Eco 2740: Economic Statistics. Estimation and Confidence Intervals (Chapter 10)

Estimation by sample analog

Estimation: approximate the population parameter using sample data.

Sample analog (method of moments) approach:

Estimate population parameter by analogous sample statistic.

- estimate the population mean μ by sample mean $\overline{x}$
- Estimate population variance s^2 by sample variance S^2
- Estimate population covariance by sample covariance

Maximum Likelihood Method: An alternative, more complicated estimation method for a more advanced course.

Qualities of Estimators: Unbiasedness

An ***unbiased estimator*** of a population parameter is an estimator whose expected value is equal to that parameter.

E.g. the sample mean $\overline{X}$ is an ***unbiased*** estimator of the population mean μ , since:

$$E(\overline{X}) = \mu$$

Qualities of Estimators: Consistency

An estimator is said to be ***consistent*** if the difference between the estimator and the parameter grows smaller as the sample size grows larger.

E.g. $\bar{X}$ is a ***consistent*** estimator of μ because:

$E(X) = \mu$ and

$V(\bar{X})$ is σ^2/n

That is, as n grows larger, the variance of $\bar{X}$ grows smaller and the distribution "collapses" on μ.

ON BLACKBOARD: ILLUSTRATE USING PDF

Qualities of Estimators: Relative Efficiency

If there are two unbiased estimators of a parameter, the one whose variance is smaller is said to be ***relatively efficient***.

E.g. both the sample mean of the entire sample (full sample mean) and the sample mean using only the first half of the sample (half sample mean) are both unbiased estimators of the population mean. However, the full sample mean has a lower variance than the half sample mean, so it is ***relatively efficient*** when compared to the half sample mean.

E.g. If the distribution is symmetric, then both the mean and median estimate the population mean. If the data are normally distributed the mean is a more efficient estimator.

Qualities of Estimators: Robustness

- There are two meanings of a more robust estimator:
  - ➢The estimator is less sensitive to outliers OR
  - ➢The estimator is less sensitive to violations of the assumptions made by the statistician

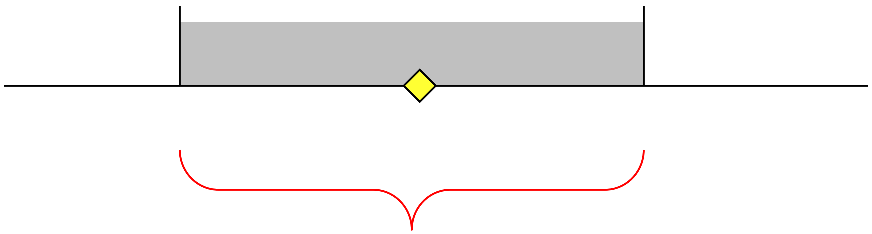
E.g. If the distribution is symmetric, both the mean and median estimate the population mean. However, the median is more robust to:

(a) outliers and

(b) violations of the assumption that the data are normally distributed.

Confidence Interval

A ***confidence interval (or interval estimator)*** is an interval which includes the true unknown population parameter with a pre-specified level of confidence, i.e. with a pre-specified probability.



That is we say (with some ___% certainty) that the population parameter of interest is between some lower and upper bounds.

Point Estimates & Confidence Intervals

For example, suppose we want to estimate the mean summer income of a class of business students. For n = 25 students, $\bar{x}$ is calculated to be 400 $/week.

point estimate

confidence interval

An alternative statement is:

With 95 percent confidence, the mean income is ***between*** 380 and 420 $/week.

Confidence Interval

- A **1-α Confidence Interval** is an interval designed to contain the true population parameter with probability 1-α

  P(Confidence Interval contains Pop. Parameter) = 1-α

- 1-α is referred to as the **confidence level**
- A common choice is 1-α = 0.95, i.e. α= 0.05
- In the next few slides we will design a confidence interval for the population mean m

Standardizing the Sample Mean

- Recall that if $X \sim \mathrm{N}\left(E\left[X\right], VAR(X)\right)$
- Then we standardize X by:

$$Z = \frac{X - E\left[X\right]}{\sqrt{VAR(X)}} \sim N\left(0,1\right)$$

- For large n, we know from the last set of slides that

$$\overline{X} \sim \text{Approx } N\left(m, \frac{s^2}{n}\right)$$

- So, noting that $VAR\left(\overline{X}\right) = s^2 / n$ we standardize it by:

$$Z = \frac{\overline{X} - m}{s / \sqrt{n}} \sim \text{Approx } N\left(0,1\right)$$

