Statistical Analysis in Fin Mkts

MSF 502

Li Cai

CHAPTER

8 Estimation

Chapter 8 Learning Objectives (LOs)

LO 8.1: Discuss point estimators and their desirable properties.

LO 8.2: Explain an interval estimator.

LO 8.3: Calculate a confidence interval for the population mean when the population standard deviation is known.

LO 8.4: Describe the factors that influence the width of a confidence interval.

LO 8.5: Discuss features of the *t* distribution.

LO 8.6: Calculate a confidence interval for the population mean when the population standard deviation is not known.

LO 8.7: Calculate a confidence interval for the population proportion.

LO 8.8: Select a sample size to estimate the population mean and the population proportion.

Fuel Usage of “Ultra-Green” Cars

- A car manufacturer advertises that its new “ultra-green” car obtains an average of 100 mpg and, based on its fuel emissions, has earned an A+ rating from the Environmental Protection Agency.
- Pinnacle Research, an independent consumer advocacy firm, obtains a sample of 25 cars for testing purposes.
- Each car is driven the same distance in identical conditions in order to obtain the car’s mpg.

Fuel Usage of "Ultra-Green" Cars

- The mpg for each "Ultra-Green" car is given below:

97	117	93	79	97
87	78	83	94	96
102	98	82	96	113
113	111	90	101	99
112	89	92	96	98

- Jared would like to use the data in this sample to:
  - Estimate with 90% confidence
    - The mean mpg of all ultra-green cars.
    - The proportion of all ultra-green cars that obtain over 100 mpg.
  - Determine the sample size needed to achieve a specified level of precision in the mean and

8.1 Point Estimators and Their Properties

LO 8.1 Discuss point estimators and their desirable properties.

- **Point Estimator**
  - **A function of the random sample used to make inferences about the value of an unknown population parameter.**
  - **For example, $\bar{X}$ is a point estimator for μ and $\bar{P}$ is a point estimator for *p*.**
- **Point Estimate**
  - **The value of the point estimator derived from a given sample.**
  - **For example, $\bar{x} = 96.5$ is a point estimate of the mean mpg for all ultra-green cars.**

LO 8.1

8.1 Point Estimators and Their Properties

- **Example:**

A statistics section at a large university has 100 students. The scores of 10 randomly selected final exams are:

66	72	40	85	75	90	92	60	82	38

Calculate the point estimate for the population mean.

SOLUTION: We calculate $\bar{x} = \frac{66 + 72 + \cdots + 38}{10} = \frac{700}{10} = 70$. Therefore, a score of 70 is a point estimate of the population mean.

LO 8.1

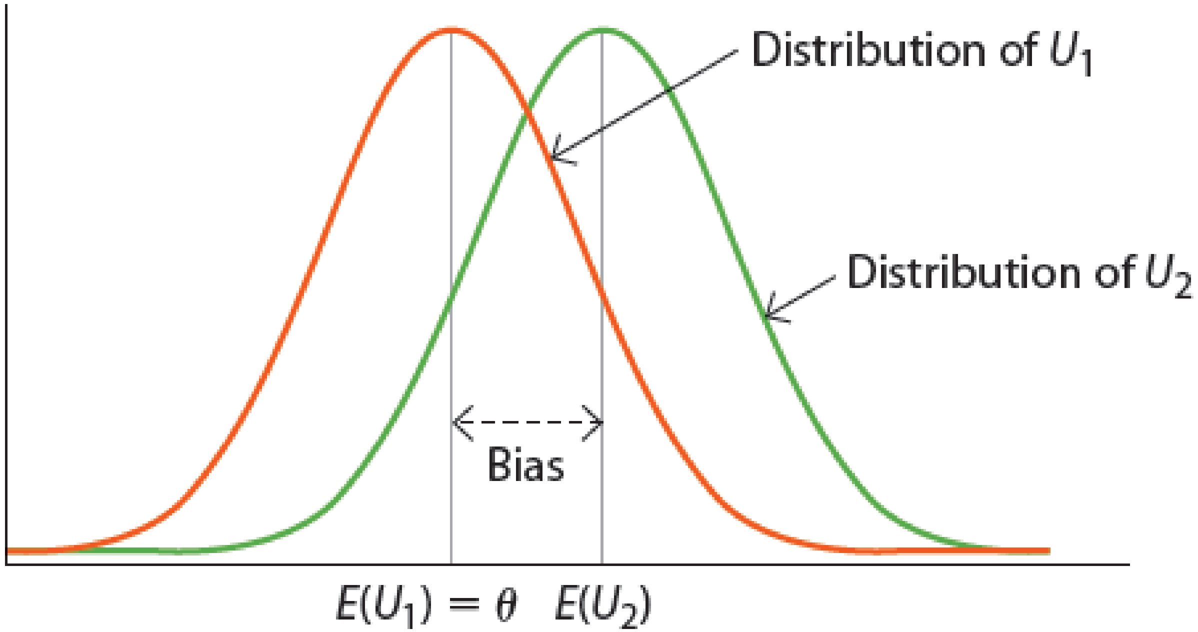
8.1 Point Estimators and Their Properties

