Chapter 14. Confidence Intervals: The Basics

Note. We are now transitioning into the heart of statistics. The main idea is to estimate properties of a population based on properties of a sample. As the book says: "The usual reason for taking a sample is not to learn about the individuals in the sample but to *infer* from the sample data some conclusion about the wider population that the sample represents. ... Statistical inference uses the language of probability to say how trustworthy our conclusions are." (page 343)

Definition. Statistical inference provides methods for drawing conclusions about a population from sample data.

Note. Our first population parameter to be estimated (or "inferred") is the mean μ . We do so by taking a simple random sample (SRS). Initially, we require the following assumptions about the sample and the population:

- 1. We have a SRS from the population of interest. There is no nonresponse or other practical difficulty.
- 2. The variable we measure has a perfectly normal distribution $N(\mu, \sigma)$ in the population.
- 3. We don't know the population mean μ . But we do know

the population standard deviation σ .

The books states "The conditions that we have a perfect SRS, that the population is perfectly Normal and that we know the population σ are all unrealistic." (page 344)

Estimating with Confidence

Note. "The big idea" in this chapter is that \overline{x} as calculated from a sample should be close to the population mean μ . "Close" will be quantified and related to the size of the sample. Based on the sample, we will have an interval of the form: estimate \pm margin of error. We will have a level of confidence that reflects the probability that the population mean μ lies in this interval.

Definition. A level *C* confidence interval for a parameter has two parts:

- An interval from the data, usually of the form: estimate \pm margin of error.
- A confidence level C, which gives the probability that the interval will capture the true parameter value in repeated samples. That is, the confidence level is the success rate for the method.

Note. Figure 14.3 illustrates these ideas graphically. It has the graph of the sampling distribution of \overline{x} at the top, and the graphs of confidence intervals based on 25 different samples at the bottom. The 25 confidence intervals are 95% confidence intervals and one of the intervals does not include the actual population mean μ .



Figure 14.3 page 347.

Example. Exercise 14.1 page 348.

Confidence Intervals for the Mean μ

Note. Suppose that we want to find an interval which we are 95% confident will contain a population mean μ . If we take a sample of size n, then the sampling distribution is (from Chapter 11) $N(\mu, \sigma/\sqrt{n})$. Therefore we can use the normal distribution N(0, 1) to calculate the level of confidence C = 95%. This requires us to go the Table A, find the entry 0.975 in the table, and read off the corresponding z-score. This gives a z-score of 1.96, which we denote as z^* and call a **critical value** (notice that the 68-95-99.7 Rule would imply a z-score of 2, so this rule is actually an approximation). We use the sample mean \overline{x} as an estimate for μ and then we are C = 95% that the population mean μ lies in the interval

from
$$\overline{x} - z^* \frac{\sigma}{\sqrt{n}}$$
 to $\overline{x} + z^* \frac{\sigma}{\sqrt{n}}$

or $\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}}$. In general, we want a level of confidence C, then we can use Table A to find a corresponding critical value z^* . In fact, Table C can be used to directly find z^* for certain common values of C. The relationship between C and z^* are as given in Figure 14.4.



Figure 14.4 page 349.

Definition. If a simple random sample of size n is drawn from a normal population having unknown mean μ and known standard deviation σ , the a **level** C **confidence interval** for μ is

$$\overline{x} \pm z^* \frac{\partial}{\sqrt{n}}.$$

Here, \overline{x} is the sample mean and z^* is the **critical value** and is found in Table C (third line from the bottom) and is based on the level of confidence C.

Example S.14.1. Stooge Confidence Intervals.

Use a random number generator to select 10 Stooge films from the population of 190. Use the number of slaps data from *The Three Stooges: An Illustrated History* to find 95% and 99% confidence intervals for the mean number of slaps per film. Assume a population standard deviation of $\sigma = 7.00$.

Solution. Using the random number generator at

http://www.random.org/integers/

we generate 10 integers between 1 and 190 to get (2/22/2009):

21 22 46 57 67 82 117 146 152 159 Using these numbers as the number of Stooge films, we get the following data:

#	Title	# Slaps
21.	Dizzy Doctors	7
22.	Three Dumb Clucks	8
46.	A Plumbing We Will Go	4
57.	An Ache in Every Stake	11
67.	They Stooge to Conga	9
82.	No Dough, Boys	18
117.	Malice in the Palace	12
146.	Loose Loot	26
152.	Goof on the Roof	7
159.	Fling in the Ring	43

We find that for this sample $\overline{x} = 14.5$. The sampling distribution standard deviation is $\sigma/\sqrt{n} = 7.00/\sqrt{10} = 2.21$. For a C = 95% confidence interval, the critical value is $z^* = 1.96$ and therefore the 95% confidence interval is

$$\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 14.5 \pm 1.96 \times 2.21 = 14.5 \pm 4.33,$$

of "from 10.17 to 18.83." For a C = 99% confidence interval, the critical value from Table C is $z^* = 2.576$ and therefore the 99% confidence interval is

$$\overline{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 14.5 \pm 2.576 \times 2.21 = 14.5 \pm 5.69,$$

of "from 8.81 to 20.19."

Note. The book is big on setting out steps to solve and discuss problems. It lays out the following four-step process when dealing with confidence intervals:

- **State:** What is the practical question that requires estimating a parameter?
- **Formulate:** Identify the parameter and choose a level of confidence.

Solve: Carry out the work in two phases:

- (a) **Check the conditions** for the interval you plan to use.
- (b) Calculate the **confidence interval**.
- **Conclude:** Return to the practical question to describe your results in this setting.

How Confidence Intervals Behave

Note. The text mentions the following as behavior of confidence intervals: The margin of error gets smaller when

- z^* gets smaller. Smaller z^* is the same as lower confidence level C. This is illustrated in Example S.14.1.
- σ is smaller.

• n gets larger.

Choosing the Sample Size

Note. If we desire a margin of error of size m in a confidence interval for μ , then we need to have a SRS of size n where $n = \left(\frac{z^*\sigma}{m}\right)^2$.

Example S.14.2. Stooge Sample Sizes.

Suppose you want to estimate the mean number of slaps per film in the Three Stooges work to within m = 3 slaps per films with 90% confidence. What sample size would you use? What 99.9% confidence? Assume a population standard deviation of $\sigma = 7.00$.

Partial Solution. For C = 90%, $z^* = 1.645$. Therefore Sample size must be at least

$$n = \left(\frac{z^*\sigma}{m}\right)^2 = \left(\frac{1.645 \times 7.00}{3}\right)^2 = 14.73.$$

Therefore we would need a sample of size 15.

Example. Exercise 14.33 page 360.

rbg-2-22-2009