School of Oriental and African Studies **September 2013**
Department of Economics

Preliminary Statistics
Lecture 4: Estimation and Confidence Intervals (Outline)

Gujarati D. *Basic Econometrics*, Appendix A.7
Barrow M. *Statistics for Economics, Accounting and Business Studies* Ch. 4

1 Introduction

Statistical inference uses probability theory to judge how much confidence we have in our summary statistics. Estimation is a procedure in which we use the information included in a sample to get inferences about the true (population) parameter values of interest.

2 Point Estimation

A point estimator is a rule (formula) that tells us how to calculate the value of an estimate based on the measurements contained in a sample. An estimate is a particular value of the estimator.

Assume that $Y \sim N(\mu, \sigma^2)$, but the true population parameters (μ, σ^2) are unknown. If we have a random sample, Y_i, $i = 1, 2, ...n$ of size n from the population distribution, we may use the sample mean, $\hat{\mu} = \bar{Y}$ and the sample variance, s^2 to estimate the unknown parameters.

Due to the variability of the sampling process, an estimator is a random variable itself. Every time we get a different sample, we obtain a different estimate. With many different samples, we can form the distribution of the estimator, and hence determine the mean and the variance of the estimator. Such a distribution is known as the sampling distribution, given that the sample is random. Recall that a sample is random if all Ys are drawn independently from the same probability distribution. That is, each Y_i included in the sample has the same PDF and each Y_i is drawn independently.

2.1 Properties of (point) estimators

For the same population parameter many different estimators may be used. In the case of the population mean, apart from the sample mean, $\hat{\mu}$, the median, $\hat{m}$, or the first sample observation can be used as estimators for the true parameter μ. How can we choose among estimators? Some estimators are good and some bad. Hence, we need to establish criteria of 'goodness' to compare estimators.

Suppose we want to estimate the true parameter θ of a random variable Y from a random sample, $Y_i, i = 1, 2, ...n$.

2.1.1 Small Sample Properties

1. Linearity

   An estimator is said to be a linear estimator if it is a linear function of the sample observations. That is, $\hat{\theta}$ is linear if $\hat{\theta} = f(Y) = a_1 Y_1 + a_2 Y_2 + ... + a_n Y_n$,

where $a_1, a_2...a_n$ and constants. Linearity is a desirable property of an estimator because it is easier to treat than non-linear estimators, but not always possible.

2. Unbiasedness

   An estimator is said to be unbiased if on average (over lots of hypothetical samples) is equal to the true parameter value. That is,

   $$E(\hat{\theta}) = \theta$$

   It is important to note the difference between the following three concepts:

   - Sampling error $= \hat{\theta} - \theta$
   - Bias $= E(\hat{\theta}) - \theta$
   - Mean square error (MSE) $= E(\hat{\theta} - \theta)^2$

3. Efficiency

   An efficient estimator is one whose sampling distribution has the smallest variance amongst unbiased estimators. In general, $\hat{\theta}$ is an efficient estimator of θ if

   (a) $\hat{\theta}$ is unbiased, and

   (b) $Var(\hat{\theta}) \leq Var(\tilde{\theta})$ where $\tilde{\theta}$ is any other unbiased estimator of θ.

   Efficiency is a desirable property because the smaller variance guarantees that in repeated sampling a high fraction of values of $\hat{\theta}$ (estimates) will be closer to θ.

BLUE

When an estimator fulfills all three properties is called BLUE (Best Linear Unbiased Estimator). In general, $\hat{\theta}$ is a BLUE of θ if:

(a) $\hat{\theta}$ is a linear function of the sample observations, and

(b) $E(\hat{\theta}) = \theta$ and

(c) $Var(\hat{\theta}) \leq Var(\tilde{\theta})$ with $\tilde{\theta}$ any other unbiased estimator.

2.1.2 Asymptotic Properties

Asymptotic properties of estimators relate to the distribution of an estimator when the sample size is large and approaches infinity. Such sampling distribution is known as the asymptotic distribution.

- Asymptotic unbiasedness

  $\hat{\theta}$ is an asymptotically unbiased estimator of θ if

  $$\lim_{n \to \infty} E(\hat{\theta}) = \theta$$

  A common example of a biased but asymptotically unbiased estimator is the variance.

- Consistency

  $\hat{\theta}$ is a consistent estimator of θ if it approaches the true value of the parameter as the sample size gets larger and larger. The point on which the distribution of $\hat{\theta}$ collapses on θ is called the probability limit of $\hat{\theta}$, frequently abbreviated as $\operatorname{plim}\hat{\theta}$. Hence formally, we may say that $\hat{\theta}$ is a consistent estimator of θ if $\operatorname{plim}\hat{\theta} = \theta$.

  A more intuitive way of understanding consistency is to say that as the sample size increases biases as well as variance reduces. The distribution of estimates gets closer and closer to the true value so that with $n = \infty$ there is no dispersion at all; the estimator converges to its true value and we can estimate it exactly.