- **Properties of Point Estimators**
  - **Unbiased**
    - **An estimator is unbiased if its expected value equals the unknown population parameter being estimated.**
  - **Efficient**
    - **An unbiased estimator is efficient if its standard error is lower than that of other unbiased estimators.**
  - **Consistent**
    - **An estimator is consistent if it approaches the unknown population parameter being estimated as the sample size grows larger.**

LO 8.1

8.1 Point Estimators and Their Properties

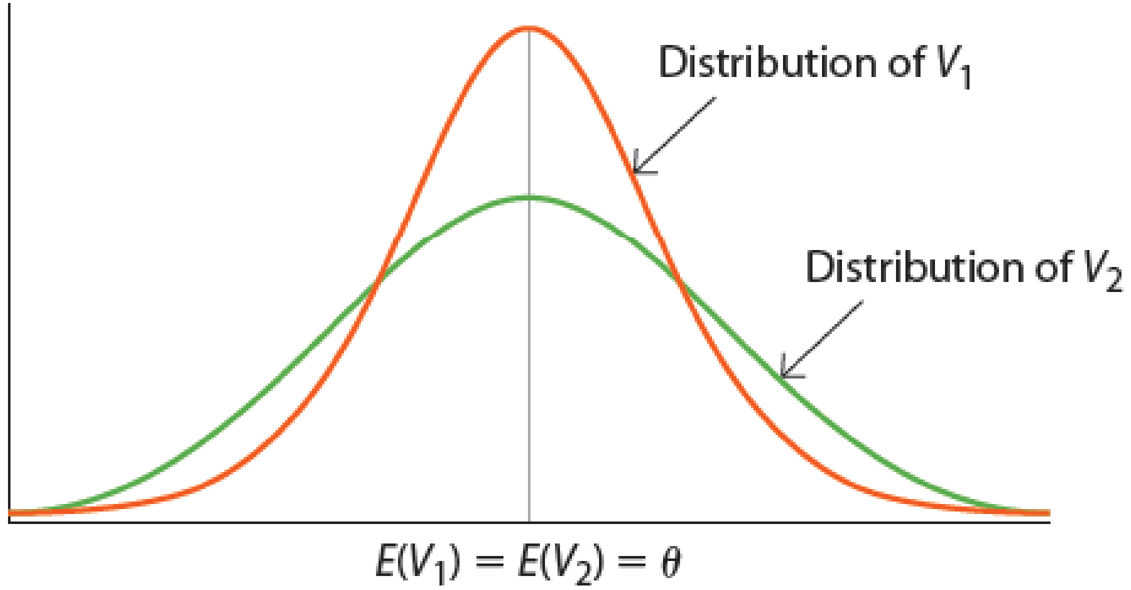
- **Properties of Point Estimators Illustrated: Unbiased Estimators**
  - **The distributions of *unbiased* (U_1) and biased (U_2) estimators.**



