Chapter 7

Confidence Intervals and Sample Size

Confidence Intervals for the Mean (σ Known or $n \geq 30$) and Sample Size

Confidence Intervals for the Mean (σ Known or $n \geq 30$)

One aspect of inferential statistics is *estimation*, which is the process of estimating the value of a parameter from information obtained from a sample.

A *point estimate* is a specific numerical value estimate of a parameter. The best point estimate of the population mean μ is the sample mean $\overline{X}$.

Confidence Intervals for the Mean (σ Known or $n \geq 30$) (Cont.)

Three Properties of a Good Estimator:

1. The estimator should be an *unbiased estimator*. That is, the expected value or the mean of the estimates obtained from samples of a given size is equal to the parameter being estimated.
2. The estimator should be consistent. For a *consistent estimator*, as sample size increases, the value of the estimator approaches the value of the parameter estimated.
3. The estimator should be *relatively efficient estimator*. That is, of all the statistics that can be used to estimate a parameter, the relatively efficient estimator has the smallest variance.

Confidence Intervals for the Mean (σ Known or $n \geq 30$) (Cont.)

An *interval estimate* of a parameter is an interval or a range of values used to estimate the parameter. This estimate may or may not contain the value of the parameter being estimated.

The *confidence level* of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter.

A *confidence interval* is a specific interval estimate of a parameter determined by using data obtained from a sample and by using the specific confidence level of the estimate.

Confidence Intervals for the Mean (σ Known)

<u>Formula for the Confidence Interval of the Mean for a Specific α:</u>

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

For a 90% confidence interval, $z_{\alpha/2}=1.645$;
for a 95% confidence interval, $z_{\alpha/2}=1.96$; and
for a 99% confidence interval, $z_{\alpha/2}=2.575$.

Confidence Intervals for the Mean (σ Known) (Cont.)

Note:

- $z_{\alpha/2}$ is the z value such that $P(z > z_{\alpha/2}) = \alpha/2$ or such that $P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1 - \alpha$.
- $(1 - \alpha)100\%$ is the level of confidence.
- $1 - \alpha$ is the confidence coefficient (level of significant).

Confidence Intervals for the Mean (σ Known) (Cont.)

The term $z_{\alpha/2}\left(\sigma/\sqrt{n}\right)$ is called maximum error of estimate. For a specific value, say, α = 0.05, 95% of the sample means will fall within this error value, $z_{\alpha/2}\left(\sigma/\sqrt{n}\right)$,on either side of the population mean.

The *maximum error of estimate* is the maximum likely difference between the point estimate of a parameter and the actual value of the parameter.

Confidence Intervals for the Mean (σ Known) (Cont.)

Example 1: Given a random sample of 40 observations from a normal population with standard deviation of 8 and the sample mean of 35.3. Construct a 95% confidence interval for population mean.

Confidence Intervals for the Mean (σ Known) (Cont.)

Example 2: Using pervious data with a sample size of 20. Construct a 99% confidence interval for the population mean.

Confidence Intervals for the Mean (σ Known) (Cont.)

Example 3 :A study of 40 English composition professors showed that they spent, on average, 12.6 minutes correcting a student's term paper. Find the 90% confidence interval of the mean time for all composition papers when $\sigma = 2.5$ minutes.

Confidence Intervals for the Mean (σ Unknown and $n \geq 30$)

If $\overline{X}$ and s are mean and standard deviation of a random sample of size $n \geq 30$ from a population with unknown standard deviation σ, then the formula for the confidence interval of the mean for a specific α is given by

$$\overline{X} - z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Confidence Intervals for the Mean (σ Unknown and $n \geq 30$)

Example 4: From a random sample of 70 students, the sample mean and sample standard deviation of Math scores on a certain exam are found to be 496 and 75, respectively. Determine a 95% confidence interval for the mean Math scores.

Sample size

If $\overline{X}$ is used as an estimate for μ, we can be (1 – α)100% , for example 95%, 98% ,99%, confident that the error will not exceed E where $E = z_{\alpha/2}\left(\sigma/\sqrt{n}\right)$. and this formula is solved for n as follows:

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

$$\therefore \quad n = \left(\frac{z_{\alpha/2}\cdot\sigma}{E}\right)^2$$

Final sample size always round up!!!

Sample size (Cont.)

Example 5: Survey of 50 people on average weekly income has a sample mean of \$630 and sample standard deviation of \$35.

(a) Construct a 95% confidence interval for the true mean weekly income.

(b) How large a sample is required if we want to be 95% confident that the sample mean will be within \$7 of population mean.

(c) How large a sample would be required to be 90% confident that our estimate is within \$7 of our population mean.

