Chapter 10

Introduction of Inference
Two Most Common Types of Formal Statistical Inference

- **Section 10.1**
  - Confidence Intervals
    - Estimates the value of a population parameter

- **Section 10.2**
  - Tests of Significance
    - Assess the evidence for a claim about a population
Lesson 10-1, Part 1

Estimating with Confidence
Suppose I want to know how often teenagers go to the movie. Specifically, I want to know how many times per month a typical teenage (ages 13 through 17) goes to the movies.

- Suppose I take an SRS of 100 teenagers and calculate the sample mean to be \( \bar{x} = 2.1 \)

- The sample mean is an unbiased estimator of the unknown population mean \( \mu \), so I would estimate the population mean to be approximately 2.1.

- However a different sample would have given a different sample mean, so I must consider the amount of variation in the sampling model for \( \bar{x} \).
Estimating With Confidence

- The sampling model for $\bar{x}$ is approximately normal.
- The mean of the sampling model is $\mu$.
- The standard deviation of the sampling model is $\frac{\sigma}{\sqrt{n}}$ assuming the population size at least 10n.
Confidence Intervals

- Suppose we know that the population standard deviation is \( \sigma = 0.50 \). Then the standard deviation for the sampling model is

\[
\sigma = \frac{0.50}{\sqrt{n}} = \frac{0.50}{\sqrt{100}} = 0.05
\]

- Then 95\% of our samples will produce a statistic \( \bar{x} \) that is between \( \mu - 0.10 \) and \( \mu + 0.10 \).
Margin of Error

- Therefore in 95% of our samples, the interval between $\bar{x} \pm 0.10$ will contain the parameter $\mu$.

- The **margin of error** is 0.10.
  - Shows how accurate we believe our guess is, based on the variability of the estimate.
Confidence Interval

- For our sample of 100 teenagers, $\bar{x} = 2.1$. Because the margin of error is 0.10, then we are 95% confident that the true population mean lies somewhere in the interval $2.1 \pm 0.10$, or [2.0, 2.2].

- The interval [2.0, 2.2] is a 95% confidence interval because we are 95% confident that the unknown $\mu$ lies between 2.0 and 2.2.
Level $C$ Confidence Interval

- Start with sample data.
- Compute an interval that has probability $C$ of containing the true value of the parameter.
- This is called a level $C$ confidence interval.
Young people have a better chance of full-time employment and good wages if they are good with numbers. How strong are the quantitative skills of young Americans of working age? One source of data is the National Assessment of Educational Progress (NAEP) Young Adult Literacy Assessment Survey, which is based on a nationwide probability sample of households. The NAEP survey includes a short test of quantitative skills, covering mainly basic arithmetic and the ability to apply it to realistic problems.
Scores on the test range from 0 to 500. For example, a person who scores 233 and add the amounts of two checks appearing on a bank deposit slip; some scoring 325 can determine the price of a meal from a menu; a person scoring 375 can transform a price in cents per ounce into dollars per pound.

Suppose you give the NAEP test to an SRS of 840 people from a large population in which the scores have mean 280 and standard deviation \( \sigma = 60 \). The sample mean of 840 scores will vary if you take repeated samples.
Example – Page 542, #10.2

A) Describe the shape, center and spread of the sampling distribution $\bar{x}$.

The sampling distribution of $\bar{x}$ is normal with mean $\mu = 280$ and $\sigma_{\bar{x}} = 2.1$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{60}{\sqrt{840}} = 2.1$$
Example – Page 542, #10.2

B). Sketch the normal curve that describes how $\bar{x}$ varies in many samples from this population. Mark its mean, and the values one, two and three standard deviations on either side of the mean.

