Instituto Superior de Economia e Gestão
Universidade de Lisboa
Mestrado em Econometria Aplicada e Previsão
(MSc in Applied Econometrics and Forecasting)
Econometrics
Paulo M.D.C. Parente
Notes on Statistical inference

Populations and parameters

Population: A group of entities of interest.

We are usually interested in knowing the value of a *parameter* that measures an aspect of the population.

Example: Mean earnings of all women that recently graduated from college. Here the population is all women that recently graduated from college and the parameter is mean earnings.

Suppose we wish to estimate its unknown population mean $\mu_X = E(X)$. Usually the population size is extremely large, so it is not practical to collect information on all individuals of the population. So how can we estimate this parameter?

Sampling and estimators

We obtain a sample of n independent observations $(X_1, ..., X_n)$ drawn from the population at random (known as *random sample*). Each member of the population has equal chance of being included in the sample.

Prior to the actual drawings from the populations, the X_i are random quantities. We know that they will be generated randomly from the distribution for X, but we *do not know their values* in advance.

Once we have taken the sample we will have a set of numbers $(x_1, ..., x_n)$. This is called a *realization.* The lower case is to emphasize that these are numbers, not variables.

Having obtained a sample of n observations $(X_1, ..., X_n)$, we plan to use them with a mathematical formula to estimate the unknown population mean μ_X. This formula is known as an *estimator.* We are going to use the analogy principle to propose an estimator for μ_X.

Analogy principle: As a population parameter is a feature of the population, to estimate it, use the corresponding feature of the sample.

Thus as μ_X is the mean of the population, a natural estimator for μ_X is the sample mean

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

An estimator is a random variable because it depends on the random quantities $(X_1, ..., X_n)$.

The actual number that we obtain, given the realization $(x_1, ..., x_n)$, is known as our *estimate*:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Example: City Unemployment Rates: Suppose that we obtain the following sample of unemployment rates for 10 cities in the United States.

City	1	2	3	4	5	6	7	8	9	10
Unemployment Rate	5.1	6.4	9.2	4.1	7.5	8.3	2.6	3.5	5.8	7.5

The estimate for the average city unemployment rate is

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 6.0$$

Properties of Estimators

The estimator suggested for the mean of the population μ_X was

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

Why should we use rather than some other estimator?

Alternative estimators:

- X_1 (the first observation).
- Maybe unequal weights – not simple average.

Which criteria can we use for choosing an estimator?

Criteria for an estimator

Unbiasedeness

Let us denote by θ an unknown parameter that we are interested in estimating (for example the mean of the population μ_X) and $\hat{\theta}$ its estimator (for instance, $\overline{X}$).

Definition 1 *An estimator is said to be an* unbiased *estimator of θ if*

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

If this equality does not hold, the estimator is said to be biased *and the bias is $E(\hat{\theta}) - \theta$.*

Example: $\overline{X}$ is an unbiased estimator of μ_X;

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

is a biased estimator of the population variance σ_X^2 with bias $-\sigma_X^2/n$.

To see this notice that

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu_X\right)^2 - \left(\overline{X} - \mu_X\right)^2$$

Taking expectations we have

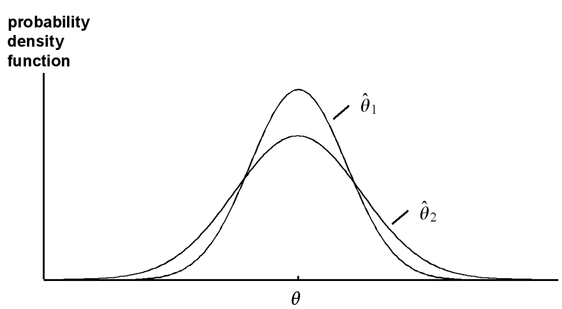
$$\begin{aligned} E[\tilde{S}^2] &= \frac{1}{n}\sum_{i=1}^{n} Var(X_i) - Var(\overline{X}) \\ &= \sigma_X^2(1 - 1/n). \end{aligned}$$

Efficiency

Definition 2 *Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be unbiased estimators of the parameter θ. If*

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

then we say that $\hat{\theta}_1$ is more efficient *than $\hat{\theta}_2$.*



How can compare estimators that are not unbiased?
A possible solution is to compute the mean squared error (MSE)

$$MSE\left(\hat{\theta}\right) = E[\left(\hat{\theta} - \theta\right)^2]$$

It is possible to show that

$$MSE\left(\hat{\theta}\right) = \underbrace{Var\left(\hat{\theta}\right)}_{\text{Variance}} + \underbrace{[E(\hat{\theta}) - \theta]^2}_{\text{Bias}^2}.$$

Large Sample properties of estimators

In some cases an estimator does not have the desirable properties when the sample size is small, but when the sample is large some of the desirable properties might hold. As we let the sample size go to infinite we call these properties *asymptotic properties*.

