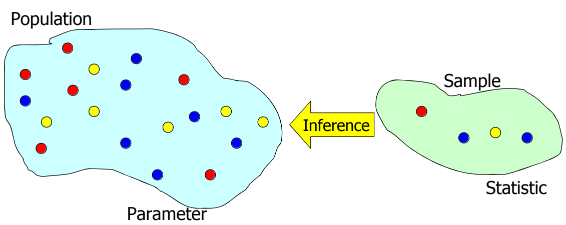
Chapter 10

Introduction to Estimation

Statistical Inference…

Statistical inference is the process by which we acquire information and draw conclusions about populations from samples.



In order to do inference, we require the skills and knowledge of descriptive statistics, probability distributions, and sampling distributions.

Estimation…

There are two types of inference: estimation and hypothesis testing; ***estimation*** is introduced first.

The objective of estimation is to determine the ***approximate value*** of a population parameter on the basis of a sample statistic.

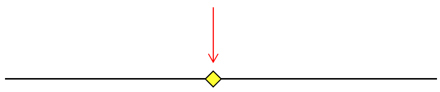
E.g., the sample mean ($\bar{x}$) is employed to ***estimate*** the population mean (μ).

Estimation…

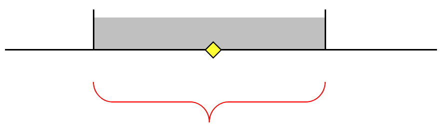
The objective of estimation is to determine the ***approximate value*** of a population parameter on the basis of a sample statistic.

There are two types of estimators:

Point Estimator

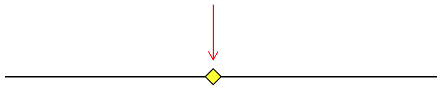


Interval Estimator



Point Estimator…

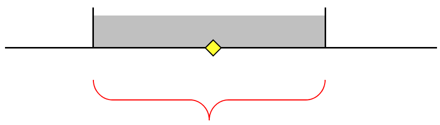
A ***point estimator*** draws inferences about a population by estimating the value of an unknown parameter using a single value or point.



We saw earlier that point probabilities in continuous distributions were virtually zero. Likewise, we'd expect that the point estimator gets closer to the parameter value with an increased sample size, but point estimators don't reflect the effects of larger sample sizes. Hence we will employ the ***interval estimator*** to estimate population parameters…

Interval Estimator…

An ***interval estimator*** draws inferences about a population by estimating the value of an unknown parameter using an interval.



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

Point & Interval Estimation…

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

interval estimate

An alternative statement is:

The mean income is ***between*** 380 and 420 \$/week.

Qualities of Estimators…

Qualities desirable in estimators include unbiasedness, consistency, and relative efficiency:

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

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

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

Unbiased Estimators…

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:

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

Unbiased Estimators…

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

E.g. the sample median is an ***unbiased*** estimator of the population mean μ since:

$$E(\text{Sample median}) = \mu$$

Consistency…

An unbiased 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. $\overline{X}$ is a ***consistent*** estimator of μ because:

$V(\overline{X})$ is σ^2/n

That is, as n grows larger, the variance of $\overline{X}$ grows smaller.

Consistency…

An unbiased 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. Sample median is a ***consistent*** estimator of μ because:

V(Sample median) is $1.57\sigma^2/n$

That is, as n grows larger, the variance of the sample median grows smaller.

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 the sample median and sample mean are unbiased estimators of the population mean, however, the sample median has a greater variance than the sample mean, so we choose $\bar{x}$ since it is ***relatively efficient*** when compared to the sample median.

Thus, the sample mean $\bar{x}$ is the "best" estimator of a population mean μ.

Estimating μ when σ is known…

In Chapter 8 we produced the following general probability statement about $\bar{X}$

$$P(-Z_{\alpha/2} < Z < Z_{\alpha/2}) = 1 - \alpha$$

And from Chapter 9 the sampling distribution of $\bar{X}$ is approximately normal with mean μ and standard deviation $\sigma / \sqrt{n}$

Thus

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

is (approximately) standard normally distributed.

Estimating μ when σ is known…

Thus, substituting Z we produce

$$P(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1-\alpha$$

In Chapter 9 (with a little bit of algebra) we expressed the following

$$P\left(\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

With a little bit of different algebra we have

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Estimating μ when σ is known…

This

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

is still a probability statement about $\bar{x}$.

