STA 4033
Mathematical Statistics
with Computer Applications
Lecture 10*

Brett Presnell
Department of Statistics
University of Florida

30 May 2001

*Copyright ©2001 by Brett Presnell
3.2. Confidence Intervals

3.2.1. Large Sample Confidence Intervals

Point estimation alone is not a very sophisticated form of statistical inference, since no information is given about the likely error in the estimates. However, when the sample size is large, the sampling distribution of many estimators

- is approximately normal,
- with negligible bias.

This implies that

$$P \left( -\frac{z_{\alpha}}{2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq \frac{z_{\alpha}}{2} \right) \approx 1 - \alpha,$$

where $z_{\alpha}$ is the upper $\alpha$th quantile of the standard normal distribution, i.e., if $Z \sim N(0, 1)$, then $P(Z > z_{\alpha}) = \alpha$.

Rearranging we obtain

$$P \left( \frac{\hat{\theta} - z_{\alpha}/2 \sigma_{\hat{\theta}}}{\sigma_{\hat{\theta}}} \leq \frac{\theta}{\hat{\theta}} \leq \frac{\hat{\theta} + z_{\alpha}/2 \sigma_{\hat{\theta}}}{\sigma_{\hat{\theta}}} \right) \approx 1 - \alpha.$$

This means that the interval

$$\left[ \hat{\theta} - \frac{z_{\alpha}}{2} \sigma_{\hat{\theta}}, \hat{\theta} + \frac{z_{\alpha}}{2} \sigma_{\hat{\theta}} \right]$$

(or $\hat{\theta} \pm \frac{z_{\alpha}}{2} \sigma_{\hat{\theta}}$ for short) is an approximate $(1 - \alpha) \times 100\%$ confidence interval for $\theta$. A verbal way of expressing this formula is

estimator $\pm$ (table value) $\times$ (standard error of estimator)
This works for all of the estimators in Table 1 from the previous lecture. Note that in practice the standard error of $\hat{\theta}$ is generally not known and must be estimated (see Table 2 from the previous lecture).

Note that $z_{\alpha/2}$ can be obtained from either

- a table of the standard normal distribution, or
- the `qnorm()` function in R.

It is common to compute 95% CIs, which use $z_{0.025} = 1.96$. Since $1.96 \approx 2$,

- $\hat{\theta} \pm 2\sigma_{\hat{\theta}}$ is roughly a 95% CI, and
- $2\sigma_{\hat{\theta}}$ is sometimes given as a *bound on the error in estimation* for the point estimator $\hat{\theta}$.

Note well, this only bounds the error in estimation with high (95%) probability, not absolutely.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimator</th>
<th>Std Err</th>
<th>Estimated Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\bar{X}$</td>
<td>$\frac{\sigma}{\sqrt{n}}$</td>
<td>$\frac{s}{\sqrt{n}}$</td>
</tr>
<tr>
<td>$p$</td>
<td>$\hat{p} = \frac{X}{n}$</td>
<td>$\sqrt{\frac{pq}{n}}$</td>
<td>$\sqrt{\frac{\hat{p}\hat{q}}{n}}$</td>
</tr>
<tr>
<td>$\mu_1 - \mu_2$</td>
<td>$\bar{X}_1 - \bar{X}_2$</td>
<td>$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$</td>
<td>$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$</td>
</tr>
<tr>
<td>$p_1 - p_2$</td>
<td>$\hat{p}_1 - \hat{p}_2$</td>
<td>$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$</td>
<td>$\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$</td>
</tr>
</tbody>
</table>

Table 1: Some common estimators and their standard errors.
3.2.2. Small Sample Confidence Intervals

When the sample size is small, CIs based on a large sample approximation may not work well. In some cases small sample confidence intervals can be derived. These often require more stringent assumptions about the population distribution than the large sample interval.

The best known example is the small sample confidence interval for the mean, $\mu$, given by

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

where $t_{\alpha}$ is the upper $\alpha$th quantile of a $t$ distribution with $n - 1$ d.f.

This interval is based on the assumption that the population distribution is normal (at least approximately).

If this the population distribution is not normal, then the $t$-interval may not have the desired coverage probability.
3.2.3. Interpretation

Note that the endpoints of a confidence interval are statistics calculated from sample data. Hence they vary from sample to sample.

Let \( \hat{\theta}_L \) and \( \hat{\theta}_U \) denote the endpoints of a \((1 - \alpha) \times 100\%\) CI for the parameter \( \theta \). Then by definition,

\[
P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha.
\]

Note that in this statement, \( \hat{\theta}_L \) and \( \hat{\theta}_U \) are random and \( \theta \) is a fixed (but unknown) constant.

This means that

- \((1 - \alpha) \times 100\%\) of all samples would yield an interval \([\hat{\theta}_L, \hat{\theta}_U]\) containing \( \theta \), and
- \(\alpha \times 100\%\) of all samples would yield an interval not containing \( \theta \).

In practice

- we draw only one sample,
- we compute the interval \([\hat{\theta}_L, \hat{\theta}_U]\) for our sample, and
- we say that we are \((1 - \alpha) \times 100\%\) confident that \( \theta \) lies in the interval thus computed.
- No probability statements are possible about the particular interval computed for our single sample.
Example  For large sample sizes, we know that $\bar{X} \pm 1.96 \frac{s}{\sqrt{n}}$ is an approximate 95% confidence interval for the population mean $\mu$, i.e.,

$$P\left(\bar{X} - 1.96 \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{s}{\sqrt{n}}\right) \approx .95.$$

Suppose we draw a random sample of size $n = 64$ and observe $\bar{x} = 100$ and $s = 16$. Then

$$100 \pm 1.96 \frac{16}{\sqrt{64}} = 100 \pm 3.92 \quad \text{or} \quad [96.18, 103.92]$$

is 95% CI for $\mu$.

- This does **NOT** mean that

$$P(96.18 \leq \mu \leq 103.92) = .95 \quad \text{FALSE!!!}$$

since the statement $96.18 \leq \mu \leq 103.92$ is either true (i.e., has probability 1) or false (i.e., has probability 0).

- It does mean that the interval $[96.18, 103.92]$ was calculated in such a way that for 95% of all samples the interval so calculated will contain $\mu$.

See R transcript for this lecture.