Mat 131 Agenda  Day 34:  11/20/02

- Section 6.4 Determining Sample Size
- Section 6.5 Estimating a Population Proportion
- Review problems from Chapter 6

Homework:  Triola 6.4 & 6.5
Sample Size for Estimating Mean $\mu$

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

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

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

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

$$n = \left( \frac{2 \sigma}{E} \right)^2$$
Sample Size for Estimating Mean $\mu$

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

(solve for $n$ by algebra)

$$n = \left[ \frac{z_{\alpha/2} \sigma}{E} \right]^2$$

$z_{\alpha/2} = \text{critical z score based on the desired degree of confidence}$

$E = \text{desired margin of error}$

$\sigma = \text{population standard deviation}$
Round-Off Rule for Sample Size $n$

When finding the sample size $n$, if the use of Formula 6-3 does not result in a whole number, always increase the value of $n$ to the next larger whole number.
Round-Off Rule for Sample Size $n$

When finding the sample size $n$, if the use of Formula 6-3 does not result in a whole number, always increase the value of $n$ to the next larger whole number.

$$n = 216.09 = 217 \ (\text{rounded up})$$
Example: If we want to estimate the mean weight of plastic discarded by households in one week, how many households must be randomly selected to be 99% confident that the sample mean is within 0.25 lb of the true population mean? (A previous study indicates the standard deviation is 1.065 lb.)

\[ E = 0.25 \]
\[ S = 1.065 \]

\[ n = \left( \frac{Z_{\alpha/2} \cdot S}{E} \right)^2 \]

\[ n = \left( \frac{2.576 \cdot 1.065}{0.25} \right)^2 \]

\[ n = 120.42 \]

Round up \( \Rightarrow 121 = n \)
Example: If we want to estimate the mean weight of plastic discarded by households in one week, how many households must be randomly selected to be 99% confident that the sample mean is within 0.25 lb of the true population mean? (A previous study indicates the standard deviation is 1.065 lb.)

\[ \alpha = 0.01 \]
\[ z_{\alpha/2} = 2.575 \]
\[ E = 0.25 \]
\[ s = 1.065 \]
**Example:** If we want to estimate the mean weight of plastic discarded by households in one week, how many households must be randomly selected to be 99% confident that the sample mean is within 0.25 lb of the true population mean? (A previous study indicates the standard deviation is 1.065 lb.)

\[ \alpha = 0.01 \]
\[ z_{\alpha/2} = 2.575 \]
\[ E = 0.25 \]
\[ s = 1.065 \]

\[ n = \left( \frac{z_{\alpha/2} \sigma}{E} \right)^2 = \left( \frac{(2.575)(1.065)}{0.25} \right)^2 \]

\[ = 120.3 = 121 \text{ households} \]
Example: If we want to estimate the mean weight of plastic discarded by households in one week, how many households must be randomly selected to be 99% confident that the sample mean is within 0.25 lb of the true population mean? (A previous study indicates the standard deviation is 1.065 lb.)

\[ n = \left( \frac{z_{\alpha/2} \sigma}{E} \right)^2 = \left( \frac{2.575 \times 1.065}{0.25} \right)^2 \]

\[ = 120.3 = 121 \text{ households} \]

If \( n \) is not a whole number, round it up to the next higher whole number.
Example: If we want to estimate the mean weight of plastic discarded by households in one week, how many households must be randomly selected to be 99% confident that the sample mean is within 0.25 lb of the true population mean? (A previous study indicates the standard deviation is 1.065 lb.)

\[
\alpha = 0.01 \\
z_{\alpha/2} = 2.575 \\
E = 0.25 \\
s = 1.065
\]

\[
\begin{align*}
n &= \left[ \frac{z_{\alpha/2} \sigma}{E} \right]^2 = \left[ \frac{(2.575)(1.065)}{0.25} \right]^2 \\
&= 120.3 = 121 \text{ households}
\end{align*}
\]

We would need to randomly select 121 households and obtain the average weight of plastic discarded in one week. We would be 99% confident that this mean is within 1/4 lb of the population mean.
What if $\sigma$ is Not Known?