- Asymptotic efficiency

  $\hat{\theta}$ is an asymptotically efficient estimator of θ if

  (a) $\hat{\theta}$ has an asymptotic distribution with finite mean and finite variance.
  (b) $\hat{\theta}$ is consistent.
  (c) No other consistent estimator of θ has a smaller asymptotic variance than $\hat{\theta}$.

3 Estimating the Expected Value of a random variable

Suppose we want to estimate the true (population) mean, μ, of a random variable Y, from a sample Y_i, $i = 1, 2, ...n$. One possible *estimator* is the sample mean $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

3.1 Properties of the estimator for μ

Based on the properties of estimators, it can be shown that the sample mean is BLUE, and hence is preferred from the other possible estimators.

- The sample mean is a linear function of the sample observations.

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + ... + \frac{1}{n}Y_n$$

- The sample mean $\hat{\mu}$ is an unbiased estimator for μ. That is, $E(\hat{\mu}) = \mu$.

$$\begin{aligned} E(\hat{\mu}) &= E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}E\left(Y_1 + Y_2 + ... + Y_n\right) \\ &= \frac{1}{n}\left[E(Y_1) + E(Y_2) + ... + E(Y_n)\right] = \frac{1}{n}[\mu + \mu + ... + \mu] = \frac{1}{n}n\mu = \mu \end{aligned}$$

- The sample mean $\hat{\mu}$ has the smallest variance among the unbiased estimators for μ. To find the variance of the sample mean we will use two properties of the variance: (i) $\text{Var}(aY) = a^2\text{Var}(Y)$; (ii) $\text{Var}(Y_1 + Y_2) = \text{Var}(Y_1) + \text{Var}(Y_2) + 2\text{Cov}(Y_1, Y_2)$ and the assumption that the random realisations of the the variable are independently

drawn, so that $\mathrm{Cov}(Y_1, Y_2) = 0$.

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}) &= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) \quad \text{(from property (i), since n is a constant)} \\
&= \frac{1}{n^2}\mathrm{Var}\left(Y_1 + Y_2 + ... + Y_n\right) = \frac{1}{n^2}\left(\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + ... + \mathrm{Var}(Y_n)\right) \\
&\quad \text{(from property (ii) and the fact that all realisations are independent)} \\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + ... + \sigma^2) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

It can be shown (Markov-Gauss Theorem) that the variance of the sample mean is the smallest amongst all the unbiased estimators for μ. The square root of the variance of the sample mean, $\sigma/\sqrt{n}$, is called the standard error of the mean. It is used to provide an indication of how accurate our estimate is. Notice when we take the square root of the variance of a **random variable** we call it a **standard deviation**; when we take the square root of a variance of an **estimator**, we call it a **standard error**.

3.2 Estimating the Variance

σ^2 is the true (unknown) variance of the random variable Y, and the variance of the sample mean (estimator for the true mean) was found to be one n-th of the true variance. So, to get an estimate for the variance of the sample mean, we need to find an estimator for σ^2.

There are two common estimators of the variance of Y:

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{\mu})^2}{n} \text{ and } s^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{\mu})^2}{n-1}$$

The first estimator, $\widehat{\sigma}^2$, usually called the population variance, which divides by n, is a biased estimator of σ^2. That is, $E(\widehat{\sigma}^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$. The second estimator, s^2, known as the sample variance, is an unbiased estimator, $E(s^2) = \sigma^2$. The bias arises because we use an estimate of the mean and the dispersion around the estimate is going to be smaller than the dispersion around the true value because the estimated mean is designed to make the dispersion as small as possible. If we used the true value of μ there would be no bias. The correction $n-1$ is called the degrees of freedom: the number of observations minus the number of parameters estimated, one in this case, $\hat{\mu}$.

This way we get an estimate for the standard error of the sample mean by

$$\widehat{SE(\hat{\mu})} = \frac{s}{\sqrt{n}}$$

Note that, similarly to the sample mean, the estimator for the variance will also be a random variable with a distribution, a mean and a variance, of its own.

3.3 Distribution of the sample mean

As an estimator, $\hat{\mu}$, is a random variance in itself. Hence, we also need to find how this random variable is distributed.

If the random variable Y is normally distributed, the sample mean will also be normally distributed. Hence,

$$\text{If} \quad Y \sim \text{N}(\mu, \sigma^2) \Rightarrow \hat{\mu} \sim \text{N}(\mu, \frac{\sigma^2}{n})$$

If the random variable Y is not normally distributed, then according to the ***Central Limit Theorem*** the sample mean will be approximately normally distributed as the sample size approaches infinity.

$$\text{If} \quad Y \sim (\mu, \sigma^2) \Rightarrow \hat{\mu} \sim \text{N}(\mu, \frac{\sigma^2}{n}) \quad \text{as} \quad n \to \infty$$

Therefore, we now have the distribution of the estimator for μ.

$$\hat{\mu} \sim \text{N}(\mu, \sigma^2/n)$$

4 Interval Estimation (Confidence Intervals)

We can also provide our inferences about the true population parameter in the form of an interval. An interval estimator is a rule specifying the method for using the sample information to calculate two endpoints that form the interval, which is intended to enclose the true parameter value.

