

In the last class, we learned about properties of individuals and properties of populations. Statistics is about the properties of populations, which we summarize with samples.

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

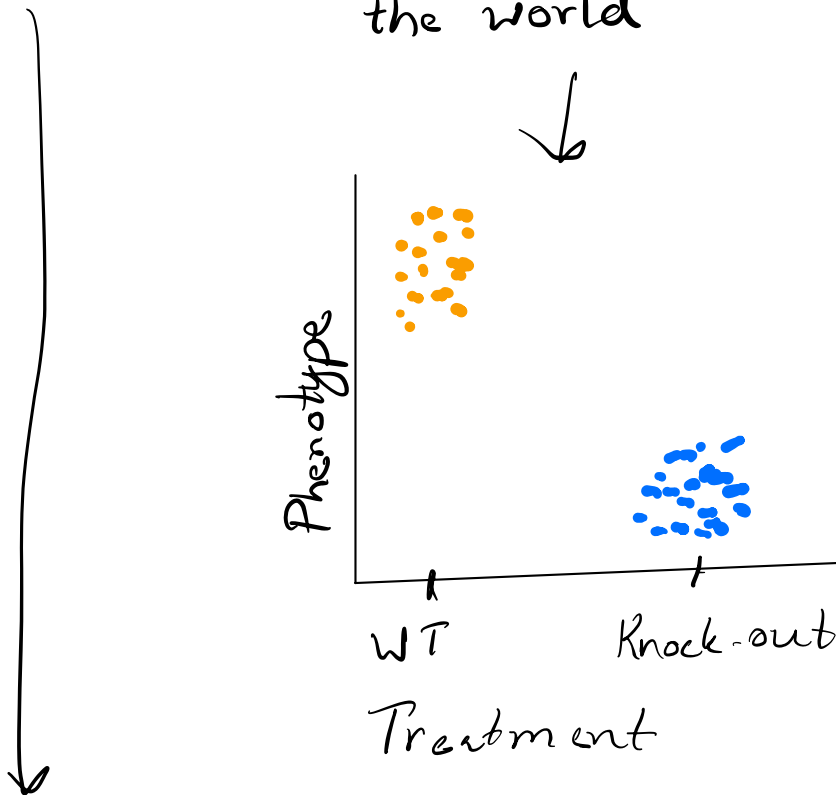
3. Species Area Relationship

Fundamental Objective

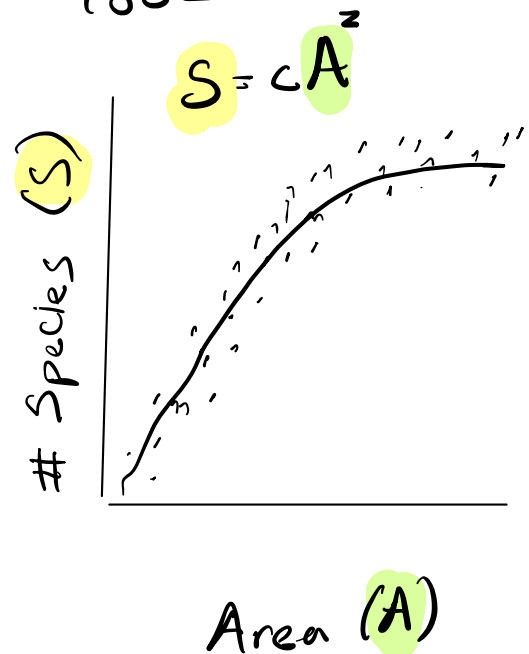
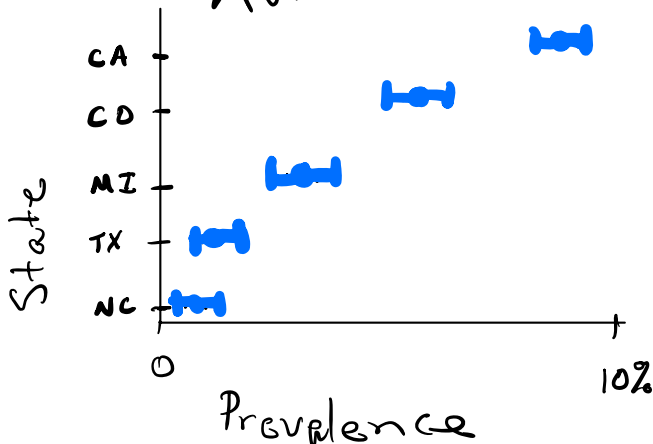
1. Descriptive

2. Understanding causal nature of the world

3. Predicting how the world will look



Avian Flu Prevalence



Prevalence is how common a disease is in a population.
 E.g., if 300 people have a disease in a population of 1000, the prevalence is $\frac{300}{1000} = 30\%$

