Chapter 7
March 23, 2014

Statistical Inference

7.1 Basic Properties of Confidence Intervals

Chapter Overview

- Basics Confidence Intervals (C.I.)
- Large-Sample Confidence Intervals for Population Mean and Proportion
  - C.I. for mean \( \mu \)
  - C.I. for proportion
  - One-sided intervals
- Intervals Based on a Normal Population Distribution for mean
  - \( t \) distribution
  - One sample \( t \) C.I.
- Confidence Intervals for the Variance and Standard Deviation of a Normal Population

What is a Confidence Interval? Point Estimate vs. Confidence Interval

To estimate a parameter of a population. Say \( \mu \) of a normal distribution. Given the observed value \( x_1, x_2, \ldots, x_n \) of a random sample \( X_1, \ldots, X_n \). We can:

- Find an point estimate of \( \mu \) using the sample mean \( \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \)
- For different observed values, we may have different estimates for \( \mu \).
- Which estimate is closer to the true value? No idea

Instead, we may provide an interval of values of \( \mu \):

- Make this interval include the true value of \( \mu \) with a certain “level of confidence” (say 0.95)
- Narrow interval \( \rightarrow \) precise estimate
- Point estimator and error of estimator combined.

Start With An Example

Example 7.1.1 Want to estimate the mean \( \mu \) of a normal population. Know \( \sigma = 2.0 \).
For a random sample of size \( n \): \( X_1, X_2, \ldots, X_n \). Let us use \( \mu = \bar{X} \).

1. What is the distribution of \( \bar{X} \)? What is the distribution of \( \frac{\bar{X} - \mu}{\sqrt{\sigma}} \)?

2. Find \( c \) such that \( P \left( -c < \frac{\bar{X} - \mu}{\sqrt{\sigma}} < c \right) = 0.95 \)? Find the interval of \( \mu \).

3. Given the observed sample \( n = 10 \): \{2,3,1,6,5,7,10,4,9,8\}. The estimate of \( \mu = \bar{x} = 5.5 \). Redo part 2.
Definition of C.I. of Normal Mean $\mu$

**Definition 1.** After observing $X_1 = x_1$, $X_2 = x_2$, \ldots, $X_n = x_n$. We compute the the observed sample mean $\bar{x}$ and the 95% C.I. for $\mu$ is:

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

or with 95% confidence:

\[
\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}
\]

**Example 7.1.2** A normal population has unknown $\mu$ and $\sigma = 2.0$, if $n = 31$ and $\bar{x} = 80.0$, what is the 95% C.I.?

\[
\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} =
\]

Defining Confidence Intervals

---

Defining Confidence Intervals

---

$\mu$
Interpreting a C.I.

- Given a 95% C.I., it is not entirely correct to say \( \mu \) falls in the C.I. with probability 0.95.
- Look at the probability \( P \left( -1.96 < \frac{X - \mu}{\sigma} < 1.96 \right) = 0.95 \), when substitute \( X \) with observed value \( \bar{x} \), no randomness left.
- A precise way to interpret C.I. is: with 95% confidence, \( \mu \) falls in the interval calculated.

Choosing a Different Confidence Level

**Example 7.1.1** We found the C.I. using \( P \left( -1.96 < \frac{X - \mu}{\sigma} < 1.96 \right) = 0.95 \), i.e., \( P \left( -z_{0.05/2} < \frac{X - \mu}{\sigma} < z_{0.05/2} \right) = 0.95 \), if we change 0.95 to 0.90, we say the confidence level is changed to 90%.

\[
P \left( -z_{0.10/2} < \frac{X - \mu}{\sigma} < z_{0.10/2} \right) = 0.90
\]

The 90% C.I. of the mean \( \mu \) is then:

\[
\bar{x} \pm z_{0.10/2} \frac{\sigma}{\sqrt{n}}
\]

**Definition 2.** A 100(1 - \( \alpha \))% C.I. for the mean \( \mu \) of a normal population when the value of \( \sigma \) is known is given by:

\[
\left( \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)
\]

**Example 7.1.3** Normal standard error \( \sigma \) = 0.100. Sample gives \( \bar{x} = 5.426 \), sample size 40. Choose \( \alpha = 0.1 \), find the (1 - \( \alpha \))100% C.I. for mean \( \mu \).