LO 8.1

8.1 Point Estimators and Their Properties

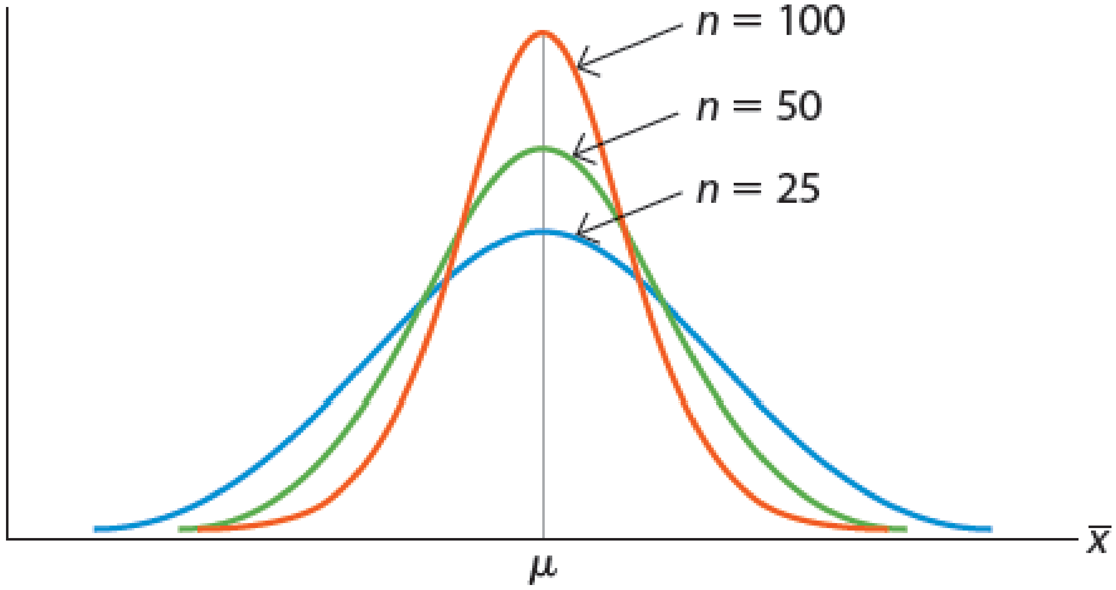
- **Properties of Point Estimators Illustrated: Efficient Estimators**
  - **The distributions of *efficient* (V_1) and less efficient (V_2) estimators.**



LO 8.1

8.1 Point Estimators and Their Properties

- **Properties of Point Estimators Illustrated: Consistent Estimator**
  - **The distribution of a *consistent* estimator $\bar{X}$ for various sample sizes.**



8.2 Confidence Interval of the Population Mean When σ Is Known

LO 8.2 Explain an interval estimator.

- **Confidence Interval**—provides a range of values that, with a certain level of confidence, contains the population parameter of interest.
  - Also referred to as an **interval estimate**.
- Construct a confidence interval as:
  Point estimate ± Margin of error.
  - **Margin of error** accounts for the variability of the estimator and the desired confidence level of the interval.

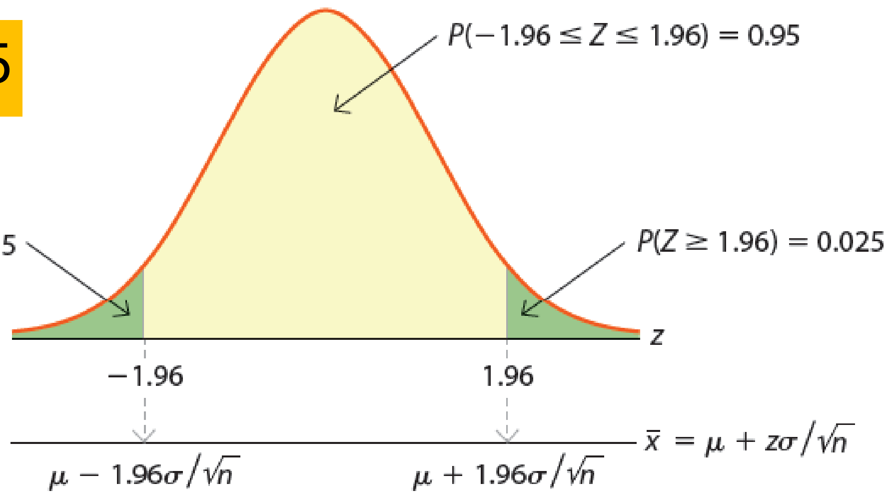
8.2 Confidence Interval of the Population Mean When σ Is Known

LO 8.3 Calculate a confidence interval for the population mean when the population standard deviation is known.

