Confidence Interval Estimation - Intro

We can use $\bar{X}$ to estimate $\mu$. But it is unlikely that $\bar{X}$ is a perfect estimate of $\mu$. ($\bar{X}$ is a point estimate of $\mu$)

Better to say that $\mu$ is between two numbers, $a$ and $b$, i.e.,

$$a < \mu < b$$
This is called an interval estimate of \( \mu \).

We can state the interval equivalently as

\[ \bar{X} \pm \text{error term} \]

What is the likelihood that \( \mu \) really is between \( a \) and \( b \)? (Confidence level)

How large is the error term?
95% Confidence Interval for $\mu$ (Large Sample)

Recall: If $n$ is large ($n \geq 20$), $\bar{X}$'s dist. is approx. normal, with mean $\mu_{\bar{X}} = \mu$ and std. dev. $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \sigma_{\bar{X}}$

\[ \mu_{\bar{X}} = 1.96 \sigma_{\bar{X}} \]

\[ = \mu - 1.96 \left( \frac{\sigma}{\sqrt{n}} \right) \]

\[ \mu_{\bar{X}} + 1.96 \sigma_{\bar{X}} \]

\[ = \mu + 1.96 \left( \frac{\sigma}{\sqrt{n}} \right) \]
So 95% of the values of $\bar{x}$ lie between $\mu - 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)$ and $\mu + 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)$.

\[ P(\mu - 1.96 \left( \frac{\sigma}{\sqrt{n}} \right) < x < \mu + 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)) \]

= .95

which can be rearranged

\[ P(\bar{x} - 1.96 \left( \frac{\sigma}{\sqrt{n}} \right) < \mu < \bar{x} + 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)) \]

= .95

So the 95% C.I. goes from $\bar{x} - 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)$ to $\bar{x} + 1.96 \left( \frac{\sigma}{\sqrt{n}} \right)$.
\[ \bar{x} \pm 1.96 \left( \frac{\sigma}{\sqrt{n}} \right) \]

Since we usually don't know \( \sigma \), we use \( s \).

\[ \bar{x} - 1.96 \left( \frac{s}{\sqrt{n}} \right) \text{ to } \bar{x} + 1.96 \left( \frac{s}{\sqrt{n}} \right) \]

Error term

Error term

**Ex:** Construct a 95\% C.I. for mean weight of full grown dogs

\[ \bar{x} = 44 \text{ lbs.} \]

\[ s = 8 \text{ lbs.} \]

\[ n = 36 \]
\[ \bar{x} \pm 1.96 \left( \frac{s}{\sqrt{n}} \right) \]

\[ = 44 \pm 1.96 \left( \frac{9}{\sqrt{36}} \right) \]

\[ = 44 \pm 2.61 \]

So our 95\% C.I is from 41.39 lbs to 46.61 lbs.

We are 95\% confident that the mean weight of all full grown dogs is between 41.39 lbs and 46.61 lbs.
95% is called the confidence level.

Interpretation - Pg 429

Note: Other confidence levels are sometimes used by researchers:

90%  
Use $z = 1.645$

99%  
Use $z = 2.578$

$$\bar{x} \pm \text{critical value} \times \frac{s}{\sqrt{n}}$$

$$\bar{x} \pm z \left( \frac{s}{\sqrt{n}} \right)$$
The error term depends on
1) the sample size $n$
   (the larger $n$ is, the smaller the error term)
2) the std. dev. $s$
   (the smaller $s$ is, the smaller the error term)
3) the confidence level
   (the higher the confidence level, the larger the crit. value is; so the larger the error term is)
Sample Size for Estimating $\mu$

The size, $n$, of the sample depends on

1) Variability of the data (std. dev. $s$)

2) Precision of the estimate, i.e., the size of the error term

$z \left( \frac{s}{\sqrt{n}} \right)$
Ex: Suppose we wish to estimate \( \mu \) with a precision of 2 units (i.e., the error term is 2) with 95% confidence. Also suppose we know \( s = 12.6 \).
Find the sample size required.

We must find \( n \) such that

\[
Z \left( \frac{s}{\sqrt{n}} \right) = 2
\]

Assuming that \( n \) is "large" for a 95% C.I.

\[
Z = 1.96
\]

So

\[
1.96 \times \frac{s}{\sqrt{n}} = 2
\]
\[ 1.96 \left( \frac{12.6}{\sqrt{n}} \right) = 2 \]

Solving for \( n \) yields
\[ n = 152.47 \]

which we round up to
\[ n = 153 \text{ obs.} \]

We need at least 153 obs.