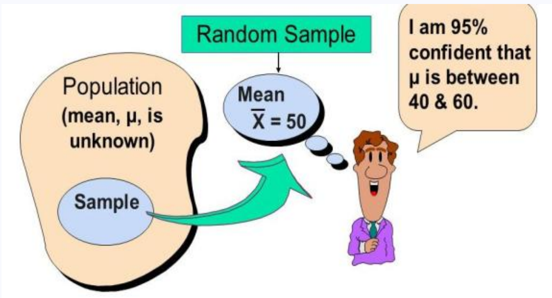
Chapter 10 –Introduction to Estimation

Chapter 10: Introduction to Estimation



1

Introduction

- There are two general procedures for making inferences about populations: estimation and hypothesis testing.
- The objective of estimation is to determine the approximate value of a population parameter based on a sample statistic.
- In this chapter, we introduce the concepts and foundations of estimation and demonstrate them with simple examples.
- In Chapter 11, we describe the fundamentals of hypothesis testing.
- Most of what we do in the remainder of this course applies the concepts of estimation and hypothesis testing.

2

2

Point Estimator

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

Drawbacks:

- Because the probability that a continuous random variable will equal a value is 0, it is virtually certain the estimate will be wrong (e.g., probability that $\bar{x}$ will exactly equal μ is 0)
- A point estimator does not reveal how close the estimate is to the unknown population parameter.
- A point estimator does not have the capacity to reflect the effects of larger sample sizes (i.e., the fact that large samples tend to produce more accurate results since it contains more information).

3

3

Interval Estimator

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

Examples:

- Television network executives want to know the proportion of television viewers who are tuned in to their networks
- An economist wants to know the mean income of university graduates
- A medical researcher who wishes to estimate the recovery rate of of heart attack victims treated with a new drug

→ Rely on random sample to estimate population parameters (mean, proportion, etc..)

What statistics can be used as an estimator?

4

4

Margin of Error and the Interval Estimate

A point estimator cannot be expected to provide the exact value of the population parameter.

An interval estimate can be computed by adding and subtracting a margin of error to the point estimate.

Point Estimate +/− Margin of Error

The purpose of an interval estimate is to provide information about how close the point estimate is to the value of the parameter.

5

5

Property 1: Unbiased

- An interval estimator is an **unbiased estimator** because its expected value equals the unknown population parameter.
- This means that if you were to take an infinite number of samples and calculate the value of the estimator in each sample, the average value of the estimators would equal the parameter.
- Sample mean is an unbiased estimator since $E(\bar{x}) = \mu$ *(Section 9.1)*
- Sample proportion is also an unbiased estimator since $E(\hat{p}) = p$

6

6

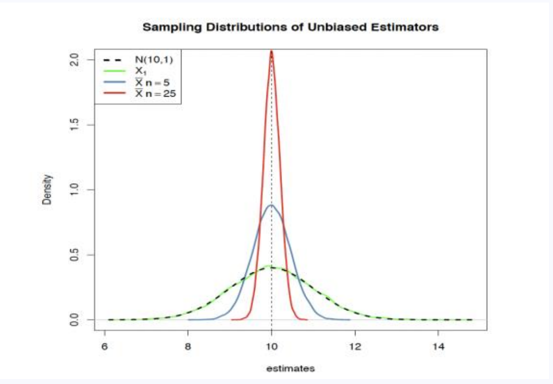
Property 2: Consistency

- How close is the estimator to the parameter?
- An unbiased estimator is **consistent** because the difference between the estimator and the parameter grows smaller as the sample size grows larger.
- More precisely, we want the probability that the estimate $(\hat{\theta})$ falls within a small interval around the true value (θ) to get increasingly closer to 1 as n grows. We write this as: $plim(\hat{\theta}) \to \theta$
- The measure used to gauge closeness is variance (or standard deviation)
- This characteristic is consistent with the sampling distribution of $\bar{X}$ and $\hat{p}$:
  - Variance of $\bar{X}$ is σ^2/n and variance of $\hat{p}$ is $p(1-p)/n$
  - Both decrease when the value of n goes up

7

7

Distribution of the Sample Mean as n Increases



8

8

Property 3: Relative efficiency

- An unbiased estimator also has **relative efficiency** when it has a variance smaller than other unbiased estimators.
- Suppose that we have two estimators, $\hat{\theta}$ and $\tilde{\theta}$, and for some given sample size n we have:

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

but $$Var\,(\hat{\theta}) < Var\,(\tilde{\theta})$$

- We then prefer to use $\hat{\theta}$ as it has lower variance than $\tilde{\theta}$, meaning that $\hat{\theta}$ is more **efficient** in using the information provided by the observations in the sample
- For instance, sample median is also an unbiased estimator but the its variance is greater than that of the sample mean → sample mean more efficient when estimating the population mean

9

9

Estimating the Population Mean

when the Population Standard Deviation is Known

- Consider the task of estimating the unknown mean μ *of* a population that is described by a continuous random variable X and has standard deviation σ.
- The first step in the estimation procedure is to draw a random sample of size n and calculate the sample mean $\bar{x}$.
- Because of the central limit theorem, if X is normally distributed or n is sufficiently large, then $\bar{X}$ is normally distributed.

