Chapter 8
Interval Estimation

- Population Mean: σ Known
- Population Mean: σ Unknown
- Determining the Sample Size
- Population Proportion

Margin of Error and the Interval Estimate

- A point estimator cannot be expected to provide the exact value of the population parameter.
- An <u>interval estimate</u> can be computed by adding and subtracting a <u>margin of error</u> 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.

Margin of Error and the Interval Estimate

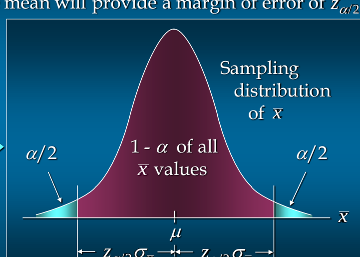
- The general form of an interval estimate of a population mean is
- $\bar{x} \pm$ Margin of Error

Interval Estimate of a Population Mean: σ Known

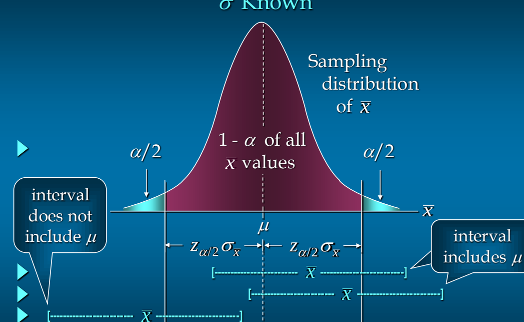
- In order to develop an interval estimate of a population mean, the margin of error must be computed using either:
  - the population standard deviation σ, or
  - the sample standard deviation s
- σ is rarely known exactly, but often a good estimate can be obtained based on historical data or other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

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 $z_{\alpha/2}\sigma_{\bar{x}}$ or less.



Interval Estimate of a Population Mean: σ Known



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

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

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.

Interval Estimate of a Population Mean: σ Known

- 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.

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.

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.

Interval Estimate of a Population Mean: σ Known

- Example: Discount Sounds

Confidence Level	Margin of Error	Interval Estimate
90%	3.29	78.71 to 85.29
95%	3.92	78.08 to 85.92
99%	5.15	76.85 to 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.

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.

Interval Estimate of a Population Mean: σ Known

- Adequate Sample Size (continued)

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.

Interval Estimate of a Population Mean: σ Unknown

- If an estimate of the population standard deviation σ cannot be developed prior to sampling, we use the sample standard deviation s to estimate σ.
- This is the σ unknown case.
- In this case, the interval estimate for μ is based on the t distribution.
- (We'll assume for now that the population is normally distributed.)

t Distribution

William Gosset, writing under the name "Student", is the founder of the t distribution.

Gosset was an Oxford graduate in mathematics and worked for the Guinness Brewery in Dublin.

He developed the t distribution while working on small-scale materials and temperature experiments.

t Distribution

The t distribution is a family of similar probability distributions.

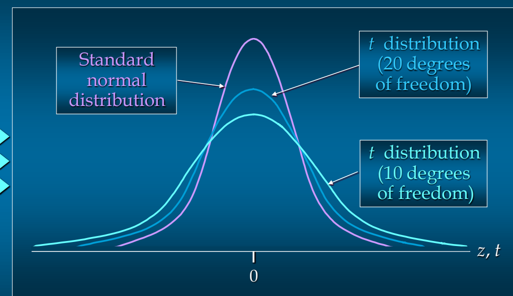
A specific t distribution depends on a parameter known as the degrees of freedom.

Degrees of freedom refer to the number of independent pieces of information that go into the computation of s.

t Distribution

- A t distribution with more degrees of freedom has less dispersion.
- As the degrees of freedom increases, the difference between the t distribution and the standard normal probability distribution becomes smaller and smaller.

t Distribution



t Distribution

- For more than 100 degrees of freedom, the standard normal z value provides a good approximation to the t value.
- The standard normal z values can be found in the infinite degrees (∞) row of the t distribution table.