1. Use the range rule of thumb to estimate the standard deviation as follows: $\sigma \approx \frac{\text{range}}{4}$
What if $\sigma$ is Not Known?

1. Use the range rule of thumb to estimate the standard deviation as follows: $\sigma \approx \frac{\text{range}}{4}$

2. Conduct a pilot study by starting the sampling process. Based on the first collection of at least 31 randomly selected sample values, calculate the sample standard deviation $s$ and use it in place of $\sigma$. That value can be refined as more sample data are obtained.
What if $\sigma$ is Not Known?

1. Use the range rule of thumb to estimate the standard deviation as follows: $\sigma \approx \frac{\text{range}}{4}$

2. Conduct a pilot study by starting the sampling process. Based on the first collection of at least 31 randomly selected sample values, calculate the sample standard deviation $s$ and use it in place of $\sigma$. That value can be refined as more sample data are obtained.

3. Estimate the value of $\sigma$ by using the results of some other study that was done earlier.
What happens when E is doubled?
What happens when $E$ is doubled?

$E = 1$:
$$n = \left( \frac{Z_{\alpha/2} \sigma}{1} \right)^2 = \frac{\left( Z_{\alpha/2} \sigma \right)^2}{1}$$

$E = 2$:
$$n = \left( \frac{Z_{\alpha/2} \sigma}{2} \right)^2 = \frac{\left( Z_{\alpha/2} \sigma \right)^2}{4}$$

\[\frac{3}{9}\]

*Sample size $n$ is decreased to $1/4$ of its original value if $E$ is doubled.*
What happens when $E$ is doubled?

$E = 1: \quad n = \left[ \frac{Z_{\alpha/2} \sigma}{1} \right]^2 = \frac{(Z_{\alpha/2} \sigma)^2}{1}$

$E = 2: \quad n = \left[ \frac{Z_{\alpha/2} \sigma}{2} \right]^2 = \frac{(Z_{\alpha/2} \sigma)^2}{4}$

- Sample size $n$ is decreased to $1/4$ of its original value if $E$ is doubled.

- Larger errors allow smaller samples.
What happens when E is doubled?

$E = 1: \quad n = \left[ \frac{z_{\alpha/2} \sigma}{1} \right]^2 = \frac{(z_{\alpha/2} \sigma)^2}{1}$

$E = 2: \quad n = \left[ \frac{z_{\alpha/2} \sigma}{2} \right]^2 = \frac{(z_{\alpha/2} \sigma)^2}{4}$

- Sample size $n$ is decreased to $1/4$ of its original value if $E$ is doubled.

- Larger errors allow smaller samples.

- Smaller errors require larger samples.
Margin of Error $E$

The margin of Error, $E$, is the maximum (with probability $1-\alpha$) likely difference between the observed sample mean $\bar{x}$ and the true value of the population mean $\mu$.

If $n \geq 30$ or $\sigma$ is known use:

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

If $n < 30$ or $\sigma$ is unknown use:

$$E = \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}$$

For a proportion use:

$$E = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$
In a binomial distribution the mean \( \mu = np \); standard deviation \( \sigma = \sqrt{npq} \).

Let us go back to the definition of \( z \): \( z = \frac{x - \mu}{\sigma} \).

Substitute the mean and the standard deviation from the binomial distribution:

\[ z = \frac{x - np}{\sqrt{npq}} \]

Divide both numerator and denominator by \( n \) (remember \( n = \sqrt{n^2} \)):
\[ z = \frac{(x-np)/n}{(\sqrt{npq})/\sqrt{n^2}} = \]
\[ = \frac{x-np}{\sqrt{npq}} \cdot \frac{n}{\sqrt{n^2}} = \frac{x-p}{\sqrt{pq}} \cdot \frac{1}{\sqrt{n}} \]

