Chapter 7: Interval Estimation: One Population

Department of Mathematics
Izmir University of Economics

Week 9-10
2014-2015

Introduction

In this chapter we will focus on

- inferential statements concerning estimation of a single population parameter, based on information contained in a random sample,
- the estimation of population mean μ, population proportion p, and population variance σ^2, and
- two estimation procedures:
  - First we estimate an unknown population parameter by a single number called a **point estimate**.
  - For most practical problems, a point estimate alone is not sufficient. So, a measure of variability is established by using an interval of values called a **confidence interval**.

Estimators

Definition:

An *estimator* of a population parameter is a random variable depending on the sample information. The value of an estimator provides an approximation to the unknown parameter. A specific value of the estimator is called as *estimate*.

We'll deal with two kinds of estimators:

- point estimator
- interval estimator

Point Estimators

Definition:

Consider a population parameter. A *point estimator* of a population parameter is a function of the sample information which produces a **single number**. This single number is called a *point estimate*.

Example:

The sample mean, $\bar{X}$, is a point estimator of the population mean, μ, and the value that $\bar{X}$ assumes for a given set of data is called the point estimate, $\bar{x}$.

Note: In general, we denote the population parameter by θ and the estimator of θ by $\hat{\theta}$.

Properties of Point Estimators

Estimators are evaluated depending on three important properties:

- unbiasedness
- consistency
- efficiency

Unbiasedness

Definition:

A point estimator is *unbiased* if its expected value is equal to the population parameter. That is, if

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

then $\hat{\theta}$ is an unbiased estimator of the parameter θ.

Example:

The sample variance (s^2) is an unbiased estimator of the population variance (σ^2) because

$$E(s^2) = \sigma^2.$$

Interpretation: An unbiased estimator can estimate the population parameter correctly on the average. In other words, if the estimator $\hat{\theta}$ is unbiased then the average $\hat{\theta}$ value is exactly correct. (It does not mean that a particular $\hat{\theta}$ is exactly correct.)

Note: **The estimators which are not unbiased are said to be *biased*.**

Definition:

Let $\hat{\theta}$ be an estimator of θ. We define the *bias* in terms of $\hat{\theta}$ as follows

$$\mathrm{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

- If $\mathrm{Bias}(\hat{\theta}) = 0$, then $\hat{\theta}$ is unbiased.
- If $\mathrm{Bias}(\hat{\theta}) \neq 0$, then $\hat{\theta}$ is biased.

Consistency

Definition:

A point estimator $\hat{\theta}$ is a *consistent* estimator of the population parameter θ if $\mathrm{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$ becomes smaller as the sample size increases.

Example:

$t^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is a consistent estimator of σ^2 because, as $n \to \infty$, t^2 approaches $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, which is an unbiased estimator of σ^2.

Efficiency

Definition:

If there are several unbiased estimators of a parameter, the estimator with the smallest variance is called the *most efficient estimator*.

Example:

Let θ be a population parameter with two point estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ satisfying $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$. Then we say that $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ because it has a smaller variance.

Properties of Selected Point Estimators

Properties of Selected Point Estimators:

Population parameter	Point estimator	Properties
Mean, μ	$\overline{X}$	unbiased, most efficient
Mean, μ	Median, $\widetilde{x}$	unbiased, not most efficient
Variance, σ^2	s^2	unbiased, most efficient

Interval Estimators

Definition:

A *confidence interval estimator* for a population parameter is a rule for determining a range or an interval that is likely to include the parameter. The corresponding estimate is called a *confidence interval estimate*.

Note: An interval estimate is better than a point estimate since it allows more information about a population characteristic.

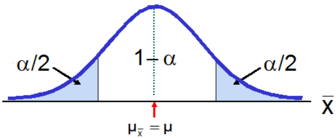
Confidence Interval

Definition:

Let θ be unknown parameter. If

$$P(a < \theta < b) = 1 - \alpha, \quad 0 < \alpha < 1,$$

then the interval (a, b) is called a $100(1-\alpha)\%$ *confidence interval (CI)* of θ and the quantity $100(1-\alpha)\%$ is called the *confidence level* of the interval.



In repeated samples of the population, the true value of the population parameter θ is contained in $100(1-\alpha)\%$ of the intervals.

Now, we will construct confidence intervals for

- population mean whenever the population is normally distributed and population variance is known,
- population mean whenever the population is normally distributed and population variance is unknown,
- population proportion, and
- population variance whenever the population is normally distributed.

CI for the mean of a population that is normally distributed: Population variance known

In this section population variance (σ^2) is known but population mean (μ) is unknown. Our aim is to find a range of values (rather than a single value) to estimate the population mean μ.

Assumptions:

- Population is normally distributed. That is consider a random sample of n observations from a normal distribution with known variance σ^2.
- If the population is not normal use a large sample.

The confidence interval for the population mean with **known variance** is

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

The confidence interval can be also written as

$$\bar{x} \pm \mathrm{ME}.$$

Here

- $z_{\frac{\alpha}{2}}$ is called the **reliability factor**,
- $SE = \frac{\sigma}{\sqrt{n}}$ is the **standard error**,
- $ME = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is the **margin of error**,
- $w = 2ME$ is the **width**,
- $UCL = \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is the **upper confidence limit**, and
- $LCL = \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ is the **lower confidence limit**.