- **Constructing a Confidence Interval for μ When σ is Known**
  - **Consider a standard normal random variable**

$$P(-1.96 \le Z \le 1.96) = 0.95$$

  -
  - **as illustrated here.**



LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Constructing a Confidence Interval for μ When σ is Known**
  - **Since** $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$
  - **We get** $P\left(-1.96 \le \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \le 1.96 \right) = 0.95$
  - **Which, after algebraically manipulating, is equal to** $P\left(\mu - 1.96\, \sigma / \sqrt{n} \le \bar{X} \le \mu + 1.96\, \sigma / \sqrt{n} \right) = 0.95$

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

Constructing a Confidence Interval for μ When σ is Known

- **Note that** $P\left(\mu - 1.96\,\sigma/\sqrt{n} \le \bar{X} \le \mu + 1.96\,\sigma/\sqrt{n}\right) = 0.95$

implies there is a 95% probability that the sample mean $\bar{X}$ **will fall within the interval** $\mu \pm 1.96\,\sigma/\sqrt{n}$

- **Thus, if samples of size *n* are drawn repeatedly from a given population, 95% of the computed sample means,** $\bar{X}$**s, will fall within the interval and the remaining 5% will fall outside the interval.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Constructing a Confidence Interval for μ When σ is Known**

- **Since we do not know μ, we cannot determine if a particular $\bar{x}$ falls within the interval or not.**
  - **However, we do know that $\bar{x}$ will fall within the interval $\mu \pm 1.96\sigma/\sqrt{n}$ if and only if μ falls within the interval $\bar{x} \pm 1.96\sigma/\sqrt{n}$.**
  - **This will happen 95% of the time given the interval construction. Thus, this is a 95% confidence interval for the population mean**

LO 8.3

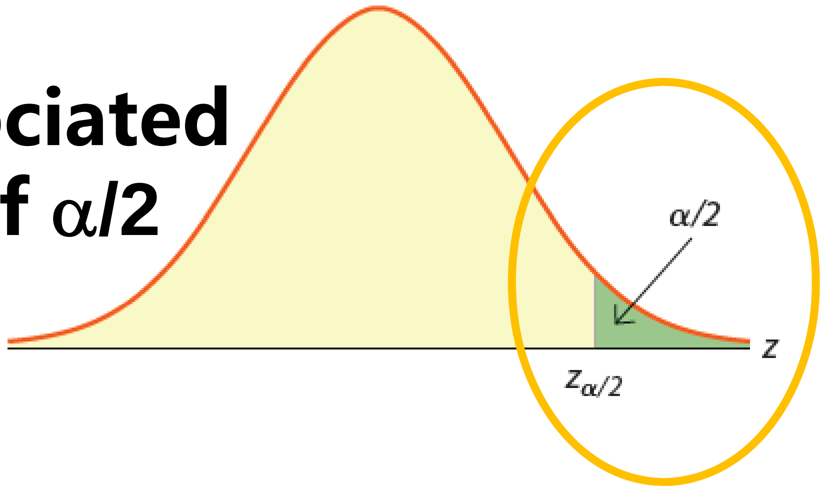
8.2 Confidence Interval of the Population Mean When σ Is Known

- **Constructing a Confidence Interval for μ When σ is Known**
  - **Level of significance (i.e., probability of error) = α.**
  - **Confidence coefficient = (1 – α)**
    α = 1 – confidence coefficient
  - **A 100(1-α)% confidence interval of the population mean μ when the standard deviation σ is known is computed as** $\bar{x} \pm z_{\alpha/2}\, \sigma/\sqrt{n}$
    or equivalently $\left[\bar{x} - z_{\alpha/2}\, \sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\, \sigma/\sqrt{n}\right]$ **.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Constructing a Confidence Interval for μ When σ is Known**
  - **$z_{\alpha/2}$ is the z value associated with the probability of $\alpha/2$ in the upper-tail.**



$$\left[\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n},\ \bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n}\right]$$