Confidence intervals when S^2 is known

$$Z = \frac{\overline{X} - m}{s / \sqrt{n}}$$

$$1 - a = P\left(-z_{a/2} < Z < z_{a/2}\right)$$

$$= P\left(-z_{a/2} < -Z < z_{a/2}\right)$$

$$= P\left(-z_{a/2} < \frac{m - \overline{X}}{s / \sqrt{n}} < z_{a/2}\right)$$

$$= P\left(\overline{X} - z_{a/2}\, s / \sqrt{n} < m < \overline{X} + z_{a/2}\, s / \sqrt{n}\right)$$

So, the probability that the confidence interval $\overline{X} \pm Z_{a/2}\, s / \sqrt{n}$ contains the true population value m is equal to 1-α

EXPLAIN ON BLACKBOARD

Confidence interval for μ when σ is known…

The interval can also be expressed as

$$\text{Lower confidence limit} = \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$\text{Upper confidence limit} = \left(\bar{x} + z_{a/2} \frac{s}{\sqrt{n}} \right)$$

The probability 1 – α is the confidence level, which is a measure of how frequently the interval will actually include μ.

Notation & Typical Confidence Interval

- Standard error: $s/\sqrt{n}$ if s known (standard deviation of the estimator)
- Critical value: $z_{a/2}$ if s known
- Point estimate: $\bar{x}$

Confidence Interval	=	Point Estimate	±	Critical Value	´	Standard Error
CI	=	$\bar{x}$	±	$Z_{a/2}$	´	$s/\sqrt{n}$

Example 10.1…

As a manager at Doll Computer Company, you need a forecast of demand to lower inventory costs.

Suppose that you (somehow) know that demand is normally distributed with population standard deviation of 75.

Using 25 recorded observations of demand you calculate a sample mean of 370.16.

Find the 95% confidence interval for the (population) mean demand.

Example 10.1…

We want a 95% confidence interval for ***mean*** (population) demand

IDENTIFY

Thus, the parameter to be estimated is the population mean: μ

And so our confidence interval estimator will be:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Example 10.1… COMPUTE

In order to use our confidence interval estimator, we need the following pieces of data:

$\bar{x}$	370.16
$z_{\alpha/2}$	1.96
σ	75
n	25

Calculated from the data… ($\bar{x}$)

$1-\alpha = .95, \ \therefore \alpha/2 = .025$

$so \ \ z_{\alpha/2} = z_{.025} = 1.96$

Given (σ, n)

therefore: $$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm z_{.025}\frac{75}{\sqrt{25}} = 370.16 \pm 1.96\frac{75}{\sqrt{25}} = 370.16 \pm 29.40$$

The **lower** and **upper** confidence limits are 340.76 and 399.56.

Example 10.1…

INTERPRET

Our confidence interval for the mean demand falls between 340.76 and 399.56, and this type of interval is correct 95% of the time. That also means that 5% of the time the interval will be incorrect. The manager can use this as input in developing an inventory policy

Interpreting the confidence Interval Estimator

Some people erroneously interpret the confidence interval estimate in Example 10.1 to mean that there is a 95% probability that the population mean lies between 340.76 and 399.56.

This interpretation is wrong because it implies that the population mean is a random variable about which we can make probability statements.

In fact, the population mean is a fixed but unknown quantity. Consequently, we cannot interpret the confidence interval estimate of μ as a probability statement about μ.

Interpreting the confidence Interval Estimator

To interpret the confidence interval estimate properly, we must remember that it is the bounds of the confidence interval $\bar{X} \pm z_{a/2} s/\sqrt{n}$ itself that is random. This is because the sample mean is random. The population mean m is constant, despite being unknown.

The confidence interval estimator is also a probability statement about the sample mean …

Interpreting the confidence Interval Estimator

It states that there is 1 - α probability that the sample mean will be equal to a value such that the interval

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

to

$$\left(\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

will include the population mean

$$m$$

If we used 100 different samples to construct 100 confidence intervals, we would expect (1 – α)100 of them to include the true value μ .