Example: For 95% confidence level find the reliability factor.

Example: Suppose that shopping times for customers at a local grocery store are normally distributed with standard deviation of 20 minutes. A random sample of 64 shoppers in the local grocery store had a mean time of 75 minutes. Find the standard error, margin of error, width, and the upper and lower confidence limits of a 95% confidence interval for the population mean μ.

Example: A process produces bags of refined sugar. The weights of the contents of these bags are normally distributed with standard deviation 1.2 ounces. The contents of a random sample of 25 bags had a mean weight of 19.8 ounces. Find the upper and lower confidence limits of 99% confidence interval for the true mean weight for all bags of sugar produced by the process.

Common Levels of Confidence

Confidence level percentage	α	$z_{\frac{\alpha}{2}}$
80	0.2	1.28
90	0.1	1.645
95	0.05	1.96
98	0.02	2.33
99	0.01	2.58

CI for the mean of a population that is normally distributed: Population variance unknown

In this section, we will consider the case in which there is no information concerning both the population mean and the population variance.

This approach provides more realistic solutions especially on small samples.

Definition:

Consider a random sample of n observations with mean $\bar{x}$ and standard deviation s from a normally distributed population with mean μ.

The random variable

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}},$$

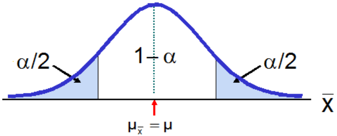
follows **Student's t distribution** with $\nu = n - 1$ degrees of freedom.

Student's *t* Distribution

Remark:

- Since the population standard deviation σ is unknown, we use the sample standard deviation *s*.
- If $n \leq 30$ we use *t*-distribution, otherwise, i.e., if $n > 30$ we can use normal distribution (*z*-distribution) by the *Central Limit Theorem*.
- The graph of density function of *t* is very similar to the one of standard normal distribution. It gets closer to the normal distribution graph as ν increases.
- The reliability factor is denoted by $t_{\nu,\frac{\alpha}{2}}$; note that

$$\text{Shaded area} = \frac{\alpha}{2} = P(t_\nu > t_{\nu,\frac{\alpha}{2}})$$



As an example, suppose that we require the number, which is exceeded with probability 0.05, by a Student's *t* random variable with 15 degrees of freedom.

Then

$$\nu = 15, \quad \frac{\alpha}{2} = 0.05$$

and by the last property

$$P(t_{15} > t_{15,0.05}) = \frac{\alpha}{2} = 0.05.$$

Using the table for Student's t distribution

$$t_{15,0.05} = 1,753.$$

CI for the mean of a normal distribution: Population variance unknown

Assumptions:

- Population standard deviation is unknown.
- Population is normally distributed.
- $t \to Z$ as n increases. That is, for large samples Student's t distribution behaves similar to standard normal distribution. Therefore if the population distribution is not normal, it is better to use a large sample.

The confidence interval for the population mean with **unknown variance** is

$$\bar{x} - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}.$$

The confidence interval can be also written as

$$\bar{x} \pm \mathrm{ME},$$

where

$$ME = t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}.$$

Example. Calculate the margin of error to estimate the population mean μ for each of the following:

a) $\alpha = 0.05$, $n = 6$, $s = 40$,

b) $\alpha = 0.1$, $n = 64$, $s = 56$,

c) 98% confidence interval, $n = 120$, $s^2 = 100$.

Example. Times in minutes that a random sample of five people spend driving to work are

30 42 35 40 45

a) Calculate the standard error.

b) Find $t_{\nu, \frac{\alpha}{2}}$ for a 95 percentage confidence interval for the true population mean.

c) Calculate width for a 95% confidence interval for the population mean time spent driving to work.

CI for the population proportion

Recall that, we denote the proportion of success of a population by p. Also, $\hat{p}$ refers the observed population proportion of successes in a random sample of n observations from a population with proportion p.

Then, if $np(1-p) > 5$, a $100(1-\alpha)\%$ confidence interval for the population proportion is given by

$$\hat{p} \pm \mathrm{ME},$$

where

$$\mathrm{ME} = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Example: Suppose that a random sample of 142 graduate - admissions personnel was asked what role scores of standardized tests (such as the GMAT or GRE) play in the consideration of a candidate for graduate school. Of these sample members, 87 answered "very important". Find a 95% confidence interval for the population proportion of graduate admissions personnel with this view.

CI for the variance of a normal population

Suppose that there is a random sample of n observations from a normally distributed population with variance σ^2.

If the observed sample variance is s^2, then the lower and upper confidence limits of a $100(1-\alpha)\%$ confidence interval for the population variance is given respectively by

$$\mathrm{LCL} = \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \text{ and } \mathrm{UCL} = \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}.$$

Example: A psychologist wants to estimate the variance of employee test scores. A random sample of 18 scores had a sample standard deviation of 10.4. Find a 90% confidence interval for the population variance.