10

10

Interval Estimate of a Population Mean: σ Known

The general form of an interval estimate of a population mean is $\bar{x} \pm$ Margin of Error

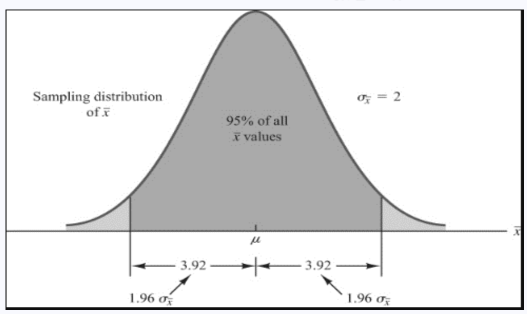
- In order to develop an interval estimate of a population mean, the margin of error must be computed using either:
  - the population standard deviation σ, or
  - the sample standard deviation s
- σ is rarely known exactly, but often a good estimate can be obtained based on historical data or other information.
- We refer to such cases as the σ known case.

11

11

Interval Estimate of a Population Mean: σ Known

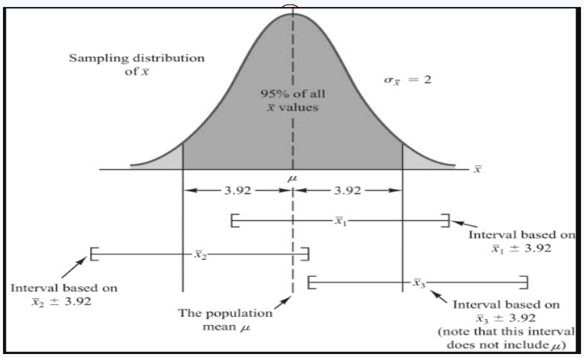
There is a $1 - \alpha$ probability that the value of a sample mean will provide a margin of error of $z_{\alpha/2}\sigma_{\bar{x}}$ less.



12

12

Interval Estimate of a Population Mean: σ Known



13

13

Interval Estimate of a Population Mean: σ Known

- **Interval Estimate of μ**

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

where: $\bar{x}$ is the sample mean
$1 - \alpha$ is the confidence coefficient
$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution
σ is the population standard deviation
n is the sample size

14

14

Interval Estimate of a Population Mean: σ Known

- Values of $z_{\alpha/2}$ for the Most Commonly Used Confidence Levels

Confidence Level	α	$\alpha/2$	Table Look-up Area	$z_{\alpha/2}$
90%	.10	.05	.9500	1.645
95%	.05	.025	.9750	1.960
99%	.01	.005	.9950	2.576

15

15

Meaning of Confidence

Because 90% of all the intervals constructed using $\bar{x} \pm 1.645\sigma_{\bar{x}}$ will contain the population mean,
we say we are 90% confident that the interval $\bar{x} \pm 1.645\sigma_{\bar{x}}$ includes the population mean μ.

We say that this interval has been established at the 90% confidence level.

The value .90 is referred to as the confidence coefficient.

16

16

Estimating the Population Mean

when the Population Standard Deviation Is Known

- We have shown in Section 9.1 that:

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

- We showed that we can build the following probability statement associated with the sampling distribution of the mean:

$$P(\mu - Z_{\alpha/2}\,\sigma/\sqrt{n} < \bar{X} < \mu + Z_{\alpha/2}\,\sigma/\sqrt{n}) = 1 - \alpha$$

- We can also using similar algebra manipulation to derive:

$$P(\bar{X} - Z_{\alpha/2}\,\sigma/\sqrt{n} < \mu < \bar{X} + Z_{\alpha/2}\,\sigma/\sqrt{n}) = 1 - \alpha$$

- The equation says that, with repeated sampling from this population, the proportion of values of $\bar{X}$ for which the interval includes the population mean μ is equal to $1 - \alpha$.

17

17

Confidence Interval Estimator of μ when σ Is Known

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- The probability $1 - \alpha$ is called the **confidence level**.
- ➢ $\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is called the **lower confidence limit (LCL)**.
- ➢ $\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is called the **upper confidence limit (UCL)**.

- We often represent the **confidence interval estimator of μ** as:

$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ where the minus sign defines LCL, and the plus sign defines UCL.

18

18

Confidence Interval Estimator of p when σ Is Known

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

(derived from the sampling distribution of a proportion)

- Again:
- ➢ $\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$ is called the **lower confidence limit (LCL).**
- ➢ $\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$ is called the **upper confidence limit (UCL).**
- **Confidence interval estimator of p as:** $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$

→where the minus sign defines LCL, and the plus sign defines UCL.

19

19

Doll Computer Company

- To lower inventory cost, Doll's operations manager wants to estimate the mean computer demand during lead time at a 95% confidence level.

- A sample of 25 lead times is collected and demand during each period recorded. Assume that the manager knows that the standard deviation of demand is 75 computers.

Identify:

The manager should build a confidence interval estimator of μ when **σ is known**:

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

We have: $n = 25$, and $\sigma = 75$.

20

20

Doll Computer Company

Compute

$\bar{x} = \sum x_i / n = \sum 9{,}254/25 = 370.16$ Xm10-01.xlsx

$1 - \alpha = .95 \Rightarrow \alpha = .05 \Rightarrow \alpha/2 = .025$

From Normal Probabilities Table: $z_{\alpha/2} = z_{.025} = 1.96$

Thus, the confidence interval estimator is: $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm 1.96\frac{75}{\sqrt{25}} = 370.16 \pm 29.40$

LCL = 340.76, UCL = 399.56

Interpret

The estimate of (340.76, 399.56) computers can be used as an input in developing an inventory policy.

Keep in mind, in Excel if you use Data Analysis the standard error is for a sample, not population.

21

21

Interpreting the Confidence Interval Estimate

Incorrect interpretation

"There is a $(1-\alpha)\%$ probability that the population parameter lies between LCL and UCL."

This interpretation is *wrong* because it implies that we can make probability statements about a population parameter, which is not a variable but a fixed and unknown quantity.

Correct interpretation

"If we repeatedly draw samples of the same size from a population, $(1-\alpha)\%$ of the interval estimates will include the unknown population parameter and $\alpha\%$ will not."

22

22

Stata Output and Instructions

Example – Estimating a Population Mean: Standard Deviation Known

1. Import the data into one column. Check **Import first row as variable names**.
2. Click **Statistics**, **Summaries, tables and tests**, **Classical tests of hypotheses**, and **z-test (mean comparison test known variance)**.
3. Select **One-sample**, select **Demand** in the **Variable name:** box and type any number in **the Hypothesized mean:** box. Type 95 for the **Confidence level**, and 75 for the **Standard deviation**.

Xm10-01.xlsx

```
One-sample z test

Variable |  Obs    Mean    Std. Err.  Std. Dev.  [95% Conf. Interval]
---------+-----------------------------------------------------------
  Demand |   25   370.16         15         75    340.7605   399.5595
---------------------------------------------------------------------
   mean = mean(Demand)                                   z =  24.6773
Ho: mean = 0

   Ha: mean < 0             Ha: mean != 0               Ha: mean > 0
 Pr(Z < z) = 1.0000    Pr(|Z| > |z|) = 0.0000       Pr(Z > z) = 0.0000
```


23

23

Information and the Width of the Interval

Interval estimation is as useful as it is efficient, and a narrower interval provides more information than a wider one.

The width of the interval estimate is a function of the population standard deviation σ, the confidence level $(1-\alpha)\%$, and the sample size n.

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

How can we reduce the interval's width?

- The sample mean $\bar{x}$ only affects the point estimate, not the width.
- σ is a parameter of the population, and we have no control over it.
- We could decrease $1-\alpha$, and therefore $z_{\alpha/2}$, but then we would have less confidence on the inference.
- We can always increase the sample size n, but there may be practical limitations to it

24

24

Information and the Width of the Interval

In the Doll Example:

Changing the confidence level

- To 90%, the interval estimate would have been:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm z_{0.05}\frac{75}{\sqrt{25}} = 370.16 \pm 1.645\frac{75}{\sqrt{25}} = 370.16 \pm 24.68$$

- To 99%, the interval estimate would have been:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm z_{0.005}\frac{75}{\sqrt{25}} = 370.16 \pm 2.575\frac{75}{\sqrt{25}} = 370.16 \pm 38.63$$

→Confidence level of 95% is considered "standard"

Changing the sample size

- To 100, the interval estimate would have been:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 370.16 \pm z_{0.025}\frac{75}{\sqrt{100}} = 370.16 \pm 1.96\frac{75}{\sqrt{100}} = 370.16 \pm 14.70$$

25

25

Interval Estimate of a Population Mean: σ Known

Example: Lloyds Department store

Each week Lloyds department store selects a simple random sample of 100 customers in order to learn about the amount spent per shopping trip. The historical data indicates that the population follows a normal distribution.

During most recent week, Lloyd's surveyed 100 customers (n = 100) and obtained a sample mean of $\bar{x}$ = \$82. Based on historical data, Lloyd's now assumes a known value of σ = \$20. The confidence coefficient to be used in the interval estimate is .95.

26

26

Interval Estimate of a Population Mean: σ Known

Example: Lloyds Department store

95% of the sample means that can be observed are within $\pm$ 1.96 $\sigma_{\bar{x}}$ of the population mean μ. The margin of error is:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\left(\frac{20}{\sqrt{100}}\right) = 3.92$$

Interval estimate of μ is:

\$82 $\pm$ \$ 3.92
or
\$78.08 to \$85.29

We are 95% confident that the interval contains the population mean.

27

27

Interval Estimate of a Population Mean: σ Known

Example: Lloyds Department store

Confidence level	Margin of Error	Interval estimate
90%	3.29	78.71 – 85.29
95%	3.92	78.08 – 85.92
99%	5.15	76.85 – 87.15

In order to have a higher degree of confidence, the margin of error and thus the width of the confidence interval must be larger.

28

28

Interval Estimate of a Population Mean: σ Known (another example)

- Example: Discount Sounds

Discount Sounds has 260 retail outlets throughout the United States. The firm is evaluating a potential location for a new outlet, based in part, on the mean annual income of the individuals in the marketing area of the new location.

A sample of size n = 36 was taken; the sample mean income is $41,100. The population is not believed to be highly skewed. The population standard deviation is estimated to be $4,500, and the confidence coefficient to be used in the interval estimate is .95.

29

29

Interval Estimate of a Population Mean: σ Known

- Example: Discount Sounds

95% of the sample means that can be observed are within $\pm$ 1.96$\sigma_{\bar{x}}$ of the population mean μ.

The margin of error is:

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\left(\frac{4,500}{\sqrt{36}}\right) = 1,470$$

Thus, at 95% confidence, the margin of error is $1,470.

30

30

Interval Estimate of a Population Mean: σ Known

- Example: Discount Sounds

Interval estimate of μ is:

$41,100 $\pm$ $1,470
or
$39,630 to $42,570

We are 95% confident that the interval contains the population mean.

31

31

Interval Estimate of a Population Mean: σ Known

- Example: Discount Sounds

Confidence Level	Margin of Error	Interval Estimate
90%	1234	$39,866 to $42,334
95%	1470	$39,630 to $42,570
99%	1932	$39,168 to $43,032

In order to have a higher degree of confidence, the margin of error and thus the width of the confidence interval must be larger.

32

32

Error of Estimation

We define the **error of estimation** as the difference between an estimator and the parameter. In this chapter, the error of estimation is expressed as the difference between $\bar{X}$ and μ *(or $\hat{p}$ and p)* :

With a simple algebraic manipulation, we express the confidence interval as:

$$|\bar{X} - \mu| = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

We interpret this to mean that $z_{\alpha/2}\,\sigma/\sqrt{n}$ is the maximum error of estimation we are willing to tolerate. We label it B, the **bound error of estimation**, that is:

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

33

33

Determining the Sample Size

If the population standard deviation σ, the confidence level *$(1 - \alpha)$%*, and the bound on the error of estimation B are known, we can solve for n and produce the required **sample size to estimate a mean**:

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

Consider that, in Example 10.1, before collecting the data, the manager needs to estimate the mean demand during lead time within 16 units. Remember also that *$1 - \alpha = .95$, and $\sigma = 75$.*

Thus:

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

Because the sample size must be an integer and we want the bound on the error of estimation to be no more than 16, any non-integer value must be rounded up. Thus, we round n to 85.

34

34

Determining the Sample Size

If the population standard deviation σ, the confidence level *$(1 - \alpha)$%*, and the bound on the error of estimation B are known, we can solve for n and produce the required **sample size to estimate a mean**:

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

Consider that, in Example 10.1, before collecting the data, the manager needs to estimate the mean demand during lead time within 16 units. Remember also that *$1 - \alpha = .95$, and $\sigma = 75$.*

Thus:

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

Because the sample size must be an integer and we want the bound on the error of estimation to be no more than 16, any non-integer value must be rounded up. Thus, we round n to 85.

35

35

Interval Estimate of a Population Mean: σ Known

- Adequate Sample Size

In most applications, a sample size of n = 30 is adequate.

If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended.

If the population is not normally distributed but is roughly symmetric, a sample size as small as 15 will suffice.

If the population is believed to be at least approximately normal, a sample size of less than 15 can be used.

36

36