Assume, again, we have a random sample $Y_1, Y_2, ...Y_n$ from a variable Y, and we are interested in obtaining an interval estimate for its unknown parameter θ. The endpoints of the interval, being functions of the sample measurements, will vary randomly from sample to sample. Hence, the interval estimator is a random variable. Our objective is to find an interval estimator capable of generating narrow intervals that have a high probability of enclosing the true parameter of interest, θ.

Interval estimators are commonly called confidence intervals. The probability that a confidence interval will enclose θ is called the confidence coefficient (conventionally set to 95%). The confidence coefficient identifies the fraction of the time (e.g. 95% of the times), in repeated sampling, that the interval constructed will contain the true parameter θ. Suppose that $\hat{\theta}_L$ and $\hat{\theta}_U$ are the (random) lower and upper endpoints for a parameter θ. Then, $\text{Prob}(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha$ represents the $(1 - \alpha)\%$ confidence interval for the parameter θ.

4.1 Confidence interval for the mean

To obtain a confidence interval for the true unknown mean (μ) of a random variable Y, given a random sample, $Y_1, Y_2, ...Y_n$:

1. Find the distribution of the estimator $\hat{\mu}$, sample mean.
   That is, find how $\hat{\mu}$ is distributed, its mean, $E(\hat{\mu})$, and its variance $\text{Var}(\hat{\mu})$, and hence its standard error, $\text{SE}(\hat{\mu})$.

2. Standardise the estimator, i.e., construct the Z-score for $\hat{\mu}$ by subtracting its mean and dividing by its standard error.

$$Z = \frac{\hat{\mu} - E(\hat{\mu})}{\text{SE}(\hat{\mu})}$$

3. Find the distribution of the standardised estimator, Z

   - If we have a **large sample** (say, $n > 30$), then Z follows the standard normal distribution, with mean zero and variance/SD one. $Z \sim SN(0, 1)$
   - If we have a **small sample** (say, $n < 30$) and the true variance of the random variable Y, σ^2, is unknown, then Z follows the t-distribution with $n-1$ degrees of freedom, $Z \sim t_{n-1}$.

   When we estimate the variance, σ^2, by s^2 our estimated standard error is $\widehat{SE(\mu)} = s/\sqrt{n}$. This adds extra uncertainty and thus $(\hat{\mu} - E(\hat{\mu})/\widehat{SE(\mu)}$ follows the t-distribution, which is more spread out, depending on the degrees of freedom, $n-1$. As the number of observations, n, becomes large the effect of estimating the variance becomes smaller and the t-distribution becomes closer to a normal distribution.

4. Find the 2 values in the tails of the Z distribution, $-z_{\alpha/2}, z_{\alpha/2}$ such that:

$$\text{Prob}\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = (1-\alpha)$$

   Substitute for Z in the probability statement:

$$\text{Prob}\left(-z_{\alpha/2} \leq \frac{\hat{\mu}-\mu}{\text{SE}(\hat{\mu})} \leq z_{\alpha/2}\right) = (1-\alpha)$$

   And rearrange to get:

$$\begin{aligned}
&\text{Prob}\left(-z_{\alpha/2}\text{SE}(\hat{\mu}) \leq \hat{\mu}-\mu \leq z_{\alpha/2}\text{SE}(\hat{\mu})\right) = (1-\alpha) \\
&\text{Prob}\left(-\hat{\mu}-z_{\alpha/2}\text{SE}(\hat{\mu}) \leq -\mu \leq -\hat{\mu}+z_{\alpha/2}\text{SE}(\hat{\mu})\right) = (1-\alpha) \\
&\text{Prob}\left(\hat{\mu}-z_{\alpha/2}\text{SE}(\hat{\mu}) \leq \mu \leq \hat{\mu}+z_{\alpha/2}\text{SE}(\hat{\mu})\right) = (1-\alpha)
\end{aligned}$$

This is the $(1-\alpha)\%$ confidence interval for μ. The 95% confidence interval for μ, assuming we have a large sample would be:

$$\text{Prob}\left(-1.96 \leq \frac{\hat{\mu}-\mu}{\text{SE}(\hat{\mu})} \leq 1.96\right) = 0.95$$

$$\hat{\mu} - 1.96\text{SE}(\hat{\mu}) \leq \mu \leq \hat{\mu} + 1.96\text{SE}(\hat{\mu}).$$

This says that we are 95% confident that the true mean (μ) lies within approximately 2 standard errors around our sample estimator. Note that this is a confident interval (not a probability) because the true parameter, μ, is an (unknown) fixed number and no probabilities can be attached to a number. μ will either lie in the interval or not. The probabilities are attached to the random variable $\hat{\mu}$, which differ in different samples. Thus, if we were to take 100 (hypothetical) samples, then in 95 of them, the true mean would lie in this interval.