Lecture #7 **Chapter 7: Estimates and sample sizes**

In this chapter, we will learn an important technique of statistical inference to use sample statistics to estimate the value of an unknown population parameter.

7-2 Estimating a population proportion

Recall: A point estimate is a single value estimate for a population parameter. The most unbiased point estimate of the population proportion is the sample proportion, $\hat{p}$.

An **interval estimate (confidence interval)** is an interval, or range of values, used to estimate a population parameter. For example 0.476<p<0.544

The **level of confidence (1-α)** is the probability that the interval estimate contains the population parameter. For example, we are 90% confident that the above interval contains the true value of p.

"We are 90% confident" means that if we were to select many different samples of size n and construct the confidence interval, 90% of them actually contain the value of the population proportion p.

We know from the central limit theorem that when n>30, the sampling distribution of sample proportion is a normal distribution. The level of confidence (**1-**α) is the area under the standard normal curve between the **critical values** -$z_{\alpha/2}$ and $z_{\alpha/2}$.

Critical values are values that separate sample statistics that are probable from sample statistics that are improbable, or unusual. (**1-**α) is the percent of the area under the normal curve between - $z_{\alpha/2}$ and $z_{\alpha/2}$.

For example, if (**1-**α) $= \mathbf{0.90}$, then $\alpha = \mathbf{0.10}$ and $\alpha/2$**=0.05.** 5% of the area lies to the left of $-z_{\alpha/2}$=-1.645 and 5% lies to the right of $z_{\alpha/2}$=1.645.

Example 1: Find the critical value $z_{\alpha/2}$ corresponding to the given degree of confidence. a) 99% b) 97%

The **margin of error**, denoted by **E**, is the greatest possible distance between the observed sample proportion $\hat{p}$ and the true value of the population proportion p.

$$E = z_{\alpha/2} \cdot \sigma_{\hat{p}} = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

Thus a (**1-**$\boldsymbol{\alpha}$) confidence interval for the population proportion is $\hat{p}$ –E <p <$\hat{p}$ +E.

Round-off rule for confidence interval estimate of p: Round the confidence interval limits for p to 3 significant digits.

Guide line for constructing a confidence interval for a population proportion:

1. Identify the sample statistics n and x.
2. Find the point estimate $\hat{p} = \frac{x}{n}$
3. Verify that the sampling distribution of $\hat{p}$ can be approximated by the normal distribution $n\hat{p} \geq 5$, $n\hat{q} \geq 5$.
4. Find the critical value $z_{\alpha/2}$ that corresponds to the given level of confidence (**1-**$\boldsymbol{\alpha}$) .
5. Find the margin of error E. $E = z_{\alpha/2} \cdot \sigma_{\hat{p}} = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$
6. Find the left and right end points and form the confidence interval.
   $\hat{p}$ –E <p <$\hat{p}$ +E

Example 2: 829 adult were surveyed in one city, and 51% of them are opposed to the use of the photo-cop for issuing traffic ticket. A) Find the best point estimate of the proportion of all adults in that city opposed to photo-cop use? B) Construct a 95% confidence interval for the proportion of adults who opposed to photo-cop use? C) base on the results, can we safely conclude that the majority of adult oppose use of the photo-cop?

Determining Sample size:

Note: E = $z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$ Solve this formula for n, $n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 \cdot \hat{p}\cdot\hat{q}}{E^2}$ when $\hat{p}$ ***is known.***

$n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 \cdot 0.25}{E^2}$ when $\hat{p}$ **is not known.**

Round-off rule for sample size n: when necessary, round up to obtain the next whole number.

Example 3: a sociologist wishes to estimate the percentage of the U.S population living in poverty. What size sample should be obtained if she whishes the estimate to be within 2 percentage points with 99% confidence a) if she uses the 2003 estimate of 12.7% obtained from the American Community Survey. b) If she no prior information suggesting a possible value of p.

Finding the point estimate and E form a confidence interval:

If we already know the confidence interval limits from either a journal article, or ti might have been generated using software or a calculator, then the sample proportion $\hat{p}$ and the margin of error E can found as follows:

$\hat{p} = \frac{\text{(upper confidence limit)}+\text{(lower confidence limit)}}{2}$,

$E = \frac{(upper\ confidence\ limit)-(lower\ confidence\ limit)}{2}$

Try it yourself: #29 on section 7-2

7-3 and 7-4 Estimating a population mean: I) σ *is known* **II)** *is not known*

In this two section, you will learn how to sue sample statistics to make an estimate of the population parameter μ.

