January 15th, 2025

Topic: What is statistics and how is it done in Biology?

Opening discussion

1. How is statistics used in different fields of biology and science/society?

2. How have you used statistics before?

3. What courses have emphasized statistics?

4. What disciplines use statistics frequently? What disciplines rarely use statistics?

5. Why do we do statistics in the first place?

The fundamental use of statistics is to quantify properties of samples and to determine whether those properties truthfully represent properties of populations.

A population is the entire set of objects relevant to a scientific study. We call the objects of study individuals.

Samples are the specific individuals under study.

Individuals are the biological units of study relevant to a stated biological question.

> Example: I study infectious diseases and how they spread. One question I have is what fraction of the US population has been infected with avian influenza.

The population for this question is the entire US population.

A sample is 500 students at WFU.
Another is 500 people from North Carolina.
Other samples:

- 500 individuals from California
- 500 farm workers in California

Given that avian influenza is much more common on farms, and in particular CA farms, it is likely that we end up with very different answers based on our sampling design. A first step in statistics is to evaluate how our sampling design leads to samples that reflect the true properties of the population under question.

Some samples are clearly better than others for inferring information about the properties of populations.

Good samples are those where individuals in the sample have properties that are randomly selected in reference to the biological question.

> Such samples are called **random samples**. Nearly all statistical methods make an assumption that the data come from a random sample.

For avian flu, exposure to the virus is important for determining whether individuals contract the virus. Exposure is not selected randomly in our sample. In the WFU and NC cases, exposure is lower than the population. In the CA farm worker example, exposure is higher than the population level.

Note:
Individuals (the statistical units under study) need not be *biological individuals*.

They can be

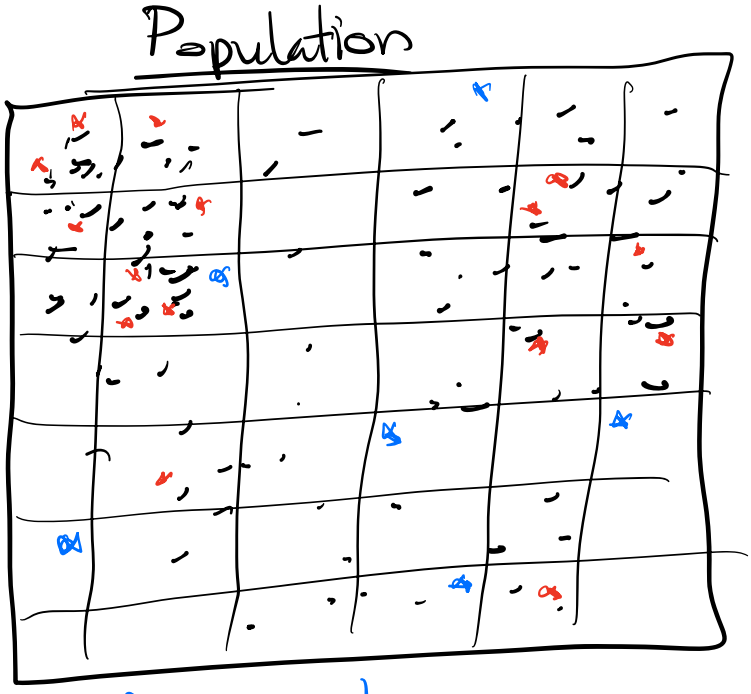
times	colonies	communities
locations	organs	organelles
individuals (in the biological sense)		
groups	societies	cultures
genes	genomes	proteomes
cells	molecules	tissues

or some combination!

Example

I do some spatial ecology and a big question is about how dense individuals are in space.

Consider trees in a grid.



Q1. How many trees per unit area?

Q2. How many trees neighbor each tree in the same sized area?

Individuals are different

For Q1, the individuals are locations because the question is about properties of space.

For Q2, the individuals are trees because the question is about properties of trees.

Sampling the wrong individuals can lead to non-random sampling of the correct individuals for the question.

Analogy to class size.

Average class size → individuals are classes

Average classmates → individuals are students.

