Chapter 4
Confidence Interval of the Mean

Section 4.1
The Confidence of an Estimate

David W. Gerbing

The School of Business
Portland State University

- The Confidence of an Estimate

Why Do We Care about the Confidence Interval?

For an obtained *m*, always construct the confidence interval

- How much information does an analysis convey regarding μ, the population mean desired as a basis for decision making?
- **Key Concept**: The estimate *m* provides useful information for decision making only if there is confidence that *m* is close to μ
- For example, evaluate the cost of a specific procedure to accomplish some task of interest over multiple occurrences
- Obtain a sample mean of $m = \$894$, which, without assessing confidence, yields a data summary of unknown quality
  - If the value of μ is likely within the interval $\$894 \pm \$539$, the confidence interval, then there is *no* actionable information
  - In contrast, if the value of μ is likely within $\$894 \pm \$5.39$, then this same sample mean provides *much* information
- Always include a confidence interval with a sample estimate, such as *m*, to indicate how close *m* is likely to μ

4.1
The Confidence of an Estimate
How Close is a Sample Value to the True Value?

How Close is One Estimate to the True Value?

Use multiple values of a statistic to assess confidence

- Usually observe a single sample
- **Sampling variation**: The value of a statistic varies from one sample to another (usually hypothetical) sample
- **Key Concept**: A statistic calculated from the one obtained sample, such as m, is useful for decision making only if it is close to its true corresponding population value, such as μ
  - Suppose a sample from a population of interest of size $n = 10$ reveals $m = 21.4$
  - We know that μ is probably not 150 or -450.3
- Although μ could be practically anything, it is likely close to m and values of 150 or -450.3 are not close to $m = 21.4$
- The question of interest: How close?

The Confidence Interval

Estimate the unknown value μ

- Just knowing the value of a single m does not provide enough information to know how close that m is to μ
- What we need instead is a *range* of values from which to better evaluate the true value of μ
- This range of values that likely contains μ is a primary concept to be applied in the analysis of *any* sample mean, m
- **Confidence Interval**: A *range of values* that likely contains the population value at a specified level of confidence
- How do we obtain this confidence interval to estimate μ?
- **Key Concept**: Obtain the confidence interval of μ from the range of variation of the usually hypothetical values of many m's, each m calculated from a different sample

Simulate Drawing Multiple Samples

The computer provides the data *as if* we have multiple samples

- **Key Question**: How does the knowledge of *m* over multiple samples provide useful information to estimate the unknown μ?
- Computer simulation allows us to directly explore what happens if we would actually have multiple random samples

- `lessR` function `simMeans()`: Simulate multiple samples from the same normal population and calculate *m* for each sample
  - *Required*: `ns`, number of samples
    `n`, size of each sample
  - *Default*: `mu=0`, population mean
    `sigma=1`, population standard deviation
  - *Optional*: see `?simMeans`, e.g., `pause=TRUE`

```
> simMeans(ns=8, n=10, mu=21, sigma=9)
```


Example 1: Multiple Values of *m* to Estimate μ

m's from 8 Different Samples, n=10 per Sample

- Take 8 samples, each of size $n = 10$, from the *same* normal population $\mu = 21$ and $\sigma = 9$, and calculate *m* for *each* sample

	1	2	3	4	5	6	7	8
#1:	15.8	17.4	20.3	20.5	21.5	22.1	23.2	25.0

- The range of variation of the obtained *m*'s extends from 15.8 all the way to 25.0, so what is μ?
- To start, μ is probably *not* around 150, or -10, but rather . . .

  μ is likely within the range of variation of the *m*'s, somewhere between around 17 to 23
- **Key Concept**: As later shown in more detail, the range of variation of *m* across repeated samples provides the information needed to construct the confidence interval to estimate μ

Example 2: Multiple Values of *m* to Estimate μ

m's from Eight Different Samples, n=1000 per Sample

- Suppose another 8 samples were taken from the *same* normal population $\mu = 21$ and $\sigma = 9$, but now each of size $n = 1000$

```
> simMeans(ns=8, n=1000, mu=21, sigma=9)
```

	1	2	3	4	5	6	7	8
#2:	20.6	20.8	20.8	21.0	21.2	21.2	21.3	21.5

- What is our best estimate of μ from this information?

  μ is likely within the range of variation of the *m*'s, somewhere between around 20.8 only to 21.3
- A smaller interval, here at $n = 1000$ for each sample, is obtained compared to the previous example at $n = 10$
- **Key Concept**: As shown, an *m* from a sample of larger size tends to provide a more precise estimate of μ

Applying the Information from Multiple Values of m

Goal is to evaluate confidence of a single sample mean m

- After the sampling variability of m has been established, suppose measurements of some variable Y for a new, *single* sample yield $m = 21.4$
- How much confidence that $m = 21.4$ is *close* to unknown μ?
- **Key Concept**: The less variable are *all* the m's, the more confidence that *any one m* is closer to μ
  - Example #1: If m typically varies 4 units, from 15.8 to 25.0, $m = 21.4$ is likely within two or more units of μ
  - Example #2: If m typically varies 1 unit, from 20.8 to 21.8, $m = 21.4$ is likely within only a half a unit of μ
- A method is needed to assess the variability of the sample means over repeated samples, the m's, presumably without actually having to obtain these additional samples

- The End

Index

Subtract 2 from each listed value to get the Slide #