Stat 400, section 6.1a Point Estimation (Theory and an Application)

notes by Tim Pilachowski

Recall back to Lecture 5.3a:

In the real world, people often want (or need) to know information about some large group of people or items: a **population**.

The numbers associated with the desired information are called the **population parameters**.

It frequently occurs that we know something about the probability distribution of a population, without knowing the value of the parameter. [We'll use the general convention and designate the generic version of the unknown parameter with the Greek letter θ.]

When the population consists of discrete values, the probability mass function can be designated $p_X(x,\theta)$ to emphasize that the distribution depends on the, as yet unknown, value of θ.

In a similar fashion, when the population consists of continuous values, the probability density function can be designated $f_X(x,\theta)$.

It is important to note that θ is *not* a variable, but rather *is* a parameter—it has a fixed value, but we don't know what that value is.

Examples of variables vs. parameters:
For a quadratic equation, $f(x) = ax^2 + bx + c$, x is the variable (continuous) and a, b and c are parameters, which take on fixed values in given specific situations.
For integration we have a general solution, $\int f(t)\,dt = F(t) + C$, in which t is the variable, and C is a parameter whose value may be determined from additional information.

Can we calculate with parameters if we don't know their values? Sure we can (and often do) to derive generic formulas that can then be applied in various specific situations.

Example A: "National NAEP reports statistical information about student performance and factors related to educational performance for the nation and for specific student groups in the population (e.g. race/ethnicity, gender). It includes students drawn from both public and nonpublic (private) schools and reports results for student achievement at grades 4, 8, and 12." Source: http://nces.ed.gov/nationsreportcard/about/national.asp

The parameter in this case would be…

Example B. A news organization polls n voters and asks, "Do you intend to vote for incumbent Senator Phillip E. Buster in the upcoming election?" Pollsters record a value of 1 for a "Yes" answer and 0 for a "No" answer.

The parameter in this case would be…

Example C. A factory needs to evaluate the lifetime of its production machinery. The lifetime of similar machinery has an exponential probability distribution.

The parameter in this case would be…

Example D. McCormick Consumer testing asks participants to rate various characteristics of food items on a scale from 1 to 10. There is a baseline assumption that, if participants simply make random choices, the values have a uniform probability distribution.

The parameter in this case would be…

As of chapter 6, we are moving into the statistics portion of this class.

We don't know and most often don't have a way to determine the precise values of population parameters (θ). However, we want to be able to at least make a "guess", called a "**point estimate**", of the value of θ.

With mathematical backup, we'll be able to develop a theory for finding a point estimate which will move us from a random guess to an informed one, and we'll be able to state with some probability how close our point estimate is to the population parameter θ.

More formally:
We begin with a sample (a subset of the population) consisting of values $x_1, x_2, x_3 \ldots x_n$.
We want to find a function $h(x_1, x_2, x_3, \ldots x_n)$ which will serve as a good point estimate, or estimator, for the population parameter θ.

Simply saying "Choose h in such a way that $h(x_1, x_2, x_3, \ldots x_n) = \theta$" doesn't work—the sample would have to include the entire population.

A slightly better approach would be to say "Can we establish quantitative criteria to help us decide whether one estimator, $h_1(x_1, x_2, x_3, \ldots x_n)$, is better than another estimator, $h_2(x_1, x_2, x_3, \ldots x_n)$?"

Sounds good, but how do we go about establishing the needed criteria?

We'll begin at a point in time before the sample is chosen and before sample values are determined.

In other words, given a sample size, n, random variables $X_1, X_2, X_3, \ldots, X_n,$ are defined representing the potential outcomes of the experiment.

Recall from Lecture 5.3a that these random variables must fulfill two (important) requirements:

1. The X_i's are independent random variables.
2. Every X_i has the same probability distribution.

Arithmetic combinations of these random variables give us another random variable

$$W = h\left(X_1, X_2, X_3, \ldots, X_n\right)$$

which will be our estimator statistic, or point estimate for the population parameter θ.

This estimator statistic is often denoted $\hat{\theta}$ (see notes below).

Important notes about notation:

$X_1, X_2, X_3, \ldots, X_n$, are *random variables*.

$x_1, x_2, x_3, \ldots, x_n$, are *numbers*, the values that come from our actual sample, i.e. the data generated which will become the grist of a statistical analysis.

In a similar fashion, we should use a capital letter $\hat{\Theta}$ to denote the random variable which is our estimator statistic, and the small letter $\hat{\theta}$ for the observed (calculated) value of our point estimate. That is,

$$\hat{\Theta} = h(X_1, X_2, X_3, \ldots, X_n)$$

$$\text{and } \hat{\theta} = h(x_1, x_2, x_3, \ldots, x_n).$$

In actual practice, however, it is very common for $\hat{\theta}$ to be used interchangeably for both the random variable and the number generated from the sample. I'll hold on to the distinction between $\hat{\Theta}$ and $\hat{\theta}$ for the rest of this lecture, then switch to the convention of using $\hat{\theta}$ for both as of lecture 6.1b.

Side note: In his lecture notes Dr. Millson sidesteps this notation issue by using W and w instead of "theta".

Now we come to the main issue of chapter 6.

Can we find an estimator statistic $\hat{\Theta} = h(X_1, X_2, X_3, \ldots, X_n)$ such that $P(\hat{\theta} = \theta)$ is maximized?

This is, on the face of it, impossible. After all, we don't know the value of the population parameter θ, so how could we hope to ever know how good our sample statistic $\hat{\theta}$ is?

However, if we weaken this criterion a little bit, we can define a process that is not only workable, but also surprisingly easy to achieve.

Definition:

An estimator statistic $\hat{\Theta} = h(X_1, X_2, X_3, \ldots, X_n)$ is an *unbiased* estimator of a population parameter θ if it satisfies $E(\hat{\Theta}) = \theta$.

Where do we go from here?

The rest of Lecture 6.1a is practice in calculating various sample statistics, including but not limited to sample mean (random variable $\overline{X}$), sample variance (random variable S^2), sample standard deviation (random variable S), and sample proportion (random variable $\hat{p}$).

Lectures 6.1b and 6.1c will be an exploration of whether these sample statistics meet our criterion for being unbiased estimators of the population parameter θ.

θ – generic version of the unknown population parameter
- μ – population mean
- $\tilde{\mu}$ – population median
- σ^2 – population variance
- σ – population standard deviation
- p – population proportion

$\hat{\theta}$ – generic version of the calculated sample statistic
- $\bar{x}$ – sample mean
- $\tilde{x}$ – sample median
- s^2 – sample variance
- s – sample standard deviation
- $\hat{p}$ – sample proportion

Example E. Consider the following set of data: 5, 1, 9, 8, 1, 8, 1, 1, 9, 6.
Identify or calculate the following: sample size, total, mean, median, 10% trimmed mean, average of extreme values, variance, and standard deviation.