POINT ESTIMATION

Video 1. Start with the introduction to estimation.

We are interested in using the values of a random sample $X_1, X_2, \ldots, X_n$ to estimate the values of population parameters (usually μ or σ^2, but also including other parameters). For example, in the COVID-19 example I used the sample mean (the sample total divided by the sample size) as an estimate of the proportion of people in Florence with COVID-19. Another example: if you wanted to know the mean age of the trees in a forest, you could select 100 random trees and calculate the ages of those 100 trees. Then you could use the mean age of trees in your sample as an estimate for the mean age of trees in the whole forest.

Definition. We call the sample statistic $\hat{X}$ (a random variable determined by the values of a random sample of size n) an **estimator** of the population parameter ρ if the value of $\hat{X}$ will be used as an estimate of ρ.

(1) If $E(\hat{X}) = \rho$, then we call $\hat{X}$ an **unbiased estimator**.

(2) If for any $c > 0$, $\lim_{n\to\infty} P(|\hat{X} - \rho| < c) = 1$, then we call $\hat{X}$ a **consistent estimator**.

Video 2. Watch the introduction to the example.

Example. We know that population is uniformly continuously distributed on the interval $[0, \beta]$, but we don't know β. We found that if $\overline{X}$ is the mean of a random sample, then $E(\overline{X}) = \frac{0+\beta}{2} = \frac{\beta}{2}$. It follows that $E(2\overline{X}) = \beta$. Thus $2\overline{X}$ is an unbiased estimator of the population parameter β.

1. Keep working with the population in the example. Let $X_1, X_2, \ldots, X_{10}$ be a random sample and let $\hat{X}$ be the maximum of the sample. Our goal is to find an unbiased estimator for β based on $\hat{X}$. In the video, I made use of the fact that the CDF of any single random variable from the population is

$$G(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \frac{x}{\beta} & \text{if } 0 < x < \beta \\ 1 & \text{if } \beta \le x \end{cases}$$

From this it follows that the CDF of the sample max $\hat{X}$ is

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \frac{x^{10}}{\beta^{10}} & \text{if } 0 < x < \beta \\ 1 & \text{if } \beta \le x \end{cases}$$

a) Differentiate the CDF to find a PDF for $\hat{X}$

b) Use the PDF to calculate $E(\hat{X})$

c) Find a number c such that $E(c\hat{X}) = \beta$

Your unbiased estimator for β is $c\hat{X}$.

Date: March 16, 2020.

I collected a random sample from a population uniformly continuously distributed on $[0, \beta]$ and got the following values:

0.8023426	1.1194331	1.5173202	1.5353951	2.1296237
2.5989737	3.4224997	3.7957449	4.8618741	5.0795233

This means $2\overline{x} = 5.372546$ and $1.1\hat{x} = 5.587476$ (I'm using lowercase letters here because I have actual values for the sample statistics and I'm no longer thinking of them as random variables). We now have the values of two unbiased estimators for β. Which should we actually use? Is one of the estimators more reliable than the other? There's a general principle that applies: **when given the choice between unbiased estimators, you should always choose the estimator with smaller variance**.

Definition. The standard deviation of an estimator is often referred to as the **standard error** of the estimate. Standard errors can be thought of as measures of the reliability of an estimator (with smaller values being better).

2. Your job now is to calculate the variances of the estimators.

a) Use the theorems of the last worksheet to calculate $\text{Var}(2\overline{X})$ (it will help to know that the population variance is $\sigma^2 = \frac{\beta^2}{12}$).

b) You'll have to calculate the variance of the other estimator by hand. You already know the expected value of $\hat{X}$, so it makes sense to use the formula $\text{Var}(\hat{X}) = E\left(\hat{X}^2\right) - \left[E(\hat{X})\right]^2$.

Which estimator has a smaller variance?