  - **Confidence Intervals:**
    - **90%, α = 0.10, $\alpha/2$ = 0.05, $z_{\alpha/2} = z_{.05}$ = 1.645.**
    - **95%, α = 0.05, $\alpha/2$ = 0.025, $z_{\alpha/2} = z_{.025}$ = 1.96.**
    - **99%, α = 0.01, $\alpha/2$ = 0.005, $z_{\alpha/2} = z_{.005}$ = 2.575.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Example: Constructing a Confidence Interval for μ When σ is Known**
  - **A sample of 25 cereal boxes of Granola Crunch, a generic brand of cereal, yields a mean weight of 1.02 pounds of cereal per box.**
  - **Construct a 95% confidence interval of the mean weight of all cereal boxes.**
  - **Assume that the weight is normally distributed with a population standard deviation of 0.03 pounds.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Constructing a Confidence Interval for μ When σ is Known**
  - **This is what we know:**

$$n = 25,\ \bar{x} = 1.02 \text{ pounds}$$

$$\alpha = (1 - .95) = .05,\ z_{\alpha/2} = 1.96$$

$$\sigma = 0.03$$

  - **Substituting these values, we get**

$$\bar{x} \pm 1.96\,\sigma/\sqrt{n} = 1.02 \pm 1.96\left(0.03/\sqrt{25}\right) = 1.02 \pm 0.012$$

or, with 95% confidence, the mean weight of all cereal boxes falls between 1.008 and 1.032 pounds.

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Interpreting a Confidence Interval**
  - **Interpreting a confidence interval requires care.**
  - **Incorrect: The probability that μ falls in the interval is 0.95.**
  - **Correct: If numerous samples of size *n* are drawn from a given population, then 95% of the intervals formed by the f $\bar{x} \pm z_{\alpha/2}\,\sigma/\sqrt{n}$ will contain μ.**
    - **Since there are many possible samples, we will be right 95% of the time, thus giving us 95% confidence.**

8.2 Confidence Interval of the Population Mean When σ Is Known

LO 8.4 Describe the factors that influence the width of a confidence interval.

- **The Width of a Confidence Interval**
  - **Margin of Error** $z_{\alpha/2}\,\sigma/\sqrt{n}$
  - **Confidence Interval Width** $2\left(z_{\alpha/2}\,\sigma/\sqrt{n}\right)$
  - **The width of the confidence interval is influenced by the:**
    - **Sample size *n*.**
    - **Standard deviation σ.**
    - **Confidence level 100(1 – α)%.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **The Width of a Confidence Interval is influenced by:**
  - I. **For a given confidence level 100(1 – α)% and sample size *n*, the width of the interval is wider, the greater the population standard deviation σ.**
  - **Example: Let the standard deviation of the population of cereal boxes of Granola Crunch be 0.05 instead of 0.03. Compute a 95% confidence interval based on the same sample information.**

$$\bar{x} \pm z_{\alpha/2}\,\sigma/\sqrt{n} = 1.02 \pm 1.96\left(0.05/\sqrt{25}\right) = 1.02 \pm 0.20$$

  - **This confidence interval width has increased from 0.024 to 2(0.020) = 0.040.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **The Width of a Confidence Interval is influenced by:**
  - **II. For a given confidence level 100(1 – α)% and population standard deviation σ, the width of the interval is wider, the smaller the sample size *n*.**
  - **Example: Instead of 25 observations, let the sample be based on 16 cereal boxes of Granola Crunch. Compute a 95% confidence interval using a sample mean of 1.02 pounds and a population standard deviation of 0.03.**

$$\bar{x} \pm z_{\alpha/2}\,\sigma/\sqrt{n} = 1.02 \pm 1.96\left(0.03/\sqrt{16}\right) = 1.02 \pm 0.015$$

  - **This confidence interval width has increased from 0.024 to 2(0.015) = 0.030.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **The Width of a Confidence Interval is influenced by:**
  - III. **For a given sample size *n* and population standard deviation σ, the width of the interval is wider, the greater the confidence level 100(1 – α)%.**
  - **Example: Instead of a 95% confidence interval, compute a 99% confidence interval based on the information from the sample of Granola Crunch cereal boxes.**

$$\bar{x} \pm z_{\alpha/2}\,\sigma/\sqrt{n} = 1.02 \pm 2.575\left(0.03/\sqrt{25}\right) = 1.02 \pm 0.015$$

  - **This confidence interval width has increased from 0.024 to 2(0.015) = 0.030.**

LO 8.3

8.2 Confidence Interval of the Population Mean When σ Is Known

- **Example:**

IQ tests are designed to yield results that are approximately normally distributed. Researchers think that the population standard deviation is 15. A reporter is interested in estimating the average IQ of employees in a large high-tech firm in California. She gathers the IQ information on 22 employees of this firm and records the sample mean IQ as 106.

a. Compute 90% and 99% confidence intervals of the average IQ in this firm.

SOLUTION:

a. For a 90% confidence interval, $z_{\alpha/2} = z_{0.05} = 1.645$. Similarly, for a 99% confidence interval, $z_{\alpha/2} = z_{0.005} = 2.575$.

The 90% confidence interval is $106 \pm 1.645 \frac{15}{\sqrt{22}} = 106 \pm 5.26$.

The 99% confidence interval is $106 \pm 2.575 \frac{15}{\sqrt{22}} = 106 \pm 8.23$.

Note that the 99% interval is wider than the 90% interval.

8.5 Confidence Interval of the Population Mean When σ Is Unknown

LO 8.5 Discuss features of the *t* distribution.

- The ***t* Distribution**
  - If repeated samples of size n are taken from a normal population with a finite variance, then the statistic T follows the t distribution with $(n - 1)$ degrees of freedom, df.

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

    - **Degrees of freedom** determine the extent of the broadness of the tails of the distribution; the fewer the degrees of freedom, the broader the tails.

LO 8.5

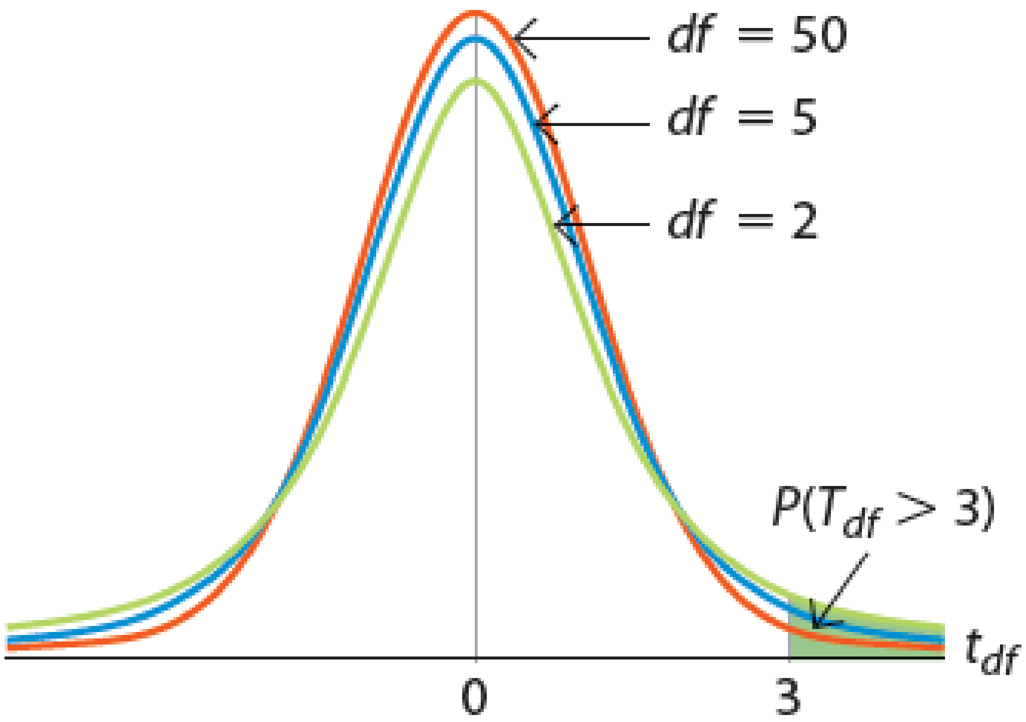
8.3 Confidence Interval of the Population Mean When σ Is Unknown

- **Summary of the t_{df} Distribution**
  - **Bell-shaped and symmetric around 0 with asymptotic tails (the tails get closer and closer to the horizontal axis, but never touch it).**
  - **Has slightly broader tails than the *z* distribution.**
  - **Consists of a family of distributions where the actual shape of each one depends on the *df*. As *df* increases, the t_{df} distribution becomes similar to the *z* distribution; it is identical to the *z* distribution when *df* approaches infinity.**