t Distribution

Degrees	Area in Upper Tail					
of Freedom	**.20**	**.10**	**.05**	**.025**	**.01**	**.005**
.	.	.	.	.	.	.
50	.849	1.299	1.676	2.009	2.403	2.678
60	.848	1.296	1.671	2.000	2.390	2.660
80	.846	1.292	1.664	1.990	2.374	2.639
100	.845	1.290	1.660	1.984	2.364	2.626
∞	.842	1.282	1.645	1.960	2.326	2.576

Standard normal z values

Interval Estimate of a Population Mean: σ Unknown

- Interval Estimate

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where: $1 - \alpha$ = the confidence coefficient

$t_{\alpha/2}$ = the t value providing an area of $\alpha/2$ in the upper tail of a t distribution with n - 1 degrees of freedom

s = the sample standard deviation

Interval Estimate of a Population Mean: σ Unknown

- Example: Apartment Rents
  - A reporter for a student newspaper is writing an article on the cost of off-campus housing. A sample of 16 efficiency apartments within a half-mile of campus resulted in a sample mean of \$750 per month and a sample standard deviation of \$55.
  - Let us provide a 95% confidence interval estimate of the mean rent per month for the population of efficiency apartments within a half-mile of campus. We will assume this population to be normally distributed.

Interval Estimate of a Population Mean: σ Unknown

- At 95% confidence, α = .05, and $\alpha/2$ = .025.
- $t_{.025}$ is based on $n - 1 = 16 - 1 = 15$ degrees of freedom.
- In the t distribution table we see that $t_{.025} = 2.131$.

Degrees	Area in Upper Tail					
of Freedom	**.20**	**.100**	**.050**	**.025**	**.010**	**.005**
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.520	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
.	.	.	.	.	.	.

Interval Estimate of a Population Mean: σ Unknown

- Interval Estimate
  - $\bar{x} \pm t_{.025}\frac{s}{\sqrt{n}}$ — Margin of Error
  - $750 \pm 2.131\frac{55}{\sqrt{16}} = 650 \pm \underline{29.30}$
  - We are 95% confident that the mean rent per month for the population of efficiency apartments within a half-mile of campus is between \$720.70 and \$779.30.

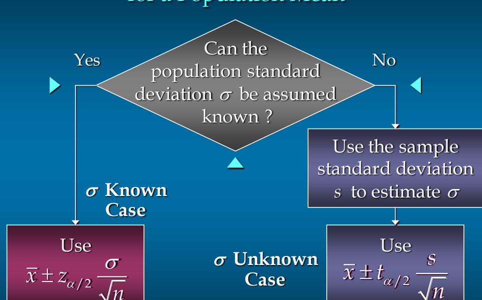
Interval Estimate of a Population Mean: σ Unknown

- Adequate Sample Size
  - In most applications, a sample size of n = 30 is adequate when using the expression $\bar{x} \pm t_{\alpha/2}\, s/\sqrt{n}$ to develop an interval estimate of a population mean.
  - If the population distribution is highly skewed or contains outliers, a sample size of 50 or more is recommended.

Interval Estimate of a Population Mean: σ Unknown

- Adequate Sample Size (continued)
  - 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.

Summary of Interval Estimation Procedures for a Population Mean



Sample Size for an Interval Estimate of a Population Mean

- Let E = the desired margin of error.
- E is the amount added to and subtracted from the point estimate to obtain an interval estimate.
- If a desired margin of error is selected prior to sampling, the sample size necessary to satisfy the margin of error can be determined.

Sample Size for an Interval Estimate of a Population Mean

- Margin of Error

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

- Necessary Sample Size

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2}$$

Sample Size for an Interval Estimate of a Population Mean

- The Necessary Sample Size equation requires a value for the population standard deviation σ.
- If σ is unknown, a preliminary or planning value for σ can be used in the equation.
  1. Use the estimate of the population standard deviation computed in a previous study.
  2. Use a pilot study to select a preliminary study and use the sample standard deviation from the study.
  3. Use judgment or a "best guess" for the value of σ.

Sample Size for an Interval Estimate of a Population Mean

- Example: Discount Sounds

Recall that Discount Sounds is evaluating a potential location for a new retail outlet, based in part, on the mean annual income of the individuals in the marketing area of the new location.

Suppose that Discount Sounds' management team wants an estimate of the population mean such that there is a .95 probability that the sampling error is $500 or less.