Confidence Level, Sample Size and Precision

- **Review** A (1 - \( \alpha \))100% C.I. for the mean \( \mu \) of a normal population when the value of \( \sigma \) is known is given by:

\[
\left( \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)
\]

- Want narrow C.I. with high confidence level.

**Example 7.2.1** Normal population, with unknown mean \( \mu \), and standard deviation \( \sigma = 2.0 \). Sample of size 25 yields \( \bar{x} = 1.0 \). What’s the 100%, 95% and 90% C.I. for \( \mu \)? Suppose a sample of size 100 also yields \( \bar{x} = 1.0 \), what’s the 90% C.I.? Which of the above is narrower?

**Example 7.2.2** Normal population, with unknown mean \( \mu \), 90% and 95% C.I.’s give: \((-0.30, 6.30), (-0.82, 6.82)\). Which one is the 95% C.I.?

Confidence Level, Sample Size and Precision

- Larger sample size \( n \) result in narrower C.I., lower confidence level result in narrower C.I.
• Confidence level cannot be too high, say, 100% will result in \((−\infty, \infty)\).

• One strategy: specify desired confidence level and interval width then determine sample size

Finding the Sample Size

Example 7.2.3 Want to estimate a normal population mean \(\mu\). Standard deviation is \(\sigma = 25\). What sample size is necessary to ensure that the 95% C.I. has a width of (at most) 10? Width of C.I. = \(\bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} - \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 2z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\) So, \(10 = 2z_{\alpha/2} \frac{25}{\sqrt{n}}\), and \(\alpha = 1 - 0.95 = 0.05\), so \(z_{0.025} = 1.96\), thus:

We got: \(n = 96.04\), so sample size is at least 97. General formula: sample size \(n\) necessary to ensure an interval width \(w\) for confidence level \((1 - \alpha)100\%\), we get \(w = 2z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\) and so:

\[
n = \left\lceil \frac{(2z_{\alpha/2} \cdot \frac{\sigma}{w})^2}{2} \right\rceil
\]

7.2 Large-Sample C.I. for a Population Mean and Proportion

Large-sample C.I. for Population Mean

Let \(X_1, X_2, \cdots, X_n\) be a random sample from an unknown population having unknown mean \(\mu\) and unknown standard deviation \(\sigma\). How to find \((1 - \alpha)100\%\) C.I.?

By CLT, \(\bar{X}\) approximately normal: mean \(\mu\), std dev \(\frac{\sigma}{\sqrt{n}}\). So,

\[
Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)
\]

Thus, we find C.I. by:

\[
P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right) \approx 1 - \alpha
\]

We get,\((1 - \alpha)100\%\) C.I. for \(\mu\) is:

\[
\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)
\]

In the expression \(\sigma\) is still unknown. For large sample sizes, we may replace it with sample standard deviation \(S\):

\[
\left(\bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}\right)
\]

For a given sample \(x_1, x_2, \cdots, x_n\), plug in \(\bar{X} = \bar{x}\) and \(S = s\) in the above expression.

Examples

Example 7.2.4 Unknown distribution, with mean \(\mu\) and standard deviation \(\sigma\). Want to find the 95% C.I. of the mean. A sample of size 196 yields \(\bar{x} = 2.0\) and \(s = 3.0\). Find the C.I.

\[
\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}
\]
Large-sample C.I. for Population Proportion \( p \)

Let \( p \) denote the proportion of "success" in a population. Let \( X \) be the number of successes in a sample of size \( n \): 
\[ X \sim \text{Binomial}(p) \]
\( n \) large, \( np > 10, n(1-p) > 10 \):
\[ X \sim \text{Normal}, \mu = np, \sigma = \sqrt{np(1-p)} \]
Estimate \( p \) using \( \hat{p} = \frac{X}{n} \), \( \hat{p} \sim \text{Normal} \) with 
\[ \mu = p, \sigma = \frac{\sqrt{p(1-p)}}{\sqrt{n}} \]

\[ \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1) \]

To find \((1-\alpha)100\%\) C.I., look at:
\[ P \left( -z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2} \right) \approx 1 - \alpha \]

Finding the C.I.