$n = 840$
$\mu = 280$
$\sigma = 60$
$\sigma_{\bar{x}} = 2.1$
According to the 68-95-99.7 rule, about 95% of all the values of $\overline{x}$ fall within **2 standard deviations** of the mean of this curve. What is the missing number? Call it $m$ for “margin of error.” Shade the region from the mean minus $m$ to the mean plus $m$ on the axis of your sketch, as in Figure 10.2

$$m = 2(2.1) = 4.2$$
Example – Page 542, #10.2
D). Whenever $\bar{x}$ falls in the region you shaded, the true value of the population mean $\mu = 280$, lies in the confidence interval between $\bar{x} - m$ and $\bar{x} + m$. Draw the confidence level below you sketch for one value of $\bar{x}$ inside the shaded region and one value of $\bar{x}$ outside the shaded region. Use figure 10.4 as a model for the drawing.
Example – Page 542, #10.2

![Gaussian distribution diagram with values 273.7, 275.8, 277.9, 280, 282.1, 284.2, 286.3]
E). In what percent of all samples will the true mean $\mu = 280$ be covered by the confidence interval $\bar{x} \pm m$?

95% (by the 68-95.99.7 rule)
Conditions for Constructing a Confidence Interval for $\mu$

- The data comes from a SRS from the population of interest, and
- The sample distribution of $\bar{x}$ is approximately normal.
  - the population is normally distributed, or
  - Central Limit Theorem ($n \geq 30$)
How do we construct confidence intervals?

- Since the sampling model of the sample mean $\bar{x}$ is approximately normal, we can use normal calculations to construct confidence intervals.
- For a 95% confidence intervals, we want the interval corresponding to the middle 95% of the normal curve.
- For a 90% confidence intervals, we want the interval corresponding to the middle 90% of the normal curve.
How do we construct confidence intervals?

- If we are using the standard normal curve, we want to find the interval using the \textit{z-values}.

- Suppose we want to find a 90\% confidence interval for a standard normal curve.
  - If the middle 90\% lies within our interval, then the remaining 10\% lies outside our interval.
  - Because the curve is symmetric, there is 5\% below the interval and 5\% above the interval.

- Find the \textit{z-values} with area 5\% below and 5\% above.
Z – Values

- These **z-values** are denoted by \(\pm z^*\).
- Because they come from the standard normal curve, they are centered at mean 0.
- \(z^*\) is called the **upper \(p\) critical value**, with probability \(p\) lying to its right under the standard normal curve.
- To find \(p\), we find the complement of \(C\) and divide it in half or \(\frac{1-C}{2}\).
Finding $Z^*$ standard deviations on either side of the mean.

The central probability $C$ under a standard normal curve lies between $-z^*$ and $z^*$. 
The number $z^*$ with probability $p$ lying to its right under the standard normal curve is called the **upper $p$ critical value** of the standard normal curve.
### Most Common Confidence Levels

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Tail Area</th>
<th>Z*</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.05</td>
<td>1.645</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>0.025</td>
<td>1.960</td>
</tr>
<tr>
<td></td>
<td>2.5%</td>
<td></td>
</tr>
<tr>
<td>99%</td>
<td>0.005</td>
<td>2.576</td>
</tr>
<tr>
<td></td>
<td>0.50%</td>
<td></td>
</tr>
</tbody>
</table>
invnorm(0.10, 0, 1) = -1.28 \quad invorm(1 - .10, 0, 1) = 1.28

\[ P = \frac{1 - .80}{2} = 0.10 \]

\[ C = .80 \]

Probability = 0.1

Probability = 0.8

Probability = 0.1

\[ z^* = 1.28 \]
Constructing a 80% Confidence Interval

- To construct a 80% confidence interval, we want to find values 1.28 standard deviations below mean and 1.28 standard deviations above the mean, or $\mu \pm 1.28\sigma$

- Using our sample data, this is $\bar{x} \pm 1.28\frac{\sigma}{\sqrt{n}}$
One-Sample z Confidence Interval for $\mu$

The general formula for a confidence interval for a population mean $\mu$ when

- $\bar{x}$ is the sample mean from a random sample,
- the sample size $n$ is large (generally $n \geq 30$), and
- $\sigma$, the population standard deviation, is known.