However, the statement is also a confidence interval estimator of μ

Estimating μ when σ is known…

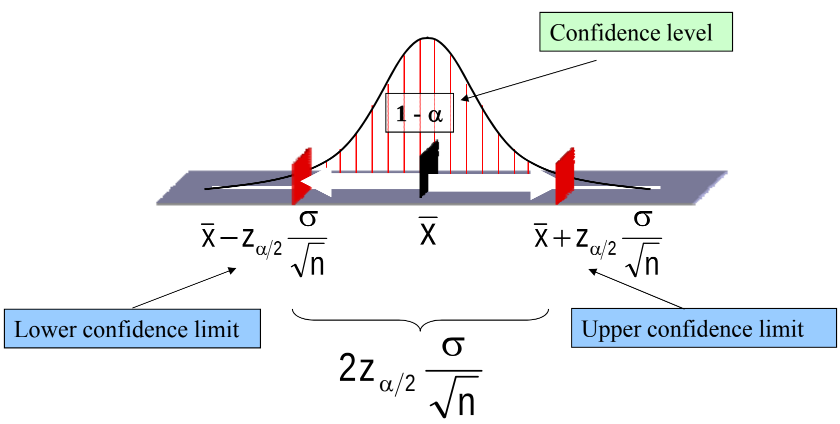
The interval can also be expressed as

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

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

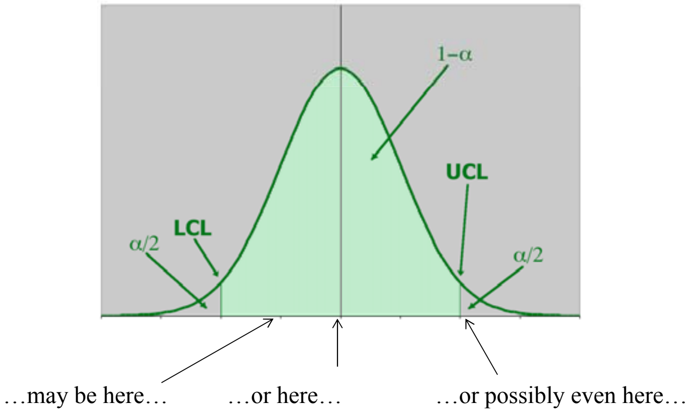
The probability $1 - \alpha$ is the confidence level, which is a measure of how frequently the interval will actually include μ.

Graphical Demonstration of the Confidence Interval for μ



Graphically…

…the ***actual*** location of the population mean μ …



The population mean is a fixed but ***unknown*** quantity. Its incorrect to interpret the confidence interval estimate as a probability statement about μ. The interval acts as the lower and upper limits of the interval ***estimate*** of the population mean.

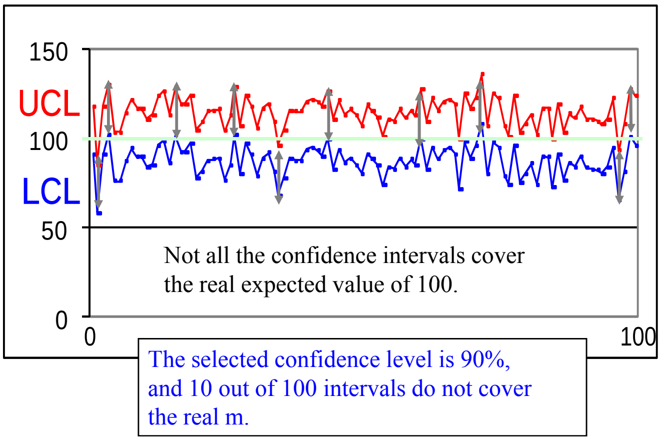
The Confidence Interval for μ (σ is known)

Four commonly used confidence levels

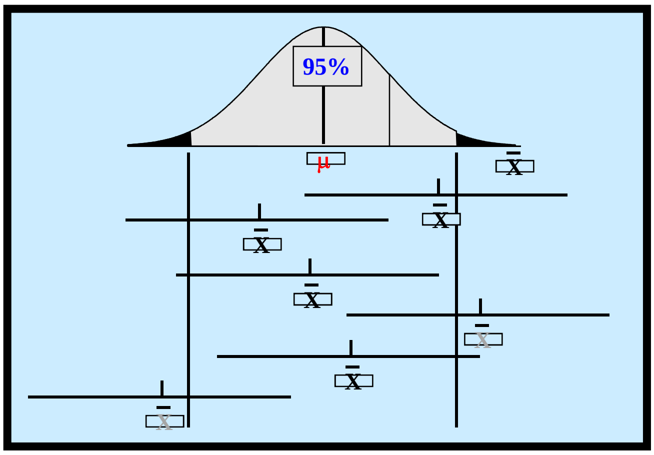
Confidence level	α	$\alpha/2$	$z_{\alpha/2}$
0.90	0.10	0.05	1.645
0.95	0.05	0.025	1.96
0.98	0.02	0.01	2.33
0.99	0.01	0.005	2.575

The confidence interval

The confidence interval are correct most, but not all, of the time.