Guide line for finding a confidence interval for population mean

($n > 30$ *or the population is normally distributed.*)

1. Find the sample statistics n and $\bar{x}$.
2. Specify σ if known. Otherwise, if n >30, find the sample standard deviation s and use it as an estimate for σ. $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
3. Find the critical value $z_{\frac{\alpha}{2}}$ that corresponds to the given level of confidence.
4. Find the margin of error E. $E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{n}$
5. Form the confidence interval.

Example 4: Starting salaries of college graduates who have taken a statistic course: n =28, $\bar{x} = \$45{,}678$, the population is normally distributed and $\sigma = \$9900$. Find a 95% confidence interval for estimating the population mean.

Round-off rule for confidence intervals used to estimate μ:

a) **If the original set of data in known, round the confidence interval limits to one more decimal place than is used for the original set of data.**
b) **When the original set of data is unknown, round the confidence interval limits to the same number of decimal places used for the sample mean.**

Example 5: A sample of 54 bears has a mean weight of 182.9 lb. Assuming that σ is known to be 121.8 lb, find a 99% confidence interval estimate of the mean of the population of all such bear weights.

Sample size for estimating mean μ:

Given a degree of confidence and a margin of error E, the minimum sample size, n, needed to estimate the population mean μ *is* $n = \left(\frac{\sigma \cdot z_{\frac{\alpha}{2}}}{E}\right)^2$

Example 6: an economist wants to estimate the mean income for the first year of work for college graduates who have taken a statistic course. How many such incomes must be found if we want to be 95% confident that the sample mean is within \$500 of the true population mean? Assume that a previous study has revealed that for such incomes, $\sigma = \$6250$.

Round-off rule for sample size n: when necessary, round up to obtain the next whole number.

Estimating a population mean: $\sigma\ is\ not\ known$

In many real-life situations, the population standard deviation is unknown. If the random variable is normally distributed (or approximately normally distributed), then we will use a t-distribution.

Def: If the distribution of a random variable x is approximately normal, then $t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ follows a t-distribution.

Critical values of t are denoted by $t_{\alpha/2}$. Several properties of the t-distribution are as follows:

1. The t-distribution is bell-shaped and symmetric about the mean.
2. The t-distribution is a family of curves, each determined by a parameter called the degrees of freedom. The **degrees of freedom** are the number of free choices left after a sample statistic such as $\bar{x}$ is calculated. Degrees of freedom = n -1
3. The total area under a t-curve is 1 or 100%.
4. The mean, median, and mode of the t-distribution are equal to zero.
5. The standard deviation of t distribution varies with the sample size, but it is greater than 1 (unlike the standard normal distribution, which has standard deviation of 1).
6. As the sample size n gets larger, the t-distribution gets closer to the standard normal distribution.

A value of $t_{\alpha/2}$ can be found in table A-3 by locating the appropriate number of degrees of freedom in the left column and proceeding across the corresponding row until reaching the number directly below the applicable value of α for area of two tails.

Example 7: find the critical value, $t_{\alpha/2}$ for a 95% confidence when the sample size is 15.

Example 8: find the critical value, $t_{\alpha/2}$ for a 90% confidence when the sample size is 22.

Guidelines: Constructing a confidence interval for μ (with σ is unknown)

(The population appears to be normally distributed or n > 30)

1. Find the sample statistics n and $\bar{x}$ *and* s.

   Sample standard deviation: $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

2. Identify the degrees of freedom, the level of confidence 1-α, and the critical value $t_{\alpha/2}$.
3. Find the margin of error E. $E = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$
4. Form the confidence interval .

Example 9: You randomly select 16 restaurants and measure the temperature of the coffee sold at each. The sample mean temperature is 162^0 F with a sample standard deviation of 10^0 F. Find the 95% confidence interval for the mean temperature. Assume the temperatures are approximately normally distributed.

Example 10: You randomly select 20 mortgage institutions and determine the current mortgage interest rate at each. The sample mean rate is 6.93% with a sample standard deviation of 0.42%. Find the 99% confidence interval for the mean mortgage interest rate. Assume the interest rates are approximately normally distributed.

Choosing the Appropriate Distribution:

Example 11: determine whether the margin of error E should be calculated using a critical value from the normal distribution, a critical value of from a t-distribution, or neither.

a) n = 150, $\bar{x}$=100, s =15, and the population has a skewed distribution.
b) n = 8, $\bar{x}$=100, s = 5, and the population has a normal distribution.
c) n=8, $\bar{x}$=100, s = 15, and the population has a very skewed distribution
d) n = 150, $\bar{x}$=100, $\sigma = 15$, and the distribution skewed
e) n = 8, $\bar{x}$=100, $\sigma = 15$, and the distribution is extremely skewed.