A level $C$ confidence level for $\mu$ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$
Confidence Interval for a Population Mean

\[ CI = estimate \pm \text{margin of error} \]
\[ CI = estimate \pm z \times \frac{\sigma}{\sqrt{n}} \]
\[ CI = \bar{x} \pm z \times \frac{\sigma}{\sqrt{n}} \]
Distribution of Sample Means with Known $\sigma$

$E = \text{Margin of Error}$
Lesson 10-1, Part 2

Estimating with Confidence
To construct a confidence interval

- Identify the population of interest and the parameter you want to draw conclusions about.
- Choose the appropriate inference procedure.
  - Verify the conditions for using the selected procedure.
- If the conditions are met carry, out the inference procedure.
  - CI = estimate ± margin of error
- Interpret your results in the context of the problem.
A study of the career paths of hotel general managers sent questionnaires to an SRS of 160 hotels belonging to major U.S. hotel chains. There were 114 responses. The average time these 114 general managers had spent with their current company was 11.78 years. Give a 99% confidence interval for the mean number of years general managers of major-chain hotels have spent with their current company. (Take it as known that the standard deviation of time with company for all general managers is 3.2 years.) *Use the Inference Toolbox as a guide.*

**Step 1 – Identify the population of interest and the parameter you want to draw conclusions about.**

\[ \mu = \text{meantime general managers spent with their current company}. \]
Step 2 – Choose the appropriate inference procedure. Verify the conditions for using the selected procedure.

One sample z-interval

Conditions:
1. SRS from the population of interest
2. The Central Limit Theorem tells us that the sampling distribution of $\bar{x}$ is approximately normal since $n \geq 30$

$\bar{x} = 11.78 \text{ years}$
$\sigma = 3.2 \text{ years}$
$n = 114$
$CL = 0.99$
Step 3 – If the conditions are met, carry out the inference procedure.

Find the confidence level

\[
CI = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 11.78 \pm 2.576 \frac{3.2}{\sqrt{114}}
\]

\[
\bar{x} = 11.78 \text{ years} \\
\sigma = 3.2 \text{ years} \\
n = 114 \\
CL = 0.99
\]

STAT/TESTS/7:ZInterval

EDIT  CALC  TESTS
1:2-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-ProPZTest...
6:2-ProPZTest...
7:ZInterval...

ZInterval
Inpt:Data
\[\begin{align*}
\bar{x} &= 11.78 \\
n &= 114 \\
C-Level &= 0.99
\end{align*}\]

ZInterval
\[\begin{align*}
\bar{x} &= 11.78 \\
n &= 114 \\
(11.008, 12.552)
\end{align*}\]
Step 4 – Interpret your results in the context of the problem.

We are 99% confident that the mean years that a general hotel manager has been with the company is between 11.01 and 12.55 years.
How Confidence Intervals Behave

- We would like **high** confidence level and **small** margin of error.
  - High confidence level means a higher percentage of all samples produce a statistic close to the true value of the parameter.
  - Smaller margin of error allows us to get closer to the true value of the parameter.

Higher confidence level requires a longer interval.
So how do we reduce the margin of error?

- Lower the confidence level (by decreasing the value of $z^*$).
- Lower the standard deviation.
- Increase the sample size.
  - To cut the margin of error in half, increase the sample size by four times the previous size.
- You can have high confidence and a small margin of error if you choose the right sample size.
Crop researchers plant 15 plots with a new variety of corn. The yields in bushels per acre are

138.0  139.1  113.0  132.5  140.7  109.7  118.9  134.8  109.6
127.3  115.6  130.4  130.2  111.7  105.5

Assume that \( \sigma = 10 \) bushels per acre

A). Find the 90% confidence interval for the mean yield \( \mu \) for this variety of corn.

Use a ZInterval
Example – Page 550, #10.8

B). Find the 95% confidence interval.

\[
\text{ZInterval} \\
\text{Inpt: List Stats} \\
\sigma: 10 \\
\text{List: L1} \\
\text{Freq: 1} \\
\text{C-Level: .9}
\]

\[
\text{ZInterval} \\
(119.55, 128.05) \\
\bar{x} = 123.8 \\
S \bar{x} = 12.26038918 \\
n = 15
\]
C). Find the 99% confidence interval.

\[
\begin{align*}
Z\text{Interval} \\
(117.15, 130.45) \\
\bar{x} = 123.8 \\
S_x = 12.26038918 \\
n = 15
\end{align*}
\]
D). How do the margin of error in a, b, and c change as the confidence level increases.

