Introduction to Statistics, Probability and Econometrics: Lecture I

I. The basic question to be answered on the first day is: What are we going to study over the next fifteen weeks and how does it fit into my graduate studies in Food and Resource Economics?

   A. The simplest (and most accurate) answer to the first question is that we are going to develop statistical reasoning using mathematical reasoning and techniques.

   B. The answer to the second part of the question requires is a little more complicated.

      1. Following Kmenta, statistical applications can be divided into two sub-groups: descriptive statistics and statistical inference.

         a) Kmenta's claim is that most statistical applications in economics involve the application of techniques for statistical inference.

         b) However, this position ignores the concept of decision making under risk.

         c) From a general statistical perspective, mathematical statistics allows for the formalization of statistical inference.

            (1) How do we formulate a test for quality (light bulb life)?

            (2) How do we develop a test for the significance of an income effect in a demand equation?

      2. Related to the general problem of statistical inference is the study of Econometrics.

         a) "Econometrics is concerned with the systematic study of economic phenomena using observed data." Spanos p.3.

         b) "Econometrics is concerned with the empirical determination of economic laws." Theil p.1.

         c) Econometrics is the systematic study of economic phenomena using observed data and economic theory.

      3. Economic theory, most particularly production economics, relies on the implicit randomness of economic variables to develop models of decision making under risk:

         a) Expected Utility Theory

         b) Capital Asset Pricing Models

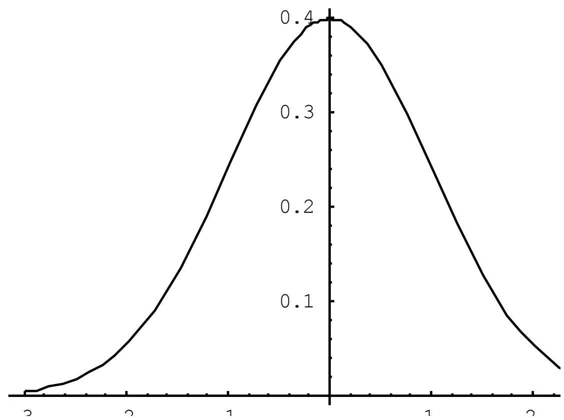
         c) Asymmetric Information

II. An Example of Inference versus Decision Making

   A. Skipping ahead a little bit, the normal distribution function depicts the probability density for a given outcome x as a function of the mean and variance of the distribution:

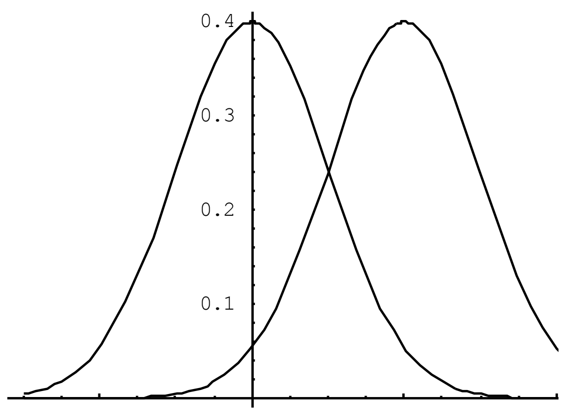
$$f(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Graphically, the shape of the function can be depicted as



Assuming that $\mu=0$ and $\sigma^2=1$.

B. Statistical inference involves testing a sample of observations drawn from this data set against an alternative assumption, for example $\mu=2$.



C. Economic applications involve the choice between the two distribution functions.

III. What is probability?

A. Two definitions:

1. Bayesian – probability expresses the degree of belief a person has about an event or statement by a number between zero and one.
2. Classical – the relative number of time that an event will occur as the number of experiments becomes very large.

$$\lim_{N\to\infty} P[O]=\frac{r_O}{N}$$

IV. What is statistics?

A. Definition I: Statistics is the science of assigning a probability of an event on the basis of experiments.

B. Definition II: Statistics is the science of observing data and making inferences about the characteristics of a random mechanism that has generated the data.

C. By random mechanisms, we are most often concerned with random variables:

1. A Discrete Random Variable is some outcome that can only take on a fixed number of values.

a) The number of dots on a die is a classic example of a discrete random variable.

b) A more abstract random variable is the number of red rice grains in a given measure of rice. It is obvious that if the measure is small, this is little different than the number of dots on the die. However, if the measure of rice becomes large (a barge load of rice), the discrete outcome becomes a countable infinity, but the random variable is still discrete in a classical sense.

2. A Continuous Random Variable represents an outcome that cannot be technically counted.

a) Amemiya uses the height of an individual as an example of a continuous random variable. This assumes an infinite precision of measurement.

b) The normally distributed random variable presented above is an example of a continuous random variable.

3. The exact difference between the two types of random variables has an effect on notions of probability.

a) The standard notions of Bayesian or Classical probability fit the discrete case well. We would anticipate a probability of 1/6 for any face of the die.

b) In the continuous scenario, the probability of any outcome is zero. However, the probability density function yields a measure of relative probability.

c) The concepts of discrete and continuous random variables are then unified under the broader concept of a probability density function.

D. Definition III: Statistics is the science of estimating the probability distribution of a random variable on the basis of repeated observations drawn from the same random variable.