EC203: Applied Econometrics
Estimators and sampling distributions

Dr. Tom Martin

University of Warwick

2019

Illustrative reading:

- Thomas: chapters 3, 4, 5 and 6.
- First year materials: $EC122/124$

Questions

Consider the following policy orientated questions:

- What is the likely outcome of the next general election?
- What is the mean and variance of child malnutrition in 2010 across all countries?
- What is the effect of micro-credit programs on child education?
- How does crime vary with capital punishment across US states?
- What is the effect of inequality on growth rates across countries in 2010?

Statistical inference in the empirical framework

In general, once you have: i) developed a question, ii) chosen a data set and method and iii) estimated an effect you should ask yourself:

- What is the direction of the effect? Does it make sense? How can you explain it? Can we give it a causal interpretation?
- Is the effect economically (practically) important? What is the size of the effect?
- Is the effect statistically significant in the **population of interest**?

The third question relates to the concept of population inference. That is, we want to use the **sample** to make statements about the **underlying population of interest**. At the heart of inference is the concept of sampling distributions.

Estimates and estimators

An **estimator** is a rule for calculating an **estimate** of some unknown population quantity. In this course we will assume the estimator is applied to a set of random observations: **a random sample**.

Given the estimator will change when we change the set of observations, implies **the estimator is a random variable**.

An **estimate** is a particular realisation calculated from a specific set of observations.

Estimates and estimators

In this lecture, we will use the mean estimator ($\bar{X}$) to get across the core ideas, where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

For example, while $\bar{X}$ is an estimator of the population mean μ, $\bar{x} = 42$ is an estimate of the population mean.

Population of interest

A good empirical analysis first requires that a population of interest is defined. In general, a population of interest is any well-defined group of:

- individuals, firms, cities, countries, ... etc.
- well-defined in the sense of a time and place: individuals in the UK between 2001 and 2010.

Suppose we are interested in estimating the mean income of the male UK working population, aged $25 - 60$ in the 2010.

Population versus random sample

Due to time and cost constraints, we cannot feasible survey all individuals in the population (of size N) and use this information to calculate the population mean (μ) directly.

Instead, we typically use a **random sample** of size n, where $n < N$, from the population. We then use this sample to estimate μ using the mean estimator $\bar{X}$.

Throughout this course we will assume we are working with random samples. This implies that:

- all individuals have an equal probability of being selected from the population of interest
- there is no dependence between them: if individual i is selected is does not affect the probability that individual j is selected.

Random sampling

Throughout this course we will assume we are working with random samples from the population of interest.

- Suppose our underlying population is represented by a random variable X with an associated pdf $f_X(x)$.
- From $f_X(x)$ we randomly make n draws.
- The resulting set $\{X_1, X_2, ..., X_n\}$ is known as random sample, where $Cov(X_i, X_j) = 0 \, \forall \, i \neq j$.[1]
- The upper cases $\{X_1, X_2, ..., X_n\}$ imply, prior to a sample being drawn, each draw X_i can theoretically take any value from the underlying population distribution $f(x)$.
- Once a random sample has been taken it is represented as $\{x_1, x_2, ..., x_n\}$.

[1] Alternatively stated, $\{X_1, X_2, ..., X_n\}$ is an independently identically distributed (iid) sample.

Sampling distributions

Conceptually, we could:

1. draw an infinite number of random samples from the population.
2. apply the mean estimator, $(\bar{X})$, to each sample.
3. plot each resulting mean estimate $\bar{x}_i$.
4. this would give a (probability) distribution of estimates: $\{\bar{x}_1, \bar{x}_2, ..., \bar{x}_m\}$.
5. the pdf of this distribution is known as the **sampling distribution**, written $f_{\bar{X}}(\bar{x})$.

Sampling distributions

Practically, we will only have one sample and one estimate, however, it is crucial to note:

1. That the estimate will be drawn from the sampling distribution.
2. Therefore, if we know the shape of the sampling distribution. For example, it could have a normal distribution.
3. We can make probabilistic statements about how close the estimate is likely to be to the population mean.