LO 8.5

8.3 Confidence Interval of the Population Mean When σ Is Unknown

- **The t_{df} Distribution with Various Degrees of Freedom**



LO 8.5

8.3 Confidence Interval of the Population Mean When σ Is Unknown

- **Example: Compute $t_{\alpha,df}$ for $\alpha = 0.025$ using 2, 5, and 50 degrees of freedom.**
  - **Solution: Turning to the Student's *t* Distribution table in Appendix A, we find that**
    - **For *df* = 2, $t_{0.025,2} = 4.303$.**
    - **For *df* = 5, $t_{0.025,5} = 2.571$.**
    - **For *df* = 50, $t_{0.025,50} = 2.009$.**
  - **Note that the t_{df} values change with the degrees of freedom. Further, as *df* increases, the t_{df} distribution begins to resemble the *z* distribution.**

8.3 Confidence Interval of the Population Mean When σ Is Unknown

LO 8.6 Calculate a confidence interval for the population mean when the population standard deviation is not known.

- **Constructing a Confidence Interval for μ When σ is Unknown**
  - A 100(1 – α)% confidence interval of the population mean μ when the population standard deviation σ is not known is computed as

$$\bar{x} \pm t_{\alpha/2,df}\, s/\sqrt{n}$$ **or equivalently** $$\left[\bar{x} - t_{\alpha/2,df}\, s/\sqrt{n},\ \bar{x} + t_{\alpha/2,df}\, s/\sqrt{n}\right]$$

where *s* is the sample standard deviation.

LO 8.6

8.3 Confidence Interval of the Population Mean When σ Is Unknown

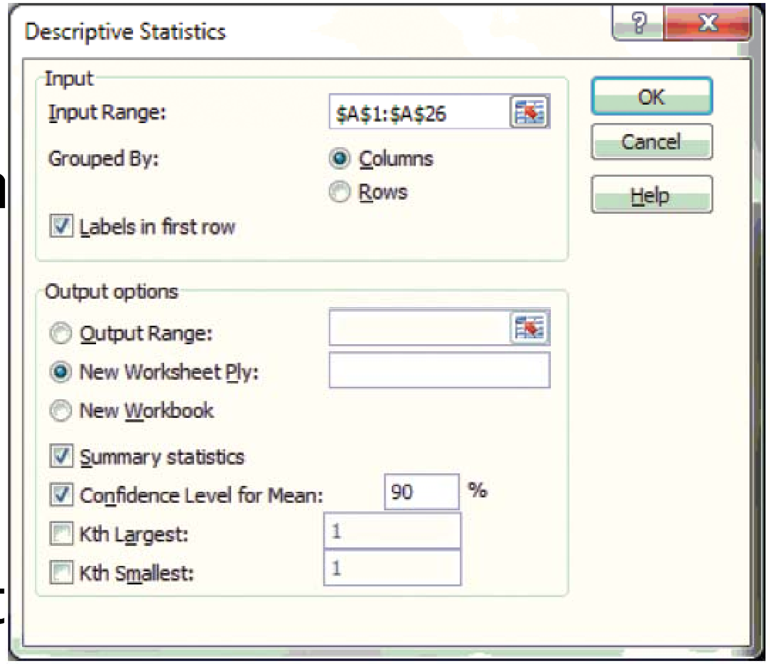
- Example: Recall that Jared Beane wants to estimate the mean mpg of all ultra-green cars. Use the sample information to construct a 90% confidence interval of the population mean. Assume that mpg follows a normal distribution.
  - Solution: Since the population standard deviation is not known, the sample standard deviation has to be computed from the sample. As a result, the 90% confidence

$$\bar{x} \pm t_{\alpha/2,df}\, s/\sqrt{n} = 96.52 \pm 1.711\left(10.70/\sqrt{25}\right) = 96.52 \pm 3.66$$

LO 8.6

8.3 Confidence Interval of the Population Mean When σ Is Unknown

- Using Excel to construct confidence intervals. The easiest way to estimate the mean when the population standard deviation is unknown is as follows:
  - Open the *MPG* data file.
  - From the menu choose Data Data Analysis > Descriptive Statistics > OK.
  - Specify the values as shown here and click OK.
  - Scroll down through the out until you see the Confidence Interval.



8.4 Confidence Interval of the Population Proportion

LO 8.7 Calculate a confidence interval for the population proportion.

- Let the parameter p represent the proportion of successes in the population, where success is defined by a particular outcome.
- $\bar{P}$ is the point estimator of the population proportion p.
- By the central limit theorem, $\bar{P}$ can be approximated by a normal distribution for large samples (i.e., $np \geq 5$ and $n(1 - p) \geq 5$).