Solving the inequality, we get the \((1-\alpha)100\%\) C.I. of \( p \):

\[ \hat{p} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{z_{\alpha/2}^2}{4n^2}}{1 + \frac{z_{\alpha/2}^2}{n}}} \]

lower confidence limit=

\[ \hat{p} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p}) + \frac{z_{\alpha/2}^2}{4n^2}}{1 + \frac{z_{\alpha/2}^2}{n}}} \]

upper confidence limit=

When \( n \) is large, we get:

lower confidence limit=\( \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

upper confidence limit=\( \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

Examples

Example 7.3.1 Among 10000 cats in IN, 20% are found to be long-hairs. What is the 99% C.I. for the proportion \( p \) of long-hairs in Indiana?

\[ \left( \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right) \]

C.I. For Normal Mean \( \mu \) with Unknown \( \mu \) and \( \sigma \)

Let \( X_1, X_2, \cdots, X_n \) be a random sample from Normal with unknown \( \mu \) and \( \sigma \). Want to find a C.I. for \( \mu \). Need to introduce a new distribution:

**Theorem 3.** When \( \bar{X} \) is the mean of a random sample of size \( n \), and \( S \) is the sample standard deviation from a normal distribution with mean \( \mu \), then the rv:

\[ T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \]

has a \( t \) distribution with degree of freedom(df) \( n-1 \), denoted by \( t_{n-1} \).
Properties of t Distribution

The density curve of a t with df $\nu$:

- Bell-shaped
- More spread out than the standard normal (z) curve - heavier tails
- t curve becomes less spread out when $\nu$ increases
- t becomes standard normal when $\nu \rightarrow \infty$

$\textit{t}_\alpha,\nu$ Notation and t Table

- The notation $t_{\alpha,\nu}$ is the value on the measurement axis for which the area under the t curve with df=$\nu$ to the right of $t_{\alpha,\nu}$ is $\alpha$; $t_{\alpha,\nu}$ is called t critical value.
- The $t_{\alpha,\nu}$ values are tabulated.
- Example 7.4.1 Find $t_{0.1,10}, t_{0.001,30}, t_{0.05,120}$.
- Example 7.4.2 Determine the t critical value that will capture the desired t curve when:
  - Central area= 0.95, df=16
  - Lower-tail area= 0.1, df=16
- Example 7.4.3 For an rv T which follows a t dist with df=$n-1$, what is $P\left(-t_{\alpha/2,n-1} < T < -t_{\alpha/2,n-1}\right)$?
The One-sample \( t \) Confidence Interval

For a random sample \( X_1, X_2, \ldots, X_n \) from \( N(\mu, \sigma^2) \). We have:

\[
T = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t_{n-1}
\]

for a given \( \alpha \):

\[
P \left( -t_{\alpha/2, n-1} < \frac{\bar{X} - \mu}{s / \sqrt{n}} < -t_{\alpha/2, n-1} \right) = 1 - \alpha
\]

Solve the inequality for \( \mu \), we get the \((1 - \alpha)100\%\) C.I. for \( \mu \).

**Proposition.** Let \( \bar{x} \) and \( s \) be the sample mean and sample standard deviation computed from a random sample from a normal population with unknown \( \mu \) and \( \sigma \). Then the \((1 - \alpha)100\%\) C.I. for \( \mu \) is:

\[
\left( \bar{x} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \right)
\]

Compare to the \((1 - \alpha)100\%\) C.I. for normal mean \( \mu \) when \( \sigma \) is known:

\[
\left( \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)
\]

**Example 7.4.4** A random sample of size 10 from a normal population yields:

2,3,1,6,5,7,10,4,9,8. Find the 95\% C.I. for the normal mean \( \mu \).

\[
\left( \bar{x} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \right)
\]

so the 95\% C.I. is:

**Meaning?**

**More Intervals on Normal Populations**

- Prediction interval (P.I.) for a single observation
- Examples
- C.I. for variance \( \sigma^2 \) and standard deviation \( \sigma \)
- Examples
Prediction Interval for a Single Obs

In many applications, we wish to predict a single value of a variable using a given sample. Suppose we have a random sample from a normal population $X_1, \cdots, X_n$. Now we wish to predict the value of $X_{n+1}$, a single future observation. We need a new rv:

$$T = \frac{\bar{X} - X_{n+1}}{S\sqrt{\frac{1}{n}}}$$

It can be shown that: $T \sim t_{n-1}$. By solving the inequality for $X_{n+1}$ below:

$$P \left( -t_{\alpha/2, n-1} < \frac{\bar{X} - X_{n+1}}{S\sqrt{\frac{1}{n}}} < t_{\alpha/2, n-1} \right) = 1 - \alpha,$$

we can find the $(1 - \alpha)100\%$ P.I.

