the same objects (individuals / samples) used in each category ? (YES / NO)

Summarizing your Variables

Defining Variables

Influence how we measure / quantify observations

Influence what types of tests can be performed

Characterizing Variables

Use frequency tables

Use graphs (histograms / boxplots)

Characterize distributions with a variety of metrics:

- range, mode
- central tendency (mean, median)
- variability (variance, interquartile range)

MARS 4910: Jan 28 / 30

Plan for Today:

Presentation of research projects by each group:

Study sites, Methods
Number of sampling days requested
Laboratory work / equipment requirements

Cruise Planning Discussion and Class Feed-back

Assignments for Next Week:

- Complete Research Plan

- Work on Research Proposal (Due Feb 11 / 13)