LO 8.7

8.4 Confidence Interval of the Population Proportion

- Thus, a 100(1–α)% confidence interval of the population proportion is

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

or

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

where $\bar{p}$ is used to estimate the population parameter *p*.

LO 8.7

8.4 Confidence Interval of the Population Proportion

- Example: Recall that Jared Beane wants to estimate the proportion of all ultra-green cars that obtain over 100 mpg. Use the sample information to construct a 90% confidence interval of the population proportion.
  - Solution: Note that $\bar{p} = 7/25 = 0.28.$ In addition, the normality assumption is met since $np \geq 5$ and $n(1 - p) \geq 5$. Thus,

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}\left(1-\bar{p}\right)}{n}} = 0.28 \pm 1.645\sqrt{\frac{0.28\left(1-0.28\right)}{28}} = 0.28 \pm 0.148$$

8.5 Selecting a Useful Sample Size

LO 8.8 Select a sample size to estimate the population mean and the population proportion.

- **Precision** in interval estimates is implied by a low margin of error.
- The larger *n* reduces the margin of error for the interval estimates.
- How large should the sample size be for a given margin of error?

LO 8.8

8.5 Selecting a Useful Sample Size

- **Selecting *n* to Estimate μ**
  - **Consider a confidence interval for μ with a known σ and let *D* denote the desired margin of error.**
  - **Since** $D = z_{\alpha/2}\,\sigma/\sqrt{n}$

    we may rearrange to get $n = \left(\dfrac{z_{\alpha/2}\sigma}{D}\right)^2$
  - **If σ is unknown, estimate it with $\hat{\sigma}$.**

LO 8.8

8.5 Selecting a Useful Sample Size

- **Selecting *n* to Estimate μ**
  - **For a desired margin of error *D*, the minimum sample size *n* required to estimate a 100(1 $-\alpha$)% confidence interval of the population mean μ is**

$$n = \left(\frac{z_{\alpha/2}\hat{\sigma}}{D} \right)^2$$

Where $\hat{\sigma}$ is a reasonable estimate of σ in the planning stage.

LO 8.8

8.5 Selecting a Useful Sample Size

- **Example: Recall that Jared Beane wants to construct a 90% confidence interval of the mean mpg of all ultra-green cars.**
  - **Suppose Jared would like to constrain the margin of error to within 2 mpg. Further, the lowest mpg in the population is 76 mpg and the highest is 118 mpg.**
  - **How large a sample does Jared need to compute the 90% confidence interval of the population mean?**

$$n = \left(\frac{z_{\alpha/2}\hat{\sigma}}{D}\right)^2 = \left(\frac{1.645 \times 10.50}{2}\right)^2 = 74.58 \text{ or } 75$$

LO 8.8

8.5 Selecting a Useful Sample Size

- **Selecting *n* to Estimate *p***
  - **Consider a confidence interval for *p* and let *D* denote the desired margin of error.**
  - **Since** $D = z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$ **where** $\bar{p}$ **is the sample proportion**

    we may rearrange to get $n = \left(\frac{z_{\alpha/2}}{D}\right)^2 \bar{p}(1-\bar{p})$
  - **Since** $\bar{p}$ **comes from a sample, we must use a reasonable estimate of *p*, that is,** $\hat{p}$ **.**

LO 8.8

8.5 Selecting a Useful Sample Size

- **Selecting *n* to Estimate *p***
  - **For a desired margin of error *D*, the minimum sample size *n* required to estimate a 100(1 – α)% confidence interval of the population proportion *p* is**

$$n = \left(\frac{z_{\alpha/2}}{D}\right)^2 \hat{p}(1-p)$$

  - **Where $\hat{p}$ is a reasonable estimate of *p* in the planning stage.**

LO 8.8

8.5 Selecting a Useful Sample Size

- **Example: Recall that Jared Beane wants to construct a 90% confidence interval of the proportion of all ultra-green cars that obtain over 100 mpg.**
  - **Jared does not want the margin of error to be more than 0.10.**
  - **How large a sample does Jared need for his analysis of the population proportion?**

$$n = \left(\frac{z_{\alpha/2}}{D}\right)^2 p(1-p) = \left(\frac{1.645}{0.10}\right)^2 0.50(1-0.50) = 67.65 \text{ or } 68$$

End of Chapter