The standard deviation of the sampling distribution of \( \hat{p} \) is
\[ \sqrt{\frac{pq}{n}} \quad \text{that is} \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \]
When $n$ is large, $\hat{p}$ can approximate the value of $p$ in the formula for $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

\[ n\hat{p} \geq 5 \]
\[ n\hat{q} \geq 5 \]
Confidence Interval for Population Proportion, \( p \)

Recall from Chapter 5, that a binomial distribution can be approximated by the normal distribution if \( np \geq 5 \) and \( nq \geq 5 \). The mean is given by \( p \) and the standard deviation is

\[
\sigma = \sqrt{\frac{pq}{n}}.
\]

With this in mind, the margin of error is

\[
E = z_{\alpha/2} \sqrt{\frac{pq}{n}}; \text{ the confidence interval is } \hat{p} - E < p < \hat{p} + E
\]
Exercise: Page 317. In a survey of 1068 American, 673 stated that they had answering machines. Find the 95% confidence interval estimate of the population proportion of all Americans who have answering machines.

\[
\hat{p} = \frac{673}{1068} = 0.630
\]

\[
\hat{q} = 1 - \hat{p} = 1 - 0.630 = 0.37
\]

\[
E = 2 \cdot \frac{\alpha}{2} \sqrt{\frac{\hat{p} \hat{q}}{n}}
\]

\[
= 1.96 \sqrt{\frac{(0.63)(0.37)}{1068}}
\]

\[
= 0.0298
\]

\[
CI = 95\% = 0.95
\]

\[
\alpha = 0.05
\]

\[
\alpha/2 = 0.025
\]

\[
Z = 1.96
\]

\[
\frac{0.5}{2} = 0.25
\]

\[
\frac{0.5}{2} = 0.25
\]

\[
\frac{0.5}{2} = 0.25
\]

\[
\frac{0.5}{2} = 0.25
\]
\[ \hat{p} - E < p < \hat{p} + E \]

\[ 0.630 - 0.029 < p < 0.630 + 0.029 \]

\[ 0.601 < p < 0.659 \]

\[ \hat{p} = \frac{\chi}{n} = 0.659 \]

\[ E = 0.0289 = \frac{0.579}{2} \]

1-PropZInt
\[ (.6012, .6591) \]
\[ \hat{p} = 0.6301498127 \]
\[ n = 1068 \]
1 - PropzInt
(0.6012, 0.6591)
\( \hat{p} = 0.6301498127 \)
n = 1068
\[ \hat{p} = \frac{673}{1068} = .63 \]
\[ \hat{q} = 1 - \hat{p} = 1 - .63 = .37 \]
\[ E = z_{a/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \]

\[ E = 1.96 \sqrt{\frac{.63 \times .37}{1068}} = .029 \]

\[ .63 - E < p < .63 + E \]

\[ .63 - .029 < p < .63 + .029 \]

\[ .601 < p < .659 \]
There is 95% probability that the proportion interval (.601, .659) or the interval between 60.1% and 65.9% contains the population proportion of all American answering machines.

Use the calculator to find the Confidence Interval.

EDIT CALC TESTS
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
8:TInterval...
9:2-SampZInt...
0:2-SampTInt...
0:1-PropZInt...

1-PropZInt
x: 673
n: 1068
C-Level:.95

1-PropZInt
(.6012, .6591)
\( \hat{p} = 0.6301498127 \)
n=1068
Sample Size
\[ E = z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \]

\[ \left( E = z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \right)^2 \]

\[ E^2 = z_{\alpha/2}^2 \times \frac{\sigma^2}{n} \]

\[ \frac{E^2}{1} = \frac{z_{\alpha/2}^2}{1} \times \frac{\sigma^2}{n} \]

\[ E^2 = z_{\alpha/2}^2 \sigma^2 \]

\[ \frac{E^2}{1} = \frac{z_{\alpha/2}^2 \sigma^2}{n} \]

\[ nE^2 = \frac{z_{\alpha/2}^2 \sigma^2}{n} \times n \]

\[ nE^2 = z_{\alpha/2}^2 \sigma^2 \]

\[ n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} \]

\[ n = \left( \frac{z_{\alpha/2}^2 \sigma^2}{E^2} \right)^2 \]
Exercise: What sample size should you use if you demand 99% Confidence that the sample mean has a margin of error of 2 given that the population standard deviation is 1.4?