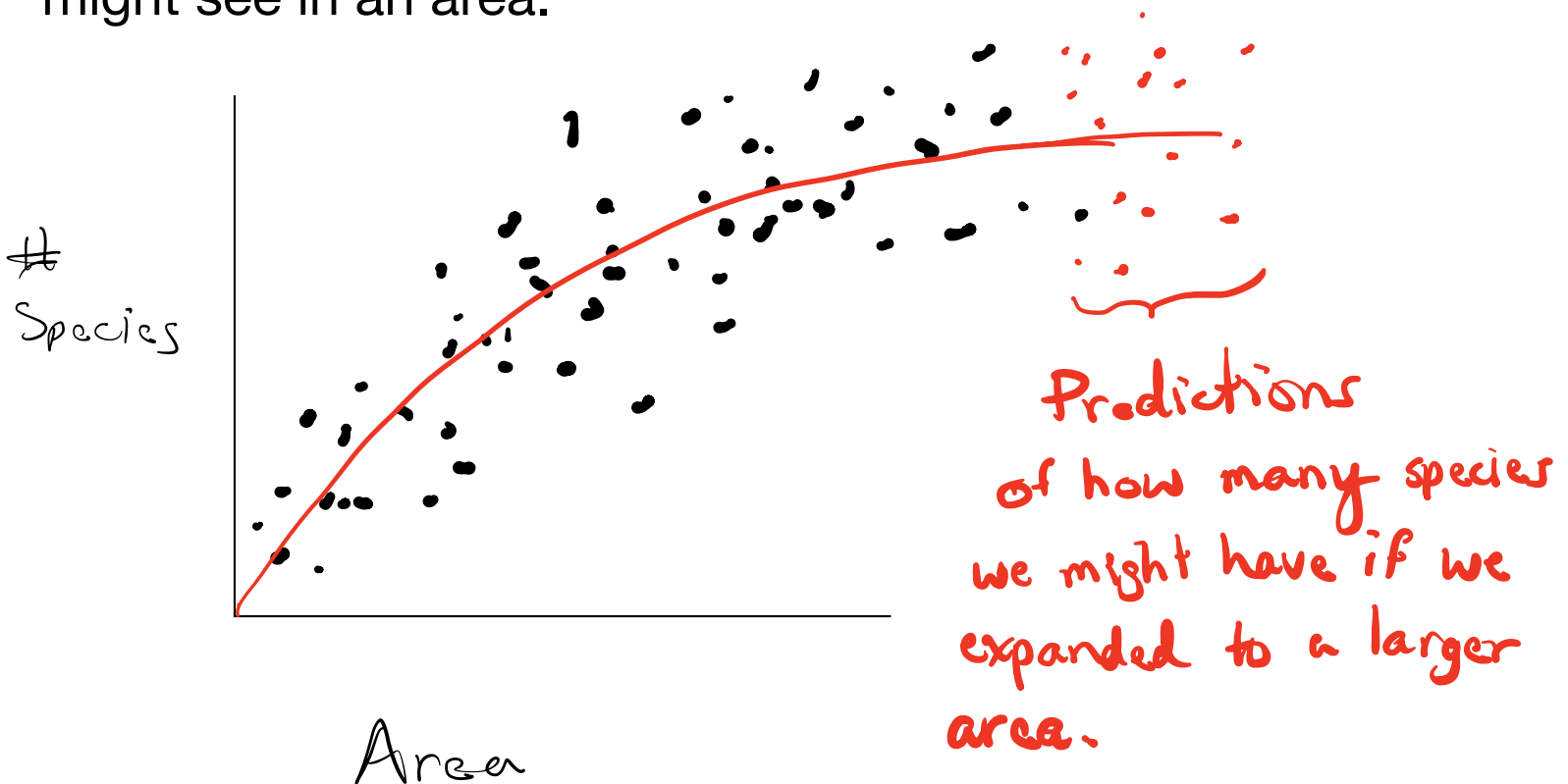
Estimating prevalence of avian flu likely comes with the goal of understanding the “state of affairs” of a particular disease. A first step to understanding a disease is knowing how common it is. We can only know how common it is from a sample of the population. We use statistics to help us make inferences about the actual prevalence in the population from what we measure in a sample. The most important statistical concept in this example is the a quantification of a range of values that encompass the ***uncertainty in our estimate*** of prevalence. This is done with ***confidence intervals*** and similar measures.

Hypothesis testing in the case of a gene knockout study is done to evaluate the evidence consistent with a particular claim about the world. We might have the hypothesis that a gene is responsible for expression of a particular phenotype. To evaluate that claim, we set up a construction where we envision how the world would look under that particular claim. To help in this goal, we ask how an experiment might look if the hypothesis were not true (the so-called “null hypothesis”). A p-value measures how likely a particular observation is under an assumption that the null hypothesis is true.

Hypothesis testing typically says nothing about how large effects are, nor whether they really exist. They only tell us whether we have sufficient evidence to be skeptical of the null.

Prediction differs from both estimation and hypothesis testing in that we don’t evaluate how the world works, we just extrapolate from the information we have to new data we

might collect in the future. The species area relations (SAR) is a robust and consistent pattern in ecology that says that the number of species (S) increases with sampling area (A) according to a power-law function, $S = cA^z$, where c and z are estimated in a given ecosystem. This means that a plot of $\log(S)$ and $\log(A)$ show a straight line relationship. Once we have this relationship, we can predict how many species we might see in an area.



For many studies, there is a role for all three goals of statistics. For example, maybe we want to estimate the species area relationship for the Hawaiian archipelago and test the idea that the Hawaiian archipelago can support more diversity than other island chains in the South Pacific Ocean. And, we might want to predict how many species might be present on the newest island as it's size increases.

These can all be done at once, and require different statistical techniques. The point is not that there are mutually exclusive

techniques, but that they all require different statistical approaches to deal with different goals of statistical analysis.