How large a sample size is needed to meet the required precision?

Sample Size for an Interval Estimate of a Population Mean

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 500$$

- At 95% confidence, $z_{.025} = 1.96$. Recall that $\sigma = 4{,}500$.

$$n = \frac{(1.96)^2(4{,}500)^2}{(500)^2} = 311.17 = 312$$

- A sample of size 312 is needed to reach a desired precision of $\pm$ \$500 at 95% confidence.

Interval Estimate of a Population Proportion

- The general form of an interval estimate of a population proportion is

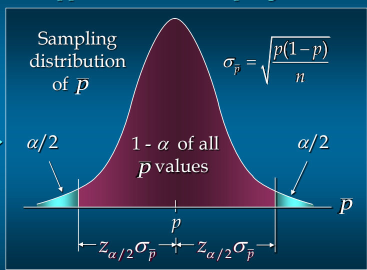
$$\bar{p} \pm \text{Margin of Error}$$

Interval Estimate of a Population Proportion

- The sampling distribution of $\bar{p}$ plays a key role in computing the margin of error for this interval estimate.
- The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $n(1 - p) \geq 5$.

Interval Estimate of a Population Proportion

- Normal Approximation of Sampling Distribution of $\bar{p}$



Interval Estimate of a Population Proportion

- Interval Estimate

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where: $1 - \alpha$ 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

$\bar{p}$ is the sample proportion

Interval Estimate of a Population Proportion

- Example: Political Science, Inc.

Political Science, Inc. (PSI) specializes in voter polls and surveys designed to keep political office seekers informed of their position in a race.

Using telephone surveys, PSI interviewers ask registered voters who they would vote for if the election were held that day.

Interval Estimate of a Population Proportion

- Example: Political Science, Inc.

In a current election campaign, PSI has just found that 220 registered voters, out of 500 contacted, favor a particular candidate. PSI wants to develop a 95% confidence interval estimate for the proportion of the population of registered voters that favor the candidate.

Interval Estimate of a Population Proportion

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

where: $n = 500$, $\bar{p} = 220/500 = .44$, $z_{\alpha/2} = 1.96$

$$.44 \pm 1.96\sqrt{\frac{.44(1-.44)}{500}} = .44 + .0435$$

PSI is 95% confident that the proportion of all voters that favor the candidate is between .3965 and .4835.

Sample Size for an Interval Estimate of a Population Proportion

- Margin of Error

$$E = z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Solving for the necessary sample size, we get

$$n = \frac{(z_{\alpha/2})^2\,\bar{p}(1-\bar{p})}{E^2}$$

However, $\bar{p}$ will not be known until after we have selected the sample. We will use the planning value p^* for $\bar{p}$.

Sample Size for an Interval Estimate of a Population Proportion

- Necessary Sample Size

$$n = \frac{(z_{\alpha/2})^2 p^*(1-p^*)}{E^2}$$

The planning value p^* can be chosen by:

1. Using the sample proportion from a previous sample of the same or similar units, or
2. Selecting a preliminary sample and using the sample proportion from this sample.
3. Use judgment or a "best guess" for a p^* value.
4. Otherwise, use .50 as the p^* value.

Sample Size for an Interval Estimate of a Population Proportion

- Example: Political Science, Inc.

Suppose that PSI would like a .99 probability that the sample proportion is within + .03 of the population proportion.

How large a sample size is needed to meet the required precision? (A previous sample of similar units yielded .44 for the sample proportion.)

Sample Size for an Interval Estimate of a Population Proportion

$$z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = .03$$

At 99% confidence, $z_{.005} = 2.576$. Recall that $\bar{p} = .44$.

$$n = \frac{(z_{\alpha/2})^2 p(1-p)}{E^2} = \frac{(2.576)^2(.44)(.56)}{(.03)^2} \cong 1817$$

A sample of size 1817 is needed to reach a desired precision of $\pm$ .03 at 99% confidence.

Sample Size for an Interval Estimate of a Population Proportion

Note: We used .44 as the best estimate of p in the preceding expression. If no information is available about p, then .5 is often assumed because it provides the highest possible sample size. If we had used p = .5, the recommended n would have been 1843.

End of Chapter 8