Properties of populations

Properties are anything you can measure about individuals that are aggregated at a population level.

Measures of Average Tendency

Means — Proportions

Medians

Quantiles

Measures of Variability

Variance

Standard Deviation (= square root of variance)

Skew

Kurtosis

Differences

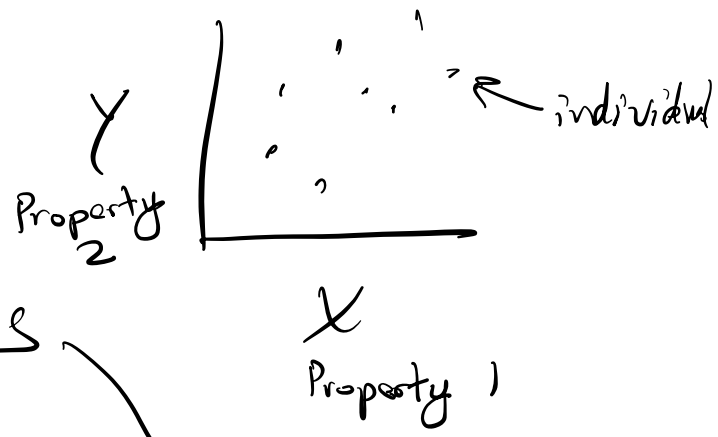
Measures of Association

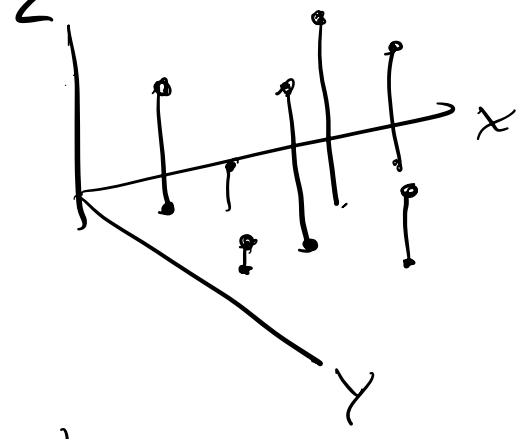
Correlations

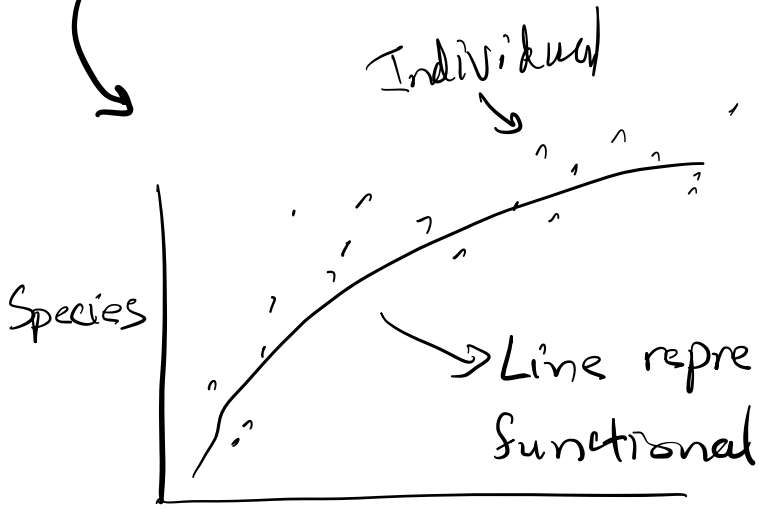
Covariances

Multivariate dissimilarities

Functional Relationships







Area

These properties are of interest for 3 main objectives in statistics.

1. Estimation

Goal: Describe the world by inferring population properties from sample properties

2. Hypothesis Testing

Goal: Evaluate the evidence consistent with a hypothesis posed in terms of population parameters.

3. Prediction

Goal: Use functional relationships from samples to predict properties of individuals.

Examples

1. Prevalence of Avian Flu
2. Gene-knock out studies
3. Species-Area Relationship

Fundamental Objective

1. Descriptive
2. Understanding causal nature of the world
3. Predicting how the world will look