Sampling distributions

Given we will never know the population mean μ (otherwise, we wouldn't need to estimate it). The practical question becomes:

- How close is the estimate $\bar{x}$ likely to be to μ?
- Using the sampling distribution and some results from statistical theory we can answer this question probabilistically.
- To do this we need to know three main properties of the sampling distribution:
  1. Its mean.
  2. Its variance.
  3. Its shape.

Stata simulation: population $X \sim N(50, 100)$

The main principles that govern the shape of the mean estimator's sampling distribution:

- Question: How do we estimate the mean of a population?
- Suppose we know the underlying population $X \sim N(50, 100)$.
- To illustrate the idea, suppose we don't know that $\mu = 50$ and we want to estimate it using a random sample of 30 observations.
- What can we learn from this single sample about the population mean, μ?

Stata simulation: population $X \sim N(50, 100)$

What can we learn about the population from a single sample?

- From the first random sample of 30 observations: $\bar{x}_1 = ...$
- How close is the sample mean to the population mean?
- For a single sample we can never know.
- To learn about how close our sample mean is likely to be to the population mean we need to take more samples.
- From the second random sample of 30 observations: $\bar{x}_2 = ...$
- From the third random sample of 30 observations: $\bar{x}_3 = ...$
- ...
- The n^{th} random sample of 30 observations: $\bar{x}_n = ...$

In theory we should do this resampling an infinite number of times. Instead we will do it $10,000$ times.

Stata simulation: population $X \sim N(50, 100)$

Intuitively:

1. Which value do you think the sampling distribution will be centered about? (What is its expected value?)
2. What do you think the spread of the sampling distribution will be? (What is its variance?)
3. What do you think will be the shape of the sampling distribution? (Will it follow any specific type of distribution?)

Given our particular estimate comes from this sampling distribution, the answer to these questions gives us information on how likely any given estimate is likely to be to the population mean.

Stata simulation: population $X \sim N(50, 100)$

What do you notice about the mean of the sampling distribution for the mean estimator?

The mean of the sampling distribution is centered about the population mean.

- This result comes from the fact that the mean estimator is an unbiased estimator of the population mean: $E[\bar{X}] = \mu$.
- This result holds independently of the size of the sample, big or small.
- It is known as a finite sample property.

Stata simulation: population $X \sim N(50, 100)$

What do you notice about the variance of the sampling distribution for the mean estimator?

The variance (and the standard deviation) of the sampling distribution is smaller than the standard deviation of the population.

- The degree of spread is a measure of the precision of our estimator (or efficiency).
- The formula outlining this result (σ^2/n) holds for all sample sizes.
- Thus, it is a finite sample property.
- Note, however, holding all else constant the precision will increase with sample size n. Consider, the previous example, but now with 100 observations.

Stata simulation: population $X \sim N(50, 100)$

What do you notice about the shape of the sampling distribution for the mean estimator?

- The shape of the sampling distribution closely approximates a normal distribution.
- Again this is a finite sample property for normal populations.
    - Because the underlying population is normal this result holds for any sample size.

Stata simulation: population $X \sim N(50, 100)$

If the underlying population is $X \sim N(50, 100)$ and we have a sample of $n = 30$, then we know the shape of the mean estimator's sampling distribution is:

$$\bar{X} \sim N(50, 100/30)$$

More generally, if the underlying population is $X \sim N(\mu, \sigma^2)$ the shape of the mean sampling distribution is:

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Showing the general result for $X \sim N(\mu, \sigma^2)$

The underlying popultion is as before: $X \sim N(\mu, \sigma^2)$. Let $\{X_1, X_2, ..., X_n\}$ be a random sample of size n.
The mean estimator is $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$. The expected value of $\bar{X}$ is:

$$\begin{aligned} E[\bar{X}] &= E[\frac{1}{n}(X_1 + X_2 + ... + X_n)] \\ &= \frac{1}{n}E[X_1] + \frac{1}{n}E[X_2] + ... + \frac{1}{n}E[X_n] \\ &= \frac{n\mu}{n} \\ &= \mu \end{aligned}$$

- This shows $\bar{X}$ is an unbiased estimator of μ.
- Note, in this calculation we made no assumption about the **shape** or **size** of the underlying population.