Consistency is a minimal requirement for an estimator:

> *"If you can't get it right as n goes to infinity, you shouldn't be in this business."*
> C.W.Granger, 2003, Nobel Prize Winner in Economics

Notice that there are estimators that are unbiased but are not consistent.

Example: if we have a sample $(Y_1, ..., Y_n)$ and we would like to estimate the population mean $\mu = E(Y)$. We consider the first observation Y_1 as an estimator for the population mean. Then, $E(Y_1) = \mu$ (and therefore it is an unbiased estimator). However, it can be shown that Y_1 is not a consistent estimator of μ.

Definition 3 *An estimator $\hat{\theta}$ is said to be a* consistent *estimator of θ if*

$$\lim_{n\to\infty} \mathcal{P}(\theta - \varepsilon < \hat{\theta} < \theta + \varepsilon) = 1,$$

for all $\varepsilon > 0$. This property is often expressed as plim $\hat{\theta} = \theta$ or $\hat{\theta} \xrightarrow{p} \theta$. If $\hat{\theta}$ is not consistent for θ, we say that it is inconsistent*.*

Proposition 4 *If*

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

and

$$\lim_{n\to\infty} Var\left(\hat{\theta}\right) = 0,$$

then plim $\hat{\theta} = \theta$.

Properties of $\overline{X}$:

1. $E\left(\overline{X}\right) = \mu_X$.
2. $Var\left(\overline{X}\right) = Var\left(X\right)/n$.
3. plim $\overline{X} = \mu_X$. (*Law of Large Numbers*)

Remark: An estimator can be biased and consistent.
Example: $\widetilde{X} = \frac{1}{n+1}\sum_{i=1}^{n} X_i$.
Remarks on Consistency:

In practice we deal with finite samples, not infinite ones. So why should we be interested in whether an estimator is consistent?

1. Sometimes it is impossible to find an estimator that is unbiased for small samples. If you can find one that is at least consistent, that may be better than having no estimate at all.
2. Often we are unable to say anything at all about the expectation of an estimator. The expected value can be applied in relatively simple contexts.

Convergence in distribution

If a random variable X has a normal distribution, its sample mean will also have a normal distribution. This fact is useful to test hypothesis on the mean of the population μ_X. However, what happens if X is not normally distributed?

Assume that T_n is a stochastic sequence (example: $T_n = \overline{X}$) that has cumulative distribution function $F_n(x)$.

Definition 5 *If there is a fixed cumulative distribution function $G(x)$ such that*

$$\lim_{n\to\infty} F_n(x) = G(x)$$

for all x at which $G(x)$ is continuous we say the T_n converges in distribution *to $G(.)$ and write $T_n \overset{D}{\to} G(.)$ where the symbol $\overset{D}{\to}$ reads "convergence in distribution".*

The following Theorem plays a central role in statistics.

Central Limit Theorem: If the X_i in the sample are all drawn independently from the same distribution (the distribution of X), and provided that this distribution has finite variance $\sigma_X^2 > 0$, then distribution of

$$Z_n = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X}$$

will converge in distribution to a standard normal distribution as n tends to infinity and we write $Z_n \overset{D}{\to} N(0,1)$.

Confidence Intervals

A point estimate by itself does not provide enough information about how close the estimate is likely to be to the population parameter.

Example: Suppose that a researcher would like to know the mean operating life of light bulbs and on the basis of a random sample of 10 observations finds that the mean operating life in the sample is 300h.

- How are we to know whether or not this is close to the mean operating life in the population of light bulbs?
- Since we do not know the population value we cannot know how close an estimate is for a particular sample.
- The questions are partially answered by constructing a *confidence interval.*

Confidence Intervals for Normal populations

The concept of *confidence interval* will be illustrated with an example:

Suppose that the population has a $N(\mu, 1)$ distribution and let $(X_1, ..., X_n)$ be a a random sample from this population. In this case the sample average $\overline{X}$ has normal distribution with mean μ and variance $1/n$: $\overline{X} \sim N(\mu, 1/n)$. It follows that

$$Z = \frac{\overline{X} - \mu}{1/\sqrt{n}} \sim N(0,1)$$

and consequently

$$\mathcal{P}\left(-1.96 < \frac{\overline{X} - \mu}{1/\sqrt{n}} < 1.96\right) = 0.95,$$

which is equivalent to

$$\mathcal{P}\left(\overline{X} - 1.96/\sqrt{n} < \mu < \overline{X} + 1.96/\sqrt{n}\right) = 0.95.$$

This equation tells us that the probability that the random interval $(\overline{X} - 1.96/\sqrt{n}, \overline{X} + 1.96/\sqrt{n})$ contains the population mean is 95%. This information allows us to construct a confidence interval which can be obtained by replacing the estimator by the estimate $(\overline{x} - 1.96/\sqrt{n}, \overline{x} + 1.96/\sqrt{n})$.

