Statistics 300: Elementary Statistics
Sections 7-2, 7-3, 7-4, 7-5

Parameter Estimation
• Point Estimate
  – Best single value to use

• Question
  – What is the probability this estimate is the correct value?

• Answer
  – zero: assuming “x” is a continuous random variable
  – Example for Uniform Distribution
If $X \sim U[100,500]$ then

- $P(x = 300) = \frac{(300-300)/(500-100)}{= 0}$

Parameter Estimation

- Pop. mean $\mu$
  - Sample mean $\bar{x}$
- Pop. proportion $p$
  - Sample proportion $\hat{p}$
- Pop. standard deviation $\sigma$
  - Sample standard deviation $s$

Problem with Point Estimates

- The unknown parameter ($\mu$, $p$, etc.) is not exactly equal to our sample-based point estimate.
- So, how far away might it be?
- An interval estimate answers this question.
Confidence Interval

• A range of values that contains the true value of the population parameter with a ...
• Specified “level of confidence”.
• [L(ower limit),U(pper limit)]

Terminology

• Confidence Level (a.k.a. Degree of Confidence)
  – expressed as a percent (%)
• Critical Values (a.k.a. Confidence Coefficients)

Terminology

• “alpha” “\(\alpha\)” = 1-Confidence
  – more about \(\alpha\) in Chapter 7
• Critical values
  – express the confidence level
Confidence Interval for $\mu$
If $\sigma$ is known (this is a rare situation)

$$= \bar{x} \pm E$$

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

Confidence Interval for $\mu$
If $\sigma$ is known (this is a rare situation)
if $x \sim N(\mu, \sigma)$

$$= \bar{x} \pm z_{\alpha/2} \cdot \left( \frac{\sigma}{\sqrt{n}} \right)$$

Why does the Confidence Interval for $\mu$ look like this?

$$= \bar{x} \pm z_{\alpha/2} \cdot \left( \frac{\sigma}{\sqrt{n}} \right)$$
\[ \bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \]

make an \( \bar{x} \) value

into a \( z \)-score.

The general \( z \)-score

equation is

\[ z = \frac{(x - \mu)}{\sigma} \]

for \( \bar{x} \),

\( \mu_{\bar{x}} \) is \( \mu \) : unchanged

and

\( \sigma_{\bar{x}} \) is \( \frac{\sigma}{\sqrt{n}} \)
so a z-score based on $\bar{x}$ is

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

**Using the Empirical Rule**

Make a probability statement:

$$P \left( -2 < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < 2 \right) = 95\%$$
Check out the “Confidence z-scores” on the WEB page.

(In pdf format.)

Use basic rules of algebra to rearrange the parts of this z-score.

Manipulate the probability statement:

\[ P\left(-2\left(\frac{\sigma}{\sqrt{n}}\right) < (\bar{x} - \mu) < 2\left(\frac{\sigma}{\sqrt{n}}\right)\right) = 0.95 \]
Manipulate the probability statement:

\[ P \left( -\bar{x} - 2 \left( \frac{\sigma}{\sqrt{n}} \right) < \mu < -\bar{x} + 2 \left( \frac{\sigma}{\sqrt{n}} \right) \right) = 0.95 \]

**Confidence = 95%**

\( \alpha = 1 - 95\% = 5\% \)

\( \alpha/2 = 2.5\% = 0.025 \)

---

Manipulate the probability statement:

multiply through by \((-1)\) and change the order of the terms

\[ P \left( \bar{x} - 2 \left( \frac{\sigma}{\sqrt{n}} \right) < \mu < \bar{x} + 2 \left( \frac{\sigma}{\sqrt{n}} \right) \right) = 0.95 \]

**Confidence = 95%**

\( \alpha = 1 - 95\% = 5\% \)

\( \alpha/2 = 2.5\% = 0.025 \)

---

**Confidence Interval for \( \mu \)**

If \( \sigma \) is not known (usual situation)

\[ \bar{x} \pm t_{\alpha/2} \cdot \left( \frac{S}{\sqrt{n}} \right) \]
Sample Size Needed to Estimate $\mu$ within $E$, with Confidence = $1-\alpha$

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

Components of Sample Size Formula when Estimating $\mu$
- $Z_{\alpha/2}$ reflects confidence level
  - standard normal distribution
- $\hat{\sigma}$ is an estimate of $\sigma$, the standard deviation of the pop.
- $E$ is the acceptable “margin of error” when estimating $\mu$

Confidence Interval for $p$
- The Binomial Distribution gives us a starting point for determining the distribution of the sample proportion: $\hat{p}$

$$\hat{p} = \frac{x}{n} = \frac{\text{successes}}{\text{trials}}$$
For Binomial “x”

\[ \mu = np \]

\[ \sigma = \sqrt{npq} \]

For the Sample Proportion

\[ \hat{p} = \frac{x}{n} = \frac{1}{n} (x) \]

x is a random variable
n is a constant

Time Out for a Principle:

If \( \mu \) is the mean of X and “a” is a constant, what is the mean of aX?

Answer: \( a \cdot \mu \)
Apply that Principle!