Interval Width…

***A wide interval provides little information*.**

For example, suppose we estimate with 95% confidence that an accountant's average starting salary is between $15,000 and $100,000.

Contrast this with: a 95% confidence interval estimate of starting salaries between $42,000 and $45,000.

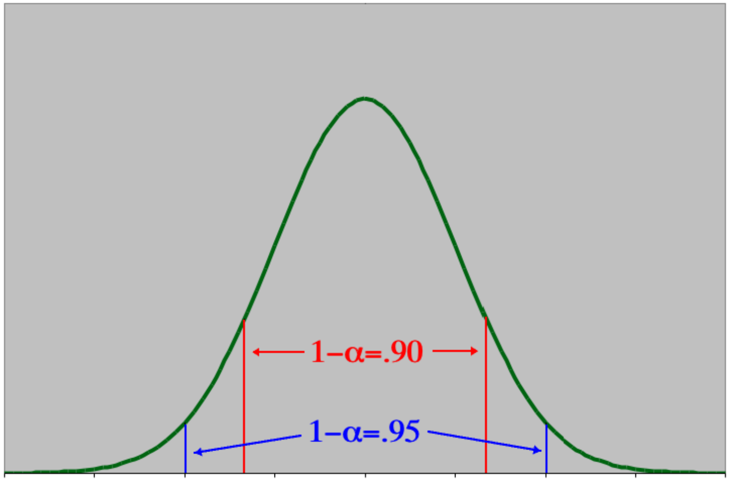
The second estimate is much narrower, providing accounting students more precise information about starting salaries.

Interval Width…

The width of the confidence interval estimate is a function of the confidence level, the population standard deviation, and the sample size…

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

A larger confidence level produces a **w i d e r** confidence interval:



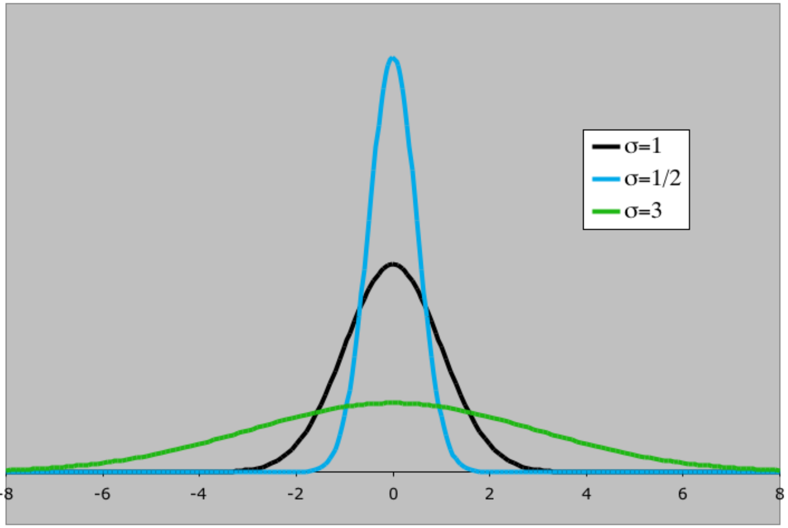
Interval Width…

The width of the confidence interval estimate is a function of the confidence level, the population standard deviation, and the sample size…

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Larger values of σ produce **w i d e r** confidence intervals

[Estimators.xls]



Interval Width…

The width of the confidence interval estimate is a function of the confidence level, the population standard deviation, and the sample size…

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Increasing the sample size decreases the width of the confidence interval while the confidence level can remain unchanged.

Note: this also increases the ***cost*** of obtaining additional data

Selecting the Sample Size…

We can select sample size to obtain a confidence interval of a certain with. From the formulas above the width of the confidence interval is 2B, where

$$B = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

is the distance between the mid-point and top of the confidence interval. For a given desired B, we can solve for n as.

$$n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2$$

Selecting the Sample Size…

To illustrate suppose that in Example 10.1 before gathering the data you decided that you needed to estimate the mean demand to within 16 units with 95 percent confidence. What sample size should he employ?

From before $1 - \alpha = .95$ and $\sigma = 75$. And now B=16, We calculate:

$$n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2 = \left(\frac{(1.96)(75)}{16}\right)^2 = 84.41$$

We round up our answer to 85.