Interpretation: If independent samples are taken repeatedly from the same population, and the confidence interval is calculated for each sample in the manner described above, then 95% of the intervals will include the true parameter μ. Our confidence interval is only one of these.

Confidence interval when the variance is known.

Suppose now that the population has a $N\left(\mu, \sigma^2\right)$ distribution and let $(X_1, ..., X_n)$ be a a random sample from this population.

In this case the sample average $\overline{X}$ has normal distribution with mean μ and variance σ^2/n : $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$. It follows that

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Consequently

$$\mathcal{P}\left(-1.96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95,$$

which is equivalent to

$$\mathcal{P}\left(\overline{X} - 1.96\sigma/\sqrt{n} < \mu < \overline{X} + 1.96\sigma/\sqrt{n}\right) = 0.95.$$

In this case the confidence interval is given by $(\overline{x} - 1.96\sigma/\sqrt{n}, \overline{x} + 1.96\sigma/\sqrt{n})$.

One can replace σ by an estimator the *sampling standard deviation* S where .

$$S^2 = \frac{1}{n-1}\sum\nolimits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

is the *Sampling Variance*.

However

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has no longer the standard normal distribution.

T has the *t-Student* distribution with *n-1* degrees of freedom, denoted as $t(n-1)$. Let us choose the value c such that

$$\mathcal{P}(T < c) = 0.975,$$

which is equivalent to choosing c such that $\mathcal{P}(-c < T < c) = 0.95$ (this can be done using the Tables for the $t(n-1)$ distribution in the book) which is equivalent to

$$\mathcal{P}\left(\overline{X} - cS/\sqrt{n} < \mu < \overline{X} + cS/\sqrt{n}\right) = 0.95.$$

In this case to obtain the confidence intervals we replace $\overline{X}$ and S by their estimates and obtain $(\overline{x} - cs/\sqrt{n}, \overline{x} + cs/\sqrt{n})$.

More generally the $100(1-\alpha)\%$ confidence interval is given by $(\overline{x} - c_{\alpha/2}s/\sqrt{n}, \overline{x} + c_{\alpha/2}s/\sqrt{n})$, where $c_{\alpha/2}$ is such that

$$\mathcal{P}\left(T < c_{\alpha/2}\right) = 1 - \alpha/2.$$

Confidence Intervals for Nonnormal populations

In some applications the distribution of the population is not normal. (**Example:** $X \sim Bernoulli(p)$). How can we construct a confidence interval in this case?

If *the sample size is large* one can use the following result given by the Central Limit Theorem

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X} \xrightarrow{D} N(0,1).$$

Remark: One can replace σ by an estimator the *sampling standard deviation* $S = \sqrt{\frac{1}{n-1}\sum\nolimits_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ and use statistic

$$T = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{S} \xrightarrow{D} N(0,1)$$

Remark: If $X \sim N\left(\mu, \sigma^2\right)$ recall that $\mu_X = \mu$ and $T \sim t(n-1)$. This is a result valid only for normal populations and *finite* n. For *large* n the $t(n-1)$ distribution is close to $N(0,1)$ distribution.

Thus we construct confidence intervals in the following way:

- T has the $N(0,1)$ distribution for *large* n. Choose the value $z_{\alpha/2}$ such that

  $$\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2,$$

  where Φ is the cumulative distribution function of the standard normal distribution (this can be done using the Tables for the $N(0,1)$ distribution in the book).

- For *large* n

$$\mathcal{P}\left(-z_{\alpha/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < z_{\alpha/2}\right) \simeq 1-\alpha,$$

which is equivalent to

$$\mathcal{P}\left(\overline{X} - z_{\alpha/2}S/\sqrt{n} < \mu < \overline{X} + z_{\alpha/2}S/\sqrt{n}\right) \simeq 1-\alpha.$$

- In this case to obtain the confidence intervals we replace $\overline{X}$ and S by their estimates and obtain $(\overline{x} - z_{\alpha/2}s/\sqrt{n}, \overline{x} + z_{\alpha/2}s/\sqrt{n})$.

Hypothesis testing

Example: Suppose that there are 2 candidates in an election, A and B. The Candidate A is reported to have received 42% and candidate B 58% of the votes. Candidate A thinks that the election was rigged and hires a consultancy agency to investigate it.

