8.1 Estimating a Population Mean

Types of estimators

Point Estimates  Interval Estimates

Point Estimate

A point estimate of a parameter is the

Ex.
Confidence Interval Estimate

with certain percentage of confidence (confidence level).

**Ex.** confidence interval for $\mu$:

\[
L \leq \mu \leq U
\]

Lower Bound \hspace{1cm} Upper Bound

If the probability of the parameter (say $\mu$) being in the interval $(L,U)$ is $1-\alpha$ then the interval is called

<table>
<thead>
<tr>
<th>Confidence Coefficient $(1-\alpha)$</th>
<th>Confidence Interval $100(1-\alpha)%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td></td>
</tr>
</tbody>
</table>
Idea: Assume you have \( m \) samples of equal size.

\[
\begin{align*}
\mu & \\
\text{Sample 1:} & \\
\text{Sample 2:} & \\
\text{Sample 3:} & \\
\vdots & \\
\text{Sample } m: & \\
\end{align*}
\]

How many times do we expect \( \mu \) to be in a \( 100(1-\alpha)\% \) CI?
8.2 Confidence Intervals (CI) for one population mean when \( \sigma \) is known.

Procedure 8.1

The one-sample z-interval procedure for a population mean.

Assumptions:

1. 

2. 
Steps:

Step 1:

For a confidence level of 1 - \( \alpha \), use table II

Step 2:

The CI for \( \mu \) is from

\[ \text{to} \]

Where:

\[ z_{\alpha/2} \] is found in
\[ n \] is the
\[ \bar{x} \] is the
Also written as:

\[
\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

or

\[
\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}
\]
What is $z_{\alpha/2}$ for the different confidence levels?

<table>
<thead>
<tr>
<th>CI</th>
<th>$1-\alpha$</th>
<th>$\alpha/2$</th>
<th>$z_{\alpha/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example

An entomologist sprayed 120 adult Melon flies with a specific low concentration of malathion and observed their survival times. The mean was found to be 18.3 days. Assume $\sigma = 5.2$ days.

a. Use this data to construct a 90% CI for the mean survival time.

$$\bar{x} =$$

$$\sigma =$$

$$n =$$

$$\alpha =$$

Since $1-\alpha =$$

Hence, the CI estimate is:

$$\bar{x} \pm z_{0.10/2} \frac{\sigma}{\sqrt{n}} =$$

$$=$$

$$=$$
b. Find a 95% CI for the mean survival time.

Since $1-\alpha = \begin{align*}
    \bar{x} \pm z_{0.05/2} \frac{\sigma}{\sqrt{n}} &= \\
    &= \\
    &= \\
\end{align*}$

Interpretation: "With 95% confidence we estimate the true mean survival time to be
c. Compare the widths of the 90% and 95% CI's for μ. What effect does increasing the confidence level have on the width of the CI when the mean and SD are fixed?

90% CI :

95% CI :

The 95% CI for μ is . Therefore, the higher the confidence level,
d. If the random sample selected was of 50 flies, find the 95% CI for $\mu$. (Use the mean and sd from part (a)).

$$\bar{x} = 18.3 \quad \sigma = 5.2$$

$$n = \quad \alpha =$$

$$\bar{x} \pm z_{0.05/2} \frac{\sigma}{\sqrt{n}} =$$

$$= $$

$$= $$
e. Compare the widths of the 95% CI's for \( \mu \). What effect does increasing the sample size have on the width of the CI when the mean and sd are the same?

95% CI when \( n=50 \):

95% CI when \( n=120 \):

The 2nd interval is \( \) . The larger the sample size the \( \) (when the confidence level and mean and sd are fixed).
f. If the interval

\[ 18.3 \pm 2.17 \frac{5.2}{\sqrt{50}} \]

is to be used for estimating \( \mu \), what level of confidence are we using?

Hence the interval would be a
8.3 Margin of Error

CI for the mean:

\[ \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]

The margin of error \( E \) for the estimate of \( \mu \) is equal to

\[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \quad \bar{x} \quad \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]
Determining the sample size

What is the sample size required to achieve a desired precision?

Need to specify

1. $E$, the
2. $1 - \alpha$, the

Equate

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

and solve for $n$.

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

$n$ must be a positive integer. (Round up if not).
Example (Continued)

g. What sample size is required to have a margin of error of 1 day when the desired level of confidence is 0.95?

\[ \sigma = 5.2 \]

\[ \alpha = \]

\[ E = \]

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

Then \( Z_{\alpha/2} = \)

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

\[ \rightarrow n = \text{flies (always round up)}. \]
NOTE:

KEY FACT 8.1 When to use the z-interval procedure? (σ is known).