95% Confidence Intervals for μ



The Confidence Interval for μ (σ is known)

E.g. Estimate the mean value of the distribution resulting from the throw of a fair die. It is known that s = 1.71. Use a 90% confidence level, and 100 repeated throws of the die

Solution: The confidence interval is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.645 \frac{1.71}{\sqrt{100}} = \bar{x} \pm .28$$

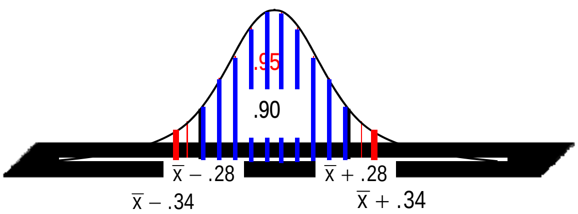
The mean values obtained in repeated draws of samples of size 100 result in interval estimators of the form
[sample mean - .28, Sample mean + .28],
90% of which cover the real mean of the distribution.

The Confidence Interval for μ (σ is known)

Recalculate the confidence interval for 95% confidence level.

Solution:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.96 \frac{1.71}{\sqrt{100}} = \bar{x} \pm .34$$



The Confidence Interval for μ (σ is known)

The width of the 90% confidence interval = 2(.28) = .56

The width of the 95% confidence interval = 2(.34) = .68

Because the 95% confidence interval is wider, it is more likely to include the value of μ.

Example 10.1…

The Doll Computer Company makes its own computers and delivers them directly to customers who order them via the Internet.

To achieve its objective of speed, Doll makes each of its five most popular computers and transports them to warehouses from which it generally takes 1 day to deliver a computer to the customer.

This strategy requires high levels of inventory that add considerably to the cost.

Example 10.1…

To lower these costs the operations manager wants to use an inventory model. He notes demand during lead time is normally distributed and he needs to know the mean to compute the optimum inventory level.

He observes 25 lead time periods and records the demand during each period. Xm10-01

The manager would like a 95% confidence interval estimate of the mean demand during lead time. Assume that the manager knows that the standard deviation is 75 computers.

Example 10.1…

"We want to estimate the ***mean*** demand over lead time with 95% confidence in order to set inventory levels…"

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…

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

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

Given

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

	A	B	C
1	**z-Estimate: Mean**		
2			
3			*Demand*
4	Mean		370.16
5	Standard Deviation		80.78
6	Observations		25
7	SIGMA		75
8	LCL		340.76
9	UCL		399.56

Example 10.1… INTERPRET

The estimation for the mean demand during lead time lies between 340.76 and 399.56 — we can use this as input in developing an inventory policy.

That is, we estimated that the mean demand during lead time falls between 340.76 and 399.56, and this type of estimator is correct 95% of the time. That also means that 5% of the time the estimator will be incorrect.

Incidentally, the media often refer to the 95% figure as "19 times out of 20," which emphasizes the ***long-run*** aspect of the confidence level.

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 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 translate the confidence interval estimate properly, we must remember that the confidence interval estimator was derived from the sampling distribution of the sample mean.

We used the sampling distribution to make probability statements about the sample mean.

Although the form has changed, 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. Once the sample mean is computed, the interval acts as the lower and upper limits of the interval estimate of the population mean.

Interpreting the Confidence Interval Estimator

As an illustration, suppose we want to estimate the mean value of the distribution resulting from the throw of a fair die.

Because we know the distribution, we also know that $\mu = 3.5$ and $\sigma = 1.71$.

Pretend now that we know only that $\sigma = 1.71$, that μ is unknown, and that we want to estimate its value.

To estimate , we draw a sample of size n = 100 and calculate. The confidence interval estimator of is

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

Interpreting the Confidence Interval Estimator

The 90% confidence interval estimator is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.645 \frac{1.71}{\sqrt{100}} = \bar{x} \pm .281$$

Interpreting the Confidence Interval Estimator

This notation means that, if we repeatedly draw samples of size 100 from this population, 90% of the values of $\bar{x}$ will be such that μ would lie somewhere

$$\bar{x} - .281 \text{ and } \bar{x} + .281$$

and that 10% of the values of $\bar{x}$ will produce intervals that would not include μ .

Now, imagine that we draw 40 samples of 100 observations each. The values of and the resulting confidence interval estimates of are shown in Table 10.2.

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.

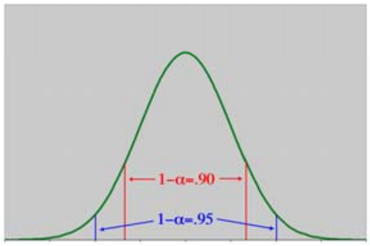
Interval Width…

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

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

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

Estimators.xls



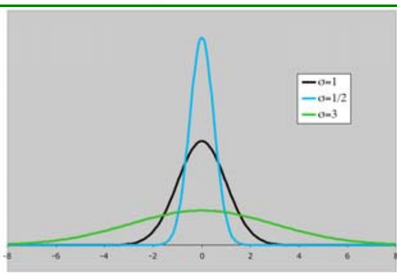
Interval Width…

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

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

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. Estimators.xls

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

Selecting the Sample Size…

In Chapter 5 we pointed out that sampling error is the difference between an estimator and a parameter.

We can also define this difference as the **error of estimation**.

In this chapter this can be expressed as the difference between $\bar{x}$ and μ.

Selecting the Sample Size…

The bound on the error of estimation is

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

With a little algebra we find the sample size to estimate a mean.

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

Selecting the Sample Size…

To illustrate suppose that in Example 10.1 before gathering the data the manager had decided that he needed to estimate the mean demand during lead time to with 16 units, which is the bound on the error of estimation.

We also have $1 - \alpha = .95$ and $\sigma = 75$. We calculate

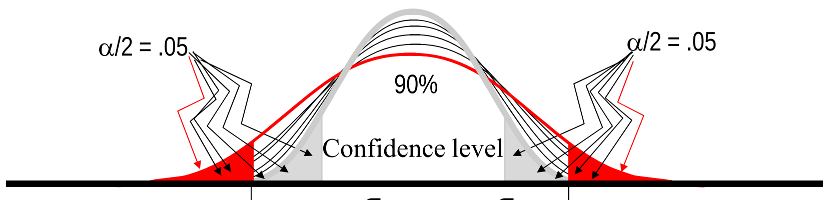
$$n = \left(\frac{z_{\alpha/2}\sigma}{B} \right)^2 = \left(\frac{(1.96)(75)}{16} \right)^2 = 84.41$$

Selecting the Sample Size…

Because n must be an integer and because we want the bound on the error of estimation to be *no more* than 16 any non-integer value must be rounded up.

Thus, the value of n is rounded to 85, which means that to be 95% confident that the error of estimation will be no larger than 16, we need to randomly sample 85 lead time intervals.

The Affects of σ on the interval width



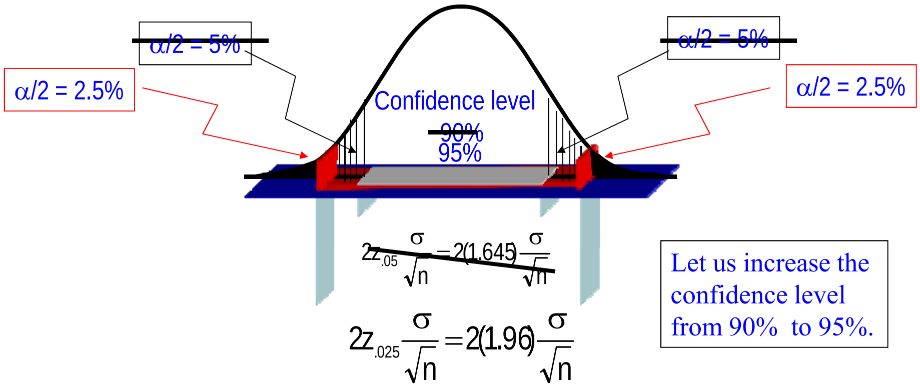
$$\cancel{2z_{.05}\frac{\sigma}{\sqrt{n}} = 2(1.645)\frac{\sigma}{\sqrt{n}}}$$

$$2z_{.05}\frac{1.5\sigma}{\sqrt{n}} = 2(1.645)\frac{1.5\sigma}{\sqrt{n}}$$

Suppose the standard deviation has increased by 50%.

To maintain a certain level of confidence, a larger standard deviation requires a larger confidence interval.

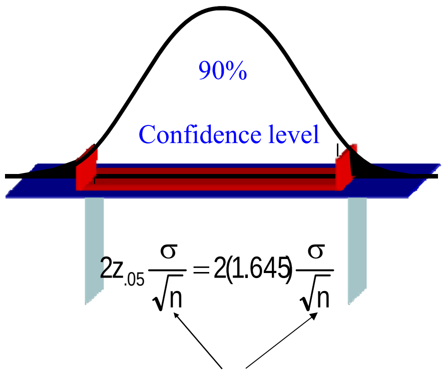
The Affects of Changing the Confidence Level



Let us increase the confidence level from 90% to 95%.

Larger confidence level produces a wider confidence interval

The Affects of Changing the Sample Size



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