1 Overview and Introduction

Lecture 0:

1. Probability Concepts (Last semester)
    - Basic Probability
    - Discerete Probability Distributions
    - Continuous Probability Distributions
    - Multivariate Probability Distributions
    - Transformations of Random Variables
2. Inferential Statistics (This Semester)
    - Review: Transformations of Random Variables
    - Central Limit Theorem
    - Estimation
    - Evaluating Estimators
    - Hypothesis Testing

2 Statistics Concepts

2.1 Why do we do Statistics?

- When we use statistics, we usually want to know something about a population
    - UCONN students
    - CT residents
    - US population
- But getting this information is not easy
    - Too expensive
    - Too time consuming
- So we take a small sample of the population and use the characteristics of the sample as an approximation of the desired characteristics of the population (This is called inferential statistics)
- In order to understand what our sample tells us about the population of interest, we need to know about the probability *Distribution* of our sample
- Last semester we discussed fundamentals of probability and discussed different common probability distributions

2.2 What kinds of things do we want to know?

- Usually we want to know about an attribute that every member of the population has
  - Height
  - Gender
  - Weight
  - Residency
- Each of these characteristics is called a *variable*
  - In general, a variable is a value that is unknown
  - For our purposes, a variable is some attribute of interest that every member of the population has
  - When we measure one of these attributes, we consider that measurement as an observation of a *random variable* that has some probability distribution.

2.3 Parameter Vs. Statistic

- When we are interested in a population we often take a sample of the population (a small subset of the population), and use our measurements of the distribution of the sample to estimate the measurements of the population.
- Because we are typically talking about the same kind of measurement of both populations and samples, we use the following definitions:
  - Defininton: A *Parameter* is a measurement of a population. This is often unknown to us, and always will be. that is why we estimate it.
  - Defininton: A *Statistic* is a measurement of a sample. This is a value we usually know (after collecting a sample) and use it to measure the parameter of interest
- When we characterize the population as having a specific distribution (like poisson or beta, for example), the parameters that define these distributions (λ for poisson; α and β for beta) are also considered the population parameters.

3 Goals

- For this semester, now that we know about probability and various distributions, we can start applying that knowledge to the kind of problems that statistics aim to answer.
- In doing this, the general procedure is as follows:
  - In a real world situation we would take a sample of size n from some population,

- this sample is assumed to be randomly selected from the population,
- thus, we can represent the sample as a sequence of random variables, $X_1, \ldots, X_n$, where each random variable is assumed to have the same distribution (the distribution of the population) and all of the random variables are independent.
- From this sample we can combine these random variables to 1) estimate the parameters of the distribution of the population and 2) answer questions about those parameters, such as "is the mean above 2.5 lbs", or "is the proportion of students with a car above 50%".
- So, learning how to (1) estimate parameters and (2) answer questions about parameters (this is called hypothesis testing) are the objectives of this course.