**Prediction Interval for a Single Obs**

**Proposition.** A P.I. for a single observation to be selected from a normal population distribution is

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

The prediction level is $(1 - \alpha)100\%$.

**Interpretation** Similar to the C.I.: we are $(1 - \alpha)100\%$ sure that a future obs will fall in the P.I. Or, when the P.I. is calculated sample after sample, in the long run, 95% will include the true value of the future obs.

**Example 7.5.1**

**Example 7.5.1** A random sample of size 10 from a normal population yields: $\bar{x} = 21.90$, $s = 4.134$, given $t_{0.025,9} = 2.262$, find:

- 95% C.I. for the normal mean $\mu$, how do you interpret the C.I.?

- 95% P.I. for a single future value, how do you interpret the P.I.?
7.3 C.I.s for Variance and Standard Deviation

C.I.s for Variance and Standard Deviation of a Normal

Given a sample from a normal population, want to estimate the C.I. of the variance $\sigma^2$ or std dev $\sigma$. Need a new distribution: $\chi^2$.

**Theorem 4.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with parameters $\mu$ and $\sigma^2$. Then the rv:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2}$$

has a $\chi^2$ distribution with $(n-1)$ df. Denoted by $\chi^2_{n-1}$.

$\chi^2$ Distribution

$\chi^2$ Distributions and $\chi^2_{\alpha,\nu}$ Notation

$\chi^2$ curves

- A group of positive distributions: $x \geq 0$. Not symmetric.
- Each has a positive skew with long upper tail.
- Different df $\nu$ has different density shape, and density curve becomes more symmetric when $\nu$ increases.

$\chi^2_{\alpha,\nu}$ Notation

- $\chi^2_{\alpha,\nu}$ is called a $\chi^2$ critical value, denote the number on the measurement axis such that $\alpha$ of the area under the $\chi^2$ curve with $\nu$ degrees of freedom lies to the right of $\chi^2_{\alpha,\nu}$.
- $\chi^2_{\alpha,\nu}$ values are tabulated.

Examples

- **Example 7.5.1** Find $\chi^2_{0.1,12}$, $\chi^2_{0.05,15}$, $\chi^2_{0.95,15}$.
- **Example 7.5.2** Determine the following
  - The 95% percentile of the $\chi^2$ dist with df $\nu = 15$. 

Purdue University Chapter7\print.tex; Last Modified: March 23, 2014 (W. Sharabati)
• Example 7.5.3 For a random sample of size $n$ from a normal population with variance $\sigma^2$. What is:

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \frac{\sigma^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

C.I. for $\sigma^2$ and $\sigma$

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \frac{\sigma^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1 - \alpha$$

Solve the inequality for $\sigma^2$, we get:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

So:

**Proposition.** A $(1 - \alpha)100\%$ C.I. for variance $\sigma^2$ of a normal population is:

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

A $(1 - \alpha)100\%$ C.I. for std dev $\sigma$ is then:

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}\right)$$

**Example 7.5.4** A random sample from a normal population of size 10 yields: 2,3,1,6,5,7,10,4,9,8. Find the 95% C.I. for the normal variance $\sigma^2$ and std dev $\sigma$.

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

95% C.I. $\Rightarrow \alpha = 0.05$, sample size $n = 10$ $\Rightarrow$ df = 9, $\chi^2_{0.025,9} = 19.022$, $\chi^2_{0.975,9} = 2.700$.

$$s^2 = \frac{2^2 + 3^2 + \cdots + 8^2 - (2+3+\cdots+8)^2}{10-1} = 9.167$$

95% C.I. for $\sigma^2$

95% C.I. for $\sigma$ is then: