Stats 11 (Fall 2004) Lecture Note — Instructor: Hongquan Xu
Introduction to Statistical Methods for Business and Economics

Chapter 8 Confidence Intervals

In Chapter 7, we used a two-standard-error interval to provide a range of likely values for an unknown parameter. Now we take a close look at this idea and consider intervals of varying widths and confidence.

Example:

Suppose that the average time to relief for a standard drug for treating migraines is 2.8 hours. A new drug is proposed to help treat migraines. In a controlled study, 100 patients who suffer from migraines are given the new drug upon the start of their next migraine attack. Would you conclude that the new drug is successful if the mean time to relief for the 100 patients is:

3 hours?

2.8 hours?

2.5 hours?

1.5 hours?

Suppose the *sample mean were 1.5 hours* – would this imply that the population mean for the new drug is 1.5? Could the true population mean for the new drug still be 2.8? How confident can we be in the value of 1.5 as our estimate for the population mean?

Some Review

- Suppose a **random sample** of size n is taken from some large population with mean ____ and standard deviation ____.
- The observations (i.e., the n data points), before you look, can be denoted ____________
- The observations are ____.
- As we have seen before, the sample mean can be written as
- Because the sample mean is computed from a random sample, then it is a random variable, with a probability distribution.

Sampling Distribution of $\overline{X}$

If $\overline{X}$ is the sample mean from a random sample of size n from a normal distribution, ___________, then the distribution of the sample mean is _______________.

Central Limit Theorem

If $\overline{X}$ is the sample mean from a random sample of size n from _________________ with mean ____ and standard deviation ____, then when n is ________, the sampling distribution of the sample mean is **approximately** _________________.

Section 8.1 Introduction

Example: Crop researchers plant 64 plots for a new variety of corn. Yields are measured in bushels of corn per acre. Suppose that the "true" distribution of the yield is normal and has $\sigma = 10$.

Draw picture:

Goal*:* Learn about the population mean ____

Point estimator: _____

But we know $\overline{X}$ may not equal μ –
in fact, the possible $\overline{X}$ values vary around μ with a standard deviation of $\frac{\sigma}{\sqrt{n}}$.

Idea*:* Find some lower and upper bounds such that:

$$P(L \leq \mu \leq U) = \text{some high number, say, } \_\_\_\_\_\_\_\_.$$

This high level is called the '______________________ ' and denoted by _________.

Common levels are 0.90, 0.95, 0.99.

Question: How to find the bounds L and U?

Can you find the values of z^* which satisfy the following probability statement?

$$P\left(-z^* \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z^*\right) = 0.95$$

Sure you can – draw a picture of the $N(0, 1)$ distribution...

So $z^* =$

Next, do the algebra to solve for μ in the middle...

$$P\left(-1.96 \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) = 0.95$$

We have just derived the formula for a 95% confidence interval for the population mean.

Example: The (r.s. of) 64 plots gave a mean yield of 125 bushels per acre. Give a 95% confidence interval for the population mean yield for this new variety of corn.

$\overline{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \Rightarrow$

We would estimate the mean yield to be between ______________ and ______________ bushels per acre with 95% confidence.

Computation is not hard ... interpreting the interval (like above) is not hard ... understanding and interpreting what that 95% confidence level means is not so easy!!!

Questions:

1. Can we say the probability that the above interval (122.55, 127.45) contains the population mean μ is 0.95? that is, $P\,(122.55 \leq \mu \leq 127.45) = 0.95$?

2. Can we say that 95% of the time the population mean μ will be in **the interval**?

The population mean is not a random quantity, it does not vary – once we have "looked" (computed) the actual interval, we cannot talk about probability for this interval anymore. The 95% confidence level applies to the procedure "before you look".

Text says (page 331): Our confidence in the interval comes from the fact that it was produced by a method that works 95% of the time.

A level C confidence interval for a parameter is an interval computed from sample data by a **method that has probability** C of producing an interval containing the true value of the parameter.

Figure 8.1.2 (page 332) is a good picture to remember the correct interpretation of the confidence LEVEL.

Interpretation for mean corn yield:

The interval (122.55, 127.45) was computed with a method which if repeated over and over .....

Notes:

- What about a **90% confidence interval** or a **99% confidence interval**?
  The value of z^* will change accordingly.

$$\overline{x} \pm z^*\left(\frac{\sigma}{\sqrt{n}}\right)$$

Example: Give the 90% confidence interval for the population mean corn yield.

Find $z^* =$

The **lower confidence level** led to a ______________ interval.

99% CI?

General form for a confidence interval:

point estimate ± margin of error

where the **margin of error** is

It depends on the level of confidence, the standard deviation in the population and the sample size. You can **solve this expression for the sample size** and have a method for determining the sample size needed to produce a CI for the mean with a certain level of confidence and a desired error margin.

Example: The corn researchers wish to estimate the mean yield within ±1 bushel with 95% confidence. How many fields must they use in the experiment?

Example: To study career paths of major-chain hotel managers. A survey was sent to 160 hotels and 114 responded. The mean time of the 114 managers who had spent with current company was 11.78 years.

a. Give a 99% confidence interval for the mean number of years. (Assume population standard deviation is 3.2 years.)

b. How large of a sample would be needed to estimate the population mean time within ±0.5 years with 99% confidence?

Example: A 99% CI for the mean time for non-major-chain hotel managers, based on a r.s. of 100 hotels: (8.0, 12.0) years.

- Is the population mean μ in this confidence interval? **Yes, No, Can't Tell**
- Is the sample mean $\overline{x}$ in this confidence interval? **Yes, No, Can't Tell; If yes, find** $\overline{x}$.
- Would a 90% CI based on the same data be wider or narrower?

Section 8.2 Means

In practice, we do NOT known σ. So we can estimate σ by ______________

Recall that, in Section 7.6, if we have a random sample of size n from a normal distribution with mean μ, then the statistic

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a ____________________

Confidence Interval for the true population mean μ

$$\overline{x} \pm t^* \text{se}\left(\overline{x}\right) = \overline{x} \pm t^* \left(\frac{s}{\sqrt{n}}\right)$$

where t^* is the value for a $t(n-1)$ distribution with area C between $-t^*$ and t^*.

This interval requires we have a **random sample from a normal population**. If the sample size is large, the assumption of normality is not so crucial and the result is approximate. (Why?)

Example: Study of DDT poisoning on rats to learn about how DDT causes tremors. Response = Refractory Period (time for nerve to recover after stimulus). Response varies normally.

Data for 4 rats: 1.6 1.7 1.8 1.9

(a) Find the mean and the standard error of the mean.

(b) Give a 90% CI for the mean refractory period for all such rats.

Section 8.3 Proportions

Recall, in section 7.3, we studied the sampling distribution of

Sample Proportion: $\widehat{P} = \frac{X}{n}$ = the proportion of *successes* in a sample of size n

Normal Approximation for a Proportion

Let $\widehat{P}$ be the sample proportion of successes in a random sample of size n

from a population with proportion of successes p.

If n is large, then $\widehat{P}$ is approximately

(rule: $np \geq 10$ and $n\left(1-p\right) \geq 10$)

So, if the sample size is sufficiently large, we can standardize the sample proportion $\widehat{P}$ (or equivalently the count X) and use Z procedures for producing a CI for the population proportion p.

Large-Sample Confidence Interval for a Population Proportion p

$$\widehat{p} \pm z^* \text{se } (\widehat{p}) = \widehat{p} \pm z^* \left(\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \right)$$

where z^* is the value for the $N(0,1)$ distribution with area C between $-z^*$ and z^*.

Example: One question on a survey of households: "Did you have a Christmas Tree last year?" Of the 500 households responding, 421 said "Yes".

(a) Find the sample proportion who said "Yes".

(b) Find the standard error for the sample proportion given in part a.

(c) Give a 95% CI for the proportion of households who had a tree this past year?

Margin of Error and Desired Sample Size

Error Margin $= m =$

Solving for the **sample size** n we have:

But we will not know the value of $\widehat{p}$ before the data is gathered.

So we will be conservative and use $p = 0.5$.

If it is not a whole number, we would round up to the next integer.

Example:

(d) How many households would need to be surveyed to estimate the proportion who say "YES" with 95% confidence and an error margin of ± 0.02 (i.e. within 2%)?