\[
\begin{align*}
\alpha &= 0.01 \\
\alpha/2 &= 0.005 \\
E &= 2 \\
\sigma &= 1.4 \\
\text{Sample size} &= 4
\end{align*}
\]

\[
n = \left( \frac{2 \times \sigma/\alpha}{E} \right)^2
\]
\[ n = \frac{Z^2_{\alpha/2} \sigma^2}{E^2} = \left( \frac{Z_{\alpha/2} \sigma}{E} \right)^2 \]

\[ n = \left( \frac{2.575 \times 1.4}{2} \right)^2 = \left( \frac{3.605}{2} \right)^2 = 1.8025^2 = 3.605 \]
Since \( n = 3.605 \) round up to 4; there should be 4 in the sample.
Determining Sample Size

\[ E = z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}} \]

\( \frac{E}{z_{\alpha/2}} = \sqrt{\frac{\hat{p} \hat{q}}{n}} \)

(solve for \( n \) by algebra)

\[ n = \frac{(z_{\alpha/2})^2 \hat{p} \hat{q}}{E^2} \]

\[ nE^2 = \frac{z^2_{\alpha/2} \hat{p} \hat{q}}{E^2} \]
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>9</th>
<th>09</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>
If you are not given an estimate for \( \hat{p} \) or \( \hat{q} \), use .5 for each.

Exercise: Page 320: You want to estimate with a margin of error of three percentage points, the percentage of driver who talk on phone while they are driving. How many drivers must you survey if you want 95% confidence level.

\[ \alpha = .05 \]
\[ \alpha/2 = .025 \]

\[ \frac{\alpha}{2} = .025 \]

\[ \begin{align*}
q &= .82 \\
\sqrt{q} &= .92 \\
\end{align*} \]

a) Use a preliminary estimate of \( \hat{p} = .18 \).

\[ n = \frac{z^2 \hat{p} \hat{q}}{\alpha^2} \]

\[ n = \frac{(1.96)^2 (.18)(.82)}{.03^2} \]

\[ = 630.02 \]

\[ \Rightarrow 631 \]

b) Assume no prior knowledge of \( \hat{p} \).

\[ n = \frac{z^2}{\alpha^2} \]

\[ n = \frac{(1.96)^2}{.03^2} \]

\[ = 1067.1 \]

\[ \Rightarrow 1068 \]
Exercise: Page 320: You want to estimate with a margin of error of three percentage points, the percentage of driver who talk on phone while they are driving. How many drivers must you survey if you want 95% confidence level.

a) Use a preliminary estimate of $\hat{p} = 0.18$.

b) Assume no prior knowledge of $\hat{p}$.
3) Assume no prior knowledge of $p$.

\[ n = \frac{z_{a/2}^2 \cdot \hat{p} \hat{q}}{E^2} \]

\[ n = \frac{(1.96)^2 \cdot .18 \cdot (1 - .18)}{3^2} = 630.0224 \]

Remember to round UP! You should use a sample size of 631 drivers.
\( n = \frac{z_{\alpha/2}^2 \cdot \hat{pq}}{E^2} \)

\( n = \frac{(1.96)^2 \cdot .5 \cdot (1 - .5)}{3^2} \)

\( n = \frac{1.96^2 \cdot .25}{3^2} = 1067.1111 \)

In this case, with no prior knowledge about \( \hat{p} \) hat, you will need 1068 randomly selected drivers.
<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>( \hat{q} )</th>
<th>( \hat{p} \hat{q} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.09</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
<td>0.16</td>
</tr>
<tr>
<td>0.3</td>
<td>0.7</td>
<td>0.21</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>0.24</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
<td>0.24</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>0.21</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2</td>
<td>0.16</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
<td>0.09</td>
</tr>
<tr>
<td>( \hat{p} )</td>
<td>( \hat{q} )</td>
<td>( \hat{p} \hat{q} )</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.09</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
<td>0.16</td>
</tr>
<tr>
<td>0.3</td>
<td>0.7</td>
<td>0.21</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>0.5</strong></td>
<td><strong>0.5</strong></td>
<td><strong>0.25</strong></td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
<td>0.24</td>
</tr>
<tr>
<td>0.7</td>
<td>0.3</td>
<td>0.21</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2</td>
<td>0.16</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Assumptions