Showing the general result for $X \sim N(\mu, \sigma^2)$

The variance of $\bar{X}$ is:

$$V(\bar{X}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}V(\sum_{i=1}^{n} X_i)$$

$$= \frac{1}{n^2}(V[X_1] + V[X_2] + ... + V[X_n] +$$

$$2cov(X_1, X_2) + 2cov(X_1, X_3) + ... + 2cov(X_{n-1,X_n})$$

$$= \frac{1}{n^2}(\sigma^2 + \sigma^2 + ... + \sigma^2)$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

- Showing explicitly, as the sample size grows the variance falls.
- Note, in this calculation we made no assumption about the **shape** or **size** of the underlying population.

Showing the general result $X \sim N(\mu, \sigma^2)$

The shape of $\bar{X}$:

1. Each X_i is normally distributed: $X_i \sim N(\mu, \sigma^2)$
2. We know that a weighted sum of normally distributed random variables is itself normally distributed.
3. The mean $\bar{X} = \sum_{i=1}^{n} X_i$ is a sum of weighted random variables each with a normal distribution.

Therefore, the mean estimator is normally distributed: $\bar{X} \sim N(\mu, \sigma^2/n)$.

- Note, this result relies, crucially, on the assumption about the **shape** (but not the **size**) of the underlying population.

Stata simulation: population $X \sim bin(1, 0.5)$

What do you think the shape of the mean sampling distribution will be if the underlying population is non-normal, $X \sim bin(1, 0.5)$, for example?

It depends on the size of the sample:

- $n = 1$
- $n = 2$
- $n = 10$
- $n = 30$
- $n = 120$ ...

The Central Limit Theorem (CLT)

What would happen if we made no assumptions about the shape of the underlying population?

Let $X_1, X_2, ..., X_n$ be an n independent random variables from ANY common population of mean μ and (finite) variance σ^2. Then define $X = X_1 + X_2 + ... + X_n$. The expected value does not depend on shape of X_i:

$$\begin{aligned} E[X] =& E[X_1 + X_2 + ... + X_n] \\ =& E[X_1] + E[X_2] + ... + E[X_n] \\ =& n\mu \end{aligned}$$

The variance does not depend on shape of X_i:

$$\begin{aligned} V[X] =& V[X_1 + X_2 + ... + X_n] \\ =& V[X_1] + V[X_2] + ... + V[X_n] \\ =& n\sigma^2 \end{aligned}$$

The Central Limit Theorem (CLT)

However, the shape of X does depend on the shape of the underlying population, X_i.

However, importantly, the CLT means that as the number of terms in the summation becomes large ($n > 30$), no matter what the common shape of the underlying population X_i the summation $X = X_1 + X_2 + ... + X_n$ is normally distributed:

$$X \sim N(.)$$

Note, the result gets stronger as the sample size grows (it asymptotically converges to the normal distribution). It is a large sample property.

The Central Limit Theorem (CLT): mean estimator

Let $X_1, X_2, ..., X_n$ be n independent random variables from ANY common population of mean μ and (finite) variance σ^2. The mean estimator is $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$. Then:

The expected value of $\bar{X}$ is:

$$
\begin{aligned}
E[\bar{X}] &= E[\frac{1}{n}(X_1 + X_2 + ... + X_n)] \\
&= \frac{1}{n}E[X_1] + \frac{1}{n}E[X_2] + ... + \frac{1}{n}E[X_n] \\
&= \frac{n\mu}{n} \\
&= \mu
\end{aligned}
$$

- This shows $\bar{X}$ is an unbiased estimator of μ.
- Note, in this calculation we made no assumption about the shape of the population distribution.