The consultancy agency obtains a sample of 100 voters and in the sample 53% voted for candidate A. Should candidate A conclude that the election was a fraud? Notice that there is some uncertainty associate with the sample estimate of 53%. How strong is the sample evidence against the official reported percentage of 42%?

One way to test this is to proceed is to use a *hypothesis test*.

Hypothesis testing consists in the following steps:

1. State the null hypothesis H_0;
2. State the alternative hypothesis H_1;
3. Choose the significance level;
4. Select a statistical test and compute the observed test statistic;
5. Find the critical value of the test statistic. Compare the critical value with the observed test statistic and decide to reject or not reject H_0.

We can commit two types of errors:

- *Type I Error* - We reject the null hypothesis when it is true.
- *Type II Error* - We accept the null hypothesis when it is false.

Hypothesis test

	Actual Situation	
Decision	H_0 is true	H_0 is false
Accept H_0	Correct Decision $(1-\alpha)$	Type II Error (β)
Reject H_0	Type I Error (α)	Correct Decision $(1-\beta)$

The *significance level* (or simply the level) is the probability of Type I error. Symbolically,

$$\mathcal{P}(\text{Reject } H_0|H_0) = \alpha.$$

Once we choose α we would like to to minimize the probability of Type II error. Equivalently, defining the *power of a test* as the probability of rejecting H_0 when it is false, we would like to maximize $(1-\beta)$. Generally power declines as the significance level declines.

Testing hypothesis about the mean in a normal population.

1- State the null hypothesis

$$H_0 : \mu = \mu_0.$$

2- State the alternative hypothesis

$$H_1 : \mu > \mu_0$$

(one sided alternative) or

$$H_1 : \mu < \mu_0$$

(one sided alternative) or

$$H_1 : \mu \neq \mu_0$$

(two sided alternative).
3- Choose the significance level: Usually we choose α as 5%, 1% or 10%.
4- Select a statistical test and compute the observed test statistic: If the population is normal, we can consider the statistic

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t\,(n-1)\,.$$

Compute its actual value in the sample $t^{act} = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$.
5- Find the critical value of the test statistic. Compare the critical value with the observed test statistic and decide to reject or not reject H_0.

- If the alternative hypothesis is $H_1 : \mu > \mu_0$ a simple procedure to decide is the following: Choose the critical value c_α such that

  $$\mathcal{P}\,(T < c_\alpha) = 1 - \alpha.$$

  Rejection rule is: reject H_0 if $t^{act} > c_\alpha$:

- If the alternative hypothesis is $H_1 : \mu < \mu_0$, choose the critical value c_α such that

  $$\mathcal{P}\,(T < c_\alpha) = 1 - \alpha.$$

  the *rejection rule* is**:** reject H_0 if $t^{act} < -c_\alpha$:

- If the alternative hypothesis is $H_1 : \mu \neq \mu_0$ we choose the critical level $c_{1-\alpha/2}$ such that

  $$\mathcal{P}\left(T < c_{\alpha/2}\right) = 1 - \alpha/2$$

  Rejection rule: reject H_0 if $|t^{act}| > c_{\alpha/2}.$:

We can use also the concept of p-value to define rejection rules for hypotheses.
P-value: The p-value is the largest significance level at which we could carry out a test and still fail to reject the null hypothesis. Mathematically:

- If the alternative hypothesis is $H_1 : \mu > \mu_0$,

  $$p - value = \mathcal{P}\left(T > t^{act}\right).$$

- If the alternative hypothesis is $H_1 : \mu < \mu_0$,

  $$p - value = \mathcal{P}\left(T < t^{act}\right).$$

- If the alternative hypothesis is $H_1 : \mu \neq \mu_0$

  $$p - value = \mathcal{P}\left(|T| > |t^{act}|\right).$$

Rejection rule:If $p - value < \alpha$ we reject the null hypothesis.

Remark: As in the case of the confidence intervals if the population is not normal the if *the sample size is large* one can invoke the Central Limit Theorem. The mechanism of hypothesis testing for population means is similar to the one described before. To test hypothesis one can use the statistic

$$Z = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{\sigma_X} \xrightarrow{D} N(0,1)$$

or

$$T = \frac{\sqrt{n}\left(\overline{X} - \mu_X\right)}{S} \xrightarrow{D} N(0,1)$$

The critical values should be obtained using the tables of the standard normal distribution.

The Relationship between Confidence Intervals and Hypothesis testing

Confidence Intervals can be used for hypothesis testing for two-sided alternatives ($H_1 : \mu \neq \mu_0$).

Suppose we construct a $100(1-\alpha)\%$ confidence interval for μ. Then if the hypothesized value of μ under H_0 is not in the confidence interval, then $H_0 : \mu_X = \mu_0$ is rejected against $H_1 : \mu \neq \mu_0$ at $\alpha\%$ level.