1. The sample is a simple random sample.

2. The conditions for the binomial distribution are satisfied (See Section 4-3.)

3. The normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$ and $nq \geq 5$ are both satisfied.
Notation for Proportions

\[ p = \ \text{population proportion} \]

\[ \hat{p} = \frac{x}{n} \ \text{sample proportion of x successes in a sample of size n} \]

(pronounced ‘p-hat’)

\[ \hat{q} = 1 - \hat{p} = \text{sample proportion of x failures in a sample size of n} \]
Definition

Point Estimate

The sample proportion $\hat{p}$ is the best point estimate of the population proportion $p$. 
Margin of Error of the Estimate of $p$

Formula 6-4

$$E = Z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}}$$
Confidence Interval for Population Proportion

\[ \hat{p} - E < p < \hat{p} + E \]

where

\[ E = z_{\alpha/2} \sqrt{\frac{\hat{p} \hat{q}}{n}} \]
Confidence Interval for Population Proportion

\[ \hat{p} - E < p < \hat{p} + E \]

\[ p = \hat{p} \pm E \]

\[ (\hat{p} - E, \hat{p} + E) \]
Round-Off Rule for Confidence Interval Estimates of $p$

Round the confidence interval limits to three significant digits.
Sample Size for Estimating Proportion $p$

When an estimate of $\hat{p}$ is known:

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

Formula 6-5

When no estimate of $p$ is known:

$$n = \left( \frac{Z_{a/2}}{} \right)^2 \frac{0.25}{E^2}$$

Formula 6-6
Two formulas for proportion sample size

\[ n = \frac{(Z_{\alpha/2})^2 \hat{p} \hat{q}}{E^2} \]

\[ n = \frac{(Z_{\alpha/2})^2 (0.25)}{E^2} \]
Example: We want to determine, with a margin of error of four percentage points, the current percentage of U.S. households using e-mail. Assuming that we want 90% confidence in our results, how many households must we survey? A 1997 study indicates 16.9% of U.S. households used e-mail.
Example: We want to determine, with a margin of error of four percentage points, the current percentage of U.S. households using e-mail. Assuming that we want 90% confidence in our results, how many households must we survey? A 1997 study indicates 16.9% of U.S. households used e-mail.

\[
n = \frac{[z_{\alpha/2}]^2 \hat{p} \hat{q}}{E^2}
\]

\[
= \frac{[1.645]^2 (0.169)(0.831)}{0.04^2}
\]

\[
= 237.51965
\]

\[
= 238 \text{ households}
\]

To be 90% confident that our sample percentage is within four percentage points of the true percentage for all households, we should randomly select and survey 238 households.
Example: We want to determine, with a margin of error of four percentage points, the current percentage of U.S. households using e-mail. Assuming that we want 90% confidence in our results, how many households must we survey? There is no prior information suggesting a possible value for the sample percentage.
Example: We want to determine, with a margin of error of four percentage points, the current percentage of U.S. households using e-mail. Assuming that we want 90% confidence in our results, how many households must we survey? There is no prior information suggesting a possible value for the sample percentage.

\[ n = \frac{[z_{\alpha/2}]^2 (0.25)}{E^2} \]

\[ = \frac{(1.645)^2 (0.25)}{0.04^2} \]

\[ = 422.81641 \]

\[ = 423 \text{ households} \]

With no prior information, we need a larger sample to achieve the same results with 90% confidence and an error of no more than 4%.
Finding the Point Estimate and E from a Confidence Interval

Point estimate of $\hat{p}$:

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

Margin of Error:

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