The Central Limit Theorem (CLT): mean estimator

The variance of $\bar{X}$ is:

$$
\begin{aligned}
V(\bar{X}) &= V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2} V(\sum_{i=1}^{n} X_i) \\
&= \frac{1}{n^2}(V[X_1] + V[X_2] + ... + V[X_n] + \\
& 2cov(X_1, X_2) + 2cov(X_1, X_3) + ... + 2cov(X_{n-1,n}) \\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + ... + \sigma^2) \\
&= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned}
$$

- Showing explicitly, as the sample size grows the variance falls.
- Note, in this calculation we made no assumption about the shape of the population distribution.

The Central Limit Theorem (CLT): mean estimator

We can use the CLT theorem to give us the shape of the $\bar{X}$ sampling distribution. The CLT tells us that no matter the underlying population:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

The sampling distribution of the mean estimator asymptotically converges to the normal distribution.

Sampling distribution of $\bar{X}$

The above examples have shown explicitly how the sampling distribution of the mean estimator depends on three main factors:

1. The sample size: n
2. Knowledge of the underlying population distribution, $X_i \sim D(.)$:, in particular: its mean, variance and shape
3. The estimator chosen. The shape, bias and spread (precision) of the sampling distribution will be different for different estimators. So far we have only considered the mean estimator.

Estimators and their sampling distributions

Let us consider a few variations. Whether we can specify the shape of the mean estimator's sampling distribution will depend on what we know (or assume) about the underlying populations: i) distribution and ii) variance σ^2, as well as our sample size n. The following summarises whether we can define (**yes** or **no**) the shape of the sampling distribution:

1. $X \sim N(.)$ and σ^2 is known, small n (**yes**) / large n (**yes**)
2. $X \sim N(.)$ and σ^2 is unknown , small n (**yes**) / large n (**yes**)
3. $X \sim D(.)$ where $D(.)$ is unknown but σ^2 is known, small n (**no**) / large n (**yes**)
4. $X \sim D(.)$ where $D(.)$ is unknown and σ^2 are unknown, small n (**no**) / large n (**yes**)

The mean estimator's sampling distribution

The shape of sampling distribution when the sample size n is small ($n < 30$):

1. $X \sim N(\mu, \sigma^2)$ with σ^2 known $\implies Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
2. $X \sim N(\mu, \sigma^2)$ with σ^2 unknown $\implies t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
3. $X \sim D(.)$ and $D(.)$ is unknown means we don't have enough information to generate shape of sampling distribution.
4. $X \sim D(.)$ and $D(.)$ is unknown means we don't have enough information to generate shape of sampling distribution.

The mean estimator's sampling distribution

Explaining result number 2. Recall, if $Z \sim N(0,1)$ and $X \sim \chi^2_n$, then the random variable defined as,

$$T = \frac{Z}{\sqrt{X/n}}$$

has a t-distribution with n degrees of freedom, written $T \sim t_n$.

- Here, the mean estimator has a standard normal distribution and the estimator for the variance has a χ^2_{n-1} sampling distribution.
- Intuitively, the use of the t-dist takes account of the added variability that comes from using an estimator for the population variance.

The mean estimator's sampling distribution

A quick note on the difference between the standard deviation and the standard error:

- The value $\sigma/\sqrt{n}$ is the standard deviation of the mean estimator.
- The value $S/\sqrt{n}$ is the standard error of the mean estimator.

The different terms are used to distinguish whether you have used: i) the actual population standard deviation (σ) to estimate the mean or ii) an estimate of the population standard deviation S.