1. For small samples, say, of size less than 15, the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.

2. For moderate size samples, say, between 15 and 30, the z-interval procedure can be used unless the data is “very badly behaved”.

3. For large samples, say, of size 30 or more, the z-interval procedure can be used essentially without restriction.
8.4 Cl for one popul’n mean when $\sigma$ is unknown

Recall: $X \sim N(\mu, \sigma)$ or $n$ large

\[
\begin{align*}
\bar{X} & \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \\
Z & = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \\
\end{align*}
\]

(Standardized version of $\bar{X}$)
Question: What if the popul’n SD is unknown?

\[ t = \frac{\bar{X} - \mu}{\hat{\sigma}} \]

(\text{Studentized version of } \bar{X})

Important: \( t \) does not follow a normal distribution. It follows a

The \textbf{t-dist’n} depends on its degrees of freedom (df) which for the \textbf{t-dist’n} is

\[ df = \]
Basic properties of $t$-curves

1. The total area under a $t$-curve is

2. A $t$-curve is

3. As number of df become larger, $t$-curves look
Using the $t$-table

Notation: $t_\alpha$ denotes the $t$-value having area $\alpha$ to the right, under the $t$-curve.

Example

For a $t$-curve with 15 degrees of freedom, determine $t_{0.10}$. That is, find the $t$-value having an area of 0.1 to its right.

$\rightarrow$ Hence, $t_{0.10} =$

What if the question asks about the $t$-value that has an area to the left of 0.1?

Answer:
The one-sample \( t \)-interval procedure for a population mean:

**Assumptions:**

1. Normal population or large sample
2. \( \sigma \) unknown

**Steps:**

**Step1:**

For a confidence level of \( 1 - \alpha \), use table IV to find \( t_{\alpha/2} \) with \( df=n-1 \), where \( n \) is the sample size.
Step 2:
The CI for $\mu$ is

$$\bar{x} \pm \frac{t_{\alpha/2, n-1}}{s}$$

where:

$$t_{\alpha/2, n-1}$$

$\bar{x}$ and $s$

The CI is exact for normal popul’ns and is approximately correct
Example:

Measurements of the acidity (pH) of rain samples were recorded at 13 sites in an industrial region. The sample mean and standard deviation were 4.546 and 0.391 respectively.

a. determine a 95% CI for the mean acidity of rain in that region.

\[
\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}} =
\]

= 

= 

= 

= 

=
99% → instead of 2.179
90% → instead of 2.179

With 95% confidence the true mean acidity of rain in the region

b. What assumption must be made in order to answer (a)?

Assume that acidity of rain is
Confidence Interval for a single population mean using Excel t-interval

Bicycle Commuting Times. A city planner working on bikeways designs a questionnaire to obtain information about local bicycle commuters. One of the questions asks how long it takes the rider to pedal from home to his or her destination. A sample of local bicycle commuters yields the following times, in minutes

22 19 24 31 29 29 21 15 27 23 37
31 30 26 16 26 12 23 48 22 29 28

Find a 90% confidence interval for the mean commuting time of all local bicycle commuters in the city.

Commands: Enter data into column A
Tools / Data Analysis / Descriptive Statistics / OK

“Data” tab in Excel 2007
Output:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Column1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td></td>
<td>Mean</td>
<td>25.81818182</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td></td>
<td>Standard Error</td>
<td>1.643083844</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td></td>
<td>Median</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td></td>
<td>Mode</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td></td>
<td>Standard Deviation</td>
<td>7.706746356</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td></td>
<td>Sample Variance</td>
<td>59.39393939</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td></td>
<td>Kurtosis</td>
<td>2.315641068</td>
</tr>
<tr>
<td>10</td>
<td>23</td>
<td></td>
<td>Skewness</td>
<td>0.83886406</td>
</tr>
<tr>
<td>11</td>
<td>37</td>
<td></td>
<td>Range</td>
<td>36</td>
</tr>
<tr>
<td>12</td>
<td>31</td>
<td></td>
<td>Minimum</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td></td>
<td>Maximum</td>
<td>48</td>
</tr>
<tr>
<td>14</td>
<td>26</td>
<td></td>
<td>Sum</td>
<td>568</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td></td>
<td>Count</td>
<td>22</td>
</tr>
<tr>
<td>16</td>
<td>26</td>
<td></td>
<td>Confidence Level(90.0%)</td>
<td>2.827325864</td>
</tr>
<tr>
<td>17</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}
\]

\[
25.82 \pm
\]

\[
25.82 \pm
\]

\[
25.82 \pm
\]

\[
(22.99, 28.66)
\]

Interpretation: We are 90% confident that the mean commuting time of local bicycle commuters in the city falls between 22.99 and 28.66 minutes.