- Let “a” be equal to “1/n”
- so \( \hat{p} = aX = \left( \frac{1}{n} \right) X = \frac{X}{n} \)
- and \( \mu_{\hat{p}} = a\mu_x = a(np) \)
  \[ = \left( \frac{1}{n} \right) np = p \]

Time Out for another Principle:

If \( \sigma_x^2 \) is the variance of X and “a” is a constant, what is the variance of aX?

Answer: \( \sigma_{aX}^2 = a^2 \sigma_x^2 \)

Apply that Principle!

- Let x be the binomial “x”
- Its variance is npq = np(1-p), which is the square of its standard deviation
Apply that Principle!

• Let “a” be equal to “1/n”

• so \( \hat{p} = aX = \left( \frac{1}{n} \right) X = \frac{X}{n} \)

• and \( \sigma^2_{\hat{p}} = a^2 \sigma^2_X = \left( \frac{1}{n} \right)^2 (npq) \)

Apply that Principle!

\[
\left( \frac{1}{n} \right)^2 npq = \frac{pq}{n} = \sigma^2_{\hat{p}}
\]

and

\[
\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}
\]

When n is Large,

\( \hat{p} \sim N \left( \mu = p, \sigma = \sqrt{\frac{pq}{n}} \right) \)
### What is a Large “n” in this situation?

- Large enough so np > 5
- Large enough so n(1-p) > 5
- Examples:
  - (100)(0.04) = 4  (too small)
  - (1000)(0.01) = 10 (big enough)

### Now make a z-score

\[ z = \frac{\hat{p} - p}{\sqrt{pq/n}} \]

And rearrange for a CI(p)

### Using the Empirical Rule

Make a probability statement:

\[
P \left( -1.96 < \frac{\hat{p} - p}{\sqrt{pq/n}} < 1.96 \right) = 95\%\]
Use basic rules of algebra to rearrange the parts of this z-score.

Manipulate the probability statement:

Step 1: Multiply through by $\sqrt{\frac{pq}{n}}$:

$$P\left(-1.96\sqrt{\frac{pq}{n}} < (\hat{p} - p) < 1.96\sqrt{\frac{pq}{n}}\right) = 0.95$$
Manipulate the probability statement:
Step 2: Subtract $\hat{p}$ from all parts of the expression:

$$P \left( -\hat{p} - 1.96 \sqrt{\frac{pq}{n}} < -p < -\hat{p} + 1.96 \sqrt{\frac{pq}{n}} \right) = 0.95$$

Manipulate the probability statement:
Step 3: Multiply through by -1:
(remember to switch the directions of $< >$)

$$P \left( \hat{p} + 1.96 \sqrt{\frac{pq}{n}} > p > \hat{p} - 1.96 \sqrt{\frac{pq}{n}} \right) = 0.95$$

Manipulate the probability statement:
Step 4: Swap the left and right sides to put in conventional $p < \hat{p}$ form:

$$P \left( \hat{p} - 1.96 \sqrt{\frac{pq}{n}} < p < \hat{p} + 1.96 \sqrt{\frac{pq}{n}} \right) = 0.95$$
Confidence Interval for p
(but the unknown p is in the formula. What can we do?)

\[ \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \]

Confidence Interval for p
(substitute sample statistic for p)

\[ \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \]

Sample Size Needed to Estimate “p” within E, with Confid.=1-\( \alpha \)

\[ n = \left( \frac{Z^2_{\alpha/2}}{E^2} \right) \cdot \hat{p}\hat{q} \]
Components of Sample Size Formula when Estimating “p”

- $Z_{a/2}$ is based on $\alpha$ using the standard normal distribution
- $p$ and $q$ are estimates of the population proportions of “successes” and “failures”
- $E$ is the acceptable “margin of error” when estimating $\mu$

Components of Sample Size Formula when Estimating “p”

- $p$ and $q$ are estimates of the population proportions of “successes” and “failures”
- Use relevant information to estimate $p$ and $q$ if available
- Otherwise, use $p = q = 0.5$, so the product $pq = 0.25$

Confidence Interval for $\sigma$
starts with this fact
if $x \sim N(\mu, \sigma)$
then

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2 \text{ (chi square)}$$
What have we studied already that connects with Chi-square random values?

\[
\frac{(n-1)s^2}{\sigma^2} \sim \chi^2 \text{ (chi square)}
\]

\[
\frac{(n-1)s^2}{\sigma^2} = \frac{(n-1) \sum (x-\mu)^2}{\sigma^2}
\]

\[
= \sum \frac{(x-\mu)^2}{\sigma^2}
\]

\[
\sum \left[ \frac{(x-\mu)^2}{\sigma^2} \right] = \sum \left[ \frac{(x-\mu)}{\sigma} \right]^2
\]

\[
= \sum z^2 \quad \text{a sum of squared standard normal values}
\]
Confidence Interval for $\sigma$

\[
LB = \sqrt{\frac{(n-1)s^2}{\chi^2_R}}
\]

\[
UB = \sqrt{\frac{(n-1)s^2}{\chi^2_L}}
\]