Confidence Intervals for the Mean (σ Unknown an $n < 30$)

Confidence Intervals for the Mean (σ Unknown and n < 30)

When the population standard deviation is not known and the sample size is less than 30, the standard deviation from a sample mean can be used in place of the population standard deviation for confidence intervals.

But a somewhat different distribution, called the *t distribution*, must be used when the sample size is less than 30 and the variable is normally or approximately normally distributed.

t Distribution

<u>Characteristics of the *t* Distribution</u>:

The *t* distribution shares some characteristics of the normal distribution and differs from it in others. The *t* distribution differs from the standard normal distribution in the following ways.

1. The variance is greater than 1.
2. The *t* distribution is actually a family of curves based on the concept of degrees of freedom, which is related to sample size.
3. As the sample size increases, the *t* distribution approaches the standard normal distribution.

Confidence Intervals for the Mean (σ Unknown and n < 30) (Cont.)

<u>Formula for the Confidence Interval of the Mean for a Specific α When σ is Unknown and n < 30</u> :

$$\overline{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

The degree of freedom (d.f.) are $n - 1$.

Confidence Intervals for the Mean (σ Unknown and n < 30) (Cont.)

Example 6 : Find the values for each.

1. $t_{\alpha/2}$ and $n = 18$ for the 99% confidence interval for the mean.
2. $t_{\alpha/2}$ and $n = 23$ for the 95% confidence interval for the mean.
3. $t_{\alpha/2}$ and $n = 25$ for the 98% confidence interval for the mean.
4. $t_{\alpha/2}$ and $n = 10$ for the 90% confidence interval for the mean.

Confidence Intervals for the Mean (σ Unknown and $n < 30$) (Cont.)

Example 7 : A sample of six adult elephants had an average weight of 12,200 pounds, with a sample standard deviation of 200 pounds. Find the 95% confidence interval of the true mean.

Confidence Intervals for the Mean (σ Unknown and $n < 30$) (Cont.)

Example 8: The following data give the speed (km/hr), as measured by radar, for 10 cars traveling between Regina and Saskatoon

109	112	100	104	125
104	135	132	112	108

Assuming that the speeds of all cars have a normal distribution, calculate a 99% confidence interval for the mean speeds of all cars.

Confidence Intervals and Sample Size for Proportions

Confidence Intervals for Proportions

If $\hat{p}$ is the proportion of successes in a random sample of large size n , that is $n\hat{p} \geq 5$ and $n\hat{q} \geq 5$, then a formula for a specific confidence interval for a proportion is given by

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where

$p =$ population proportion , $\hat{p} =$ sample proportion $= \dfrac{X}{n}$

$X =$ number of sample units that possess the characteristics of interest

Confidence Intervals for Proportions

Example 9: A random sample of 90 students is selected and 63 are found to take Math110. Construct a 99% confidence interval for the true proportion of students who take Math110.

Confidence Intervals for Proportions

Example 10 : The proportion of students in private schools is around 11%. A random sample of 450 students from a wide geographic area indicated that 55 attended private schools. Estimate the true proportion of students attending private schools with 95% confidence. How does your estimate compare to 11%?

Sample size

If $\hat{p}$ is used as an estimate for p, we can be (1 – α)100% , for example 95%, 98% ,99%, confident that the error will not exceed E where $E = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$. and the formula for the sample size is given by:

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

Final sample size always round up!!!

Note: When you have absolutely no idea about the true population proportion (p), you should assume that $p = 0.5$.

Sample Size (Cont.)

Example 11: A consumer wishes to estimate the proportion of processed food items that contain genetically modified (GM) products.

(a) If no preliminary study is available, how large sample size is need to be 95% confident the estimate is within 3% of the population proportion?

(b) In a preliminary study, 210 of 350 processed items contained GM products. How large a sample is need to construct a 95% confidence interval within 3% of the population proportion?

Confidence Intervals for Proportions and Sample Size

Example 12: A survey of women who are the main meal preparers in their households resulted in 86% know that cholesterol is a health problem. Suppose the survey consisted in 750 women.

(a) Find 90% confidence interval for the population proportion of women in this category who know cholesterol is a health problem.

(b) At the same level of confidence, what sample size would be required to decrease the error to within 1% of the true proportion?

Estimating μ:

σ is known		σ is unknown	
$n \geq 30$	$n < 30$	$n \geq 30$	$n < 30$
$\bar{X} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	$\bar{X} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	$\bar{X} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	$\bar{X} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$

Population values are assumed to be normally distributed

Estimating p (when $np \geq 5$ and $nq \geq 5$) :

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$