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Interval</th>
<th>Margin of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>119.6 to 128.0 bu/acre</td>
<td>4.2</td>
</tr>
<tr>
<td>95%</td>
<td>118.7 to 128.9 bu/acre</td>
<td>5.1</td>
</tr>
<tr>
<td>99%</td>
<td>117.1 to 130.5 bu acre</td>
<td>6.7</td>
</tr>
</tbody>
</table>

The margin of error increases with the confidence level.
Choose the Sample Size

- To determine the sample size $n$ that will yield a confidence interval for a population mean with a specific margin of error $m$, set the expression for the margin of error to less than or equal to $m$ and solve for $n$:

$$z^* \frac{\sigma}{\sqrt{n}} \leq m$$

$$n \geq \left[ \frac{z^* \sigma}{m} \right]^2$$
To assess the accuracy of a laboratory scale, a standard weight known to weigh 10 grams is weighed repeatedly. The scale readings are normally distributed with unknown mean (this mean is 10 grams if the scale has no bias). The standard deviation of the scale reading is known to be 0.0002 gram.

A). The weight is weighed five times. The mean result is 10.0023 grams. Give a 98% confidence interval for the mean of repeated measurements of the weight.

\[
\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 10.0023 \pm 2.326 \frac{0.0002}{\sqrt{5}}
\]

between 10.002 and 10.003
B). How many measurements must be averaged to get a margin of error of ±0.0001 with 98% confidence?

\[
\sigma = 0.0002 \\
m = 0.0001
\]

\[
n > \left[ \frac{z \sigma}{m} \right]^2
\]

\[
n > \left[ \frac{(2.326)0.0002}{0.0001} \right]^2
\]

\[
n > 21.64
\]

\[
n > 22
\]
Caution

- These methods only apply to certain situations. In order to construct a level $C$ confidence interval using the formula:

$$\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$$

- The data must be an SRS and we must know the population standard deviation.
- Also, we want to eliminate (if possible) any outliers.
- The margin of error only covers random sampling errors.
  - Things like undercoverage, nonresponse, and poor sampling designs can cause additional errors
A student reads that a 95% confidence interval for the mean SAT math score of California high schools seniors is 452 to 470. Asked to explain the meaning of this interval, the student says, “95% of California high school seniors have SAT Math scores between 452 to 470.” Is the student right? Justify your answer.

No, the interval refers to the mean math scores, not individual scores, which will be much more variable (if more than 95% of the students score below 470, they are not doing very well).
Lesson 10-2, Part 1

Tests of Significance
Introduction

- Use a random sample to learn something about a larger population.

- Two ways to learn about a population
  - Confidence Intervals
    - Estimates the value of a population parameter
  - Tests of Significance (Hypothesis Testing)
    - Assess the evidence for a claim about a population
Difference between the confidence intervals and tests of significance

- Confidence interval allows us to use sample data to estimate a population value, like the true mean or the true proportion.
  
  *Example*: What is the true average amount students spend weekly on gas?

- Significance testing allows us to use sample data to test a claim about a population, such as testing whether a population proportion or population mean equals some number.

  *Example*: Is the true average amount that students spend weekly on gas $20.
General Idea of Significance Testing

- A claim is made.
- Evidence (sample data) is collected in order to test the claim.
- The data are analyzed in order to support or refute the claim.
- Decide whether or not the initial claim is reasonable.
  - An outcome that would rarely happen if a claim were true is good evidence that the claim is not true.
Hypothesis

- A **hypothesis** is a statement or claim regarding a characteristic of a population.
- The claims that we test are the population mean, population proportion and the population standard deviation.
Illustrating Hypothesis Testing

- The packaging on a light bulb states that the light bulb will last 500 hours under normal use. A consumer advocate would like to know if the mean lifetime of a bulb is less than 500 hours.

- To test the claim the consumer advocate might obtain a random sample of 49 light bulbs and determines that the mean length of time before they burned out is 476 hours.

- Is this enough evidence to conclude that the bulb’s life is less than 500 hours?
Illustrating Hypothesis Testing

- If we give the manufacture the benefit of the doubt, the mean lifetime of the bulb would be $