Statistics 305

Definitions, Notes and Examples

(See Textbook Chapters 1 and 2)

The following are definitions of words and phrases which convey key basic concepts in statistics. As we progress thru the material in this course you will come to understand that these definitions are somewhat simplistic, and stronger definitions can be made. Nevertheless, the definitions given below will suffice as we begin the course. Memorize these, study the notes and examples and try to think of other examples.

Definitions

Population: A population is the entire group of numbers about which one wishes to gather information in a statistical study.

Element: The numbers in a population are called elements of the population.

Sample: A subset of the elements in a population is a sample from the population.

Random Sample: A random sample is a sample obtained so that the specific elements which comprise the sample cannot be known in advance.

Statistical Study: A statistical study is an activity which involves investigating a population by

1. taking a random sample from a population,
2. summarizing or describing the elements in the sample, and
3. making inferences about the characteristics of the population based on the like characteristics of the random sample.

Notes and Examples

Note 1: Some people prefer to think of a population as consisting of real objects, like the gears described in Example 1 on page 2 of your textbook, despite the fact that it is the associated numbers (thrust face runout) which are the center of attention. Ok, no problem, study this characteristic (thrust face runout) of the individual gears in the population of gears. Measure the thrust face runout of each gear in a random sample to get a random sample of thrust face runouts from the imagined population of thrust face runouts.
Note 2: A statistical study provides evidence about what the characteristics of a population may be. It does not provide proof, only evidence. Typically, more than one similar study is done to address questions about important populations. Each study yields evidence. As the amount of evidence increases, the actual population characteristics begin to emerge.

Note 3: It is most often the case that population(s) employed in a statistical study don’t exist. They are useful figments of our imagination.

Note 4: Section 2.2 in your textbook discusses random sampling, but only in enumerative studies. These feature a concrete population of elements. It is not, of course, possible to draw a truly random sample (in the sense of Section 2.2) for populations that are figments of our imagination. We can, however, do whatever we can to introduce randomness into the sample selection process.

Example 1: Consider Example 1 in your textbook. Imagine two different populations. One consists of the thrust face runout numbers for all gears which might be passed thru the furnace laid. The other consists of the thrust face runout numbers for all gears which might be passed suspended thru the furnace. Of course, before a gear goes thru the furnace, we can’t know the number, but we imagine one which will be realized after heat treating. We imagine these populations as containing numbers in proportions that are “true” for each process.

To obtain a random sample from the first population, we select a random sample (as best we can) of gears, send them laid thru the furnace, and measure the thrust face runout of each after emerging from the furnace. A random sample from the second population is obtained in similar fashion for suspended gears.

The objective of this statistical study is to compare the mean (average) thrust face runout of the two populations. To do this, we first compute the mean (average) of each random sample. These we take as estimates of the respective population means. Comparison of the sample means is then made and based on this comparison we have evidence which allows us to infer which population appears to have the smaller mean.

This statistical study features sampling from two populations so the situation doesn’t exactly fit the definition given of a statistical study. However, the essential ideas are there.

Example 2: A warehouse contains 5000 amplifiers manufactured by a company during the past few months. Suppose we wish to estimate the average gain of these amplifiers, but we can only actually measure gain for about 50 due to budget restrictions. This calls for a statistical study.

Imagine a population consisting of 5000 gains (one for each amplifier). To get a random sample from this population, we randomly select (say) 50 amplifiers and measure the gain for each. These 50 numbers constitute our random sample from the population of 5000 gains.

To estimate the population mean gain we compute the average sample gain. Then we infer that the mean sample gain is approximately the population mean gain. The statistical study has, you can see, produced evidence about the possible value of the population mean.
Example 3: Consider the company, from example 2, which makes amplifiers using an assembly line method. Suppose we ask, “what is the average gain of amplifiers being produced by this method”?

Here we can imagine a population of gains. Every time an amplifier is completed, its (unknown) gain is a number selected at random from this population. Our question can now be viewed as being “what is the mean of this population of gains?” Obviously, we can never know the exact answer to this question, but we can do statistical study(s) to gather evidence about its possible value. A study consists of taking a random sample (as best we can) of completed amplifiers, measuring gain for each (study step 1), computing the average of measurements (study step 2) and concluding that this is approximately the average gain produced thru the manufacturing method (study step 3).

Example 4:

**Basic question** – I quickly check TV channel 13 several times each day to see if there is important breaking news. How likely is it that I will find a commercial playing? (Channel 13 commercial broadcast plans are not available.)

**A question we will address** – If one tunes to TV Channel 13 at 10 different arbitrarily selected times during any day of any week, on average how many of the 10 tune-ins will find a commercial being broadcast?

**An Associated Population** – For any set of 10 tune-ins the number of commercials found is either 0, 1, 2, …, or 10. Imagine a population consisting of these 11 integers in proportions which match what is “true” in this situation. The mean of this population is what we would like to know. We can’t know it because this population doesn’t exist, but we can estimate it using a statistical study.

**Possible Statistical Study** – At 10 randomly selected times in a selected day, tune in and count the number of commercials found. This number can be thought of as approximating a random sample of 1 from the population of interest. Repeat this as many days (say n) as you wish. This completes Step 1 of the Statistical Study. Step 2 is to summarize by computing the average of the n numbers. Step 3 consists of stating that the mean computed in Step 2 is the approximate average number of times in 10 tune-ins that one can expect to find a commercial.

Example 5:

**Basic question** – What is the average mint year of pennies in circulation in Ames?

**An Associated Population** – Every penny has a date, like (say) 1935 stamped on it. Imagine a population of year numbers, one for each penny in circulation. This population does (essentially) exist but we cannot acquire it. Its mean is what we would like to know, but of course we cannot know it.

**Possible Statistical Study** – Acquire pennies in some way which is viewed as arbitrary and might reasonably be thought of as a random sample of pennies from the entire set. Record the date on each. This is taken as a random sample from the population of year numbers. The mean of these is the approximate answer to the basic question.
Example 6:

Basic question – I like either fresh squeezed orange juice or fresh squeezed sweetened lemon juice for breakfast. To get the most juice (on a per ounce of fruit basis) should I buy oranges or lemons?

(What would you do?)

What will we do this semester?

We will study various elements connected with the 3 steps in a Statistical Study.

- The first two or three classes will give more details about populations, especially those that are derived in mathematical theory. This material is given in Handouts 2, 3 and 4. The idea of population is a part of Step 1 in a Statistical Study. Textbook Chapter 1 defines the term “population”.

- Next, probably in class 3, we will discuss random sampling. This discussion will be brief because much more will be said when we get to textbook Chapter 5. For now, this will be our coverage of Step 1 in a Statistical Study. Textbook Chapter 2 briefly discusses aspects of “random sampling”.

- Chapter 3 in your textbook gives graphical and numerical methods for summarizing the sample elements. Summarization is done in Step 2 of a Statistical Study. Several classes will deal with this material.

- Next, parts of the material in Chapter 5 will be discussed. This contains the theory of randomness and random sampling. Details not mentioned in class 3 will be given in several classes which address Chapter 5 material. This will take us through about ½ of the semester, and provide the basic knowledge needed to study methods for making inference about population characteristics as is required in Step 3 of a Statistical Study.

- The second half of the semester will cover methods for statistical inference in different kinds of situations. Chapters 6, 7 then 4 and 9 will be addressed. These methods are used in Step 3 of a Statistical Study.