The mean estimator's sampling distribution

The shape of sampling distribution when the sample size n is large:

1. $X \sim N(\mu, \sigma^2)$ with σ^2 known $\implies Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
2. $X \sim N(\mu, \sigma^2)$ with σ^2 unknown $\implies Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$
3. $X \sim ?(\mu_X, \sigma^2)$ with σ^2 known $\implies Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
4. $X \sim ?(\mu_X, \sigma^2)$ with σ^2 unknown $\implies Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$

- Result 2 uses the standard normal rather than the t-distribution, recognising the t-distribution tends to the standard normal as n increases.
- Result 2 and 3 results invoke the CLT.

Sampling distribution: summary

- The population parameter μ can be estimated using the estimator $\bar{X}$.
- $\bar{X}$ has an associated sampling distribution.
- The shape of which depends on: i) the parameter being estimated, ii) the shape of underlying population distribution and iii) the size of the sample.

Why do we care about sampling distributions?

1. To compare between estimators. For example, what are the statistical reasons for using the estimator $\bar{X}$?
2. For population inference.

Comparing between estimators

Why did we choose the mean estimator $\frac{1}{n}\sum_{i=1}^{n} X_i$? Why did we not use some other estimator?

More generally, for any given population parameter θ there are an infinite number of estimators available to estimate it. How do you choose between them? Suppose, we have a generic estimator which is a function of the sample, $W = h(X_1, ..., X_n)$. The following criteria are used:

- Bias: $Bias(W) = E[W] - \theta$. An estimator is said to be unbiased if $E[W] = \theta$.
- Efficiency: given two estimators W_1 and W_2, W_1 is efficient relative to W_2 if $V(W_1) \leq V(W_2)$.
- Mean squared error: $MSE(W) = E[(W - \theta)^2] = V(W) + Bias(W)^2$.

Comparing between estimators

To illustrate, consider the following example:

- We draw a random sample of size n from a population with mean μ and variance σ^2.
- Suppose we want to select from the following three linear estimators:[2]
  1. $W_1 = X_3$
  2. $W_2 = \frac{1}{n}\sum_{i=1}^{n} X_i$
  3. $W_3 = \frac{1}{2n}\sum_{i=1}^{n} X_i$

---

[2] Linear in the sense that they are all a weighted averages of the form: $\sum_{i=1}^{n} w_i X_i$. Note all estimators in this course will be linear weighted averages.

Comparing between estimators

The bias of W_1, W_2 and W_3:

- $E[W_1] = E[X_3] = \mu$
- $E[W_2] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu$
- $E[W_3] = E[\frac{1}{2n}\sum_{i=1}^{n} X_i] = \frac{\mu}{2}$
- Thus, W_1 and W_2 are unbiased.

In this course we are only concerned with unbiased estimators.
So we drop W_3.

Comparing between estimators

The variance of W_1 and W_2:

- $V[W_1] = V(X_3) = \sigma^2$
- $V[W_2] = V[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\mu}{n^2} = \frac{\sigma^2}{n}$
- Given $V[W_2] < V[W_1]$, using the minimum variance criteria among unbiased estimators, we would choose W_2.

Note, it is possible to prove that W_2 is the best linear unbiased estimator (BLUE). That is, out of all unbiased linear estimators, W_2 is the most efficient.[3]

[3] Note,
$V[W_3] = V[\frac{1}{2n}\sum_{i=1}^{n} X_i] = \frac{1}{4n^2}\sum_{i=1}^{n} V[X_i] = \frac{1}{4n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\mu}{4n^2} = \frac{\sigma^2}{4n}$.
Further, if we didn't restrict our attention to unbiased estimators, $MSE(W) = E[(W-\theta)^2] = V(W) + Bias(W)^2$ is another common criteria.

Stata simulation: comparing between estimators

Simulation to show above results:

- Bias: W_1 and W_2 sampling distributions are centered around the population mean (unbiased).
- Variance: the variance of W_2 is much lower than the variance of W_2.
- (Note: W_3 has the minimum variance but it is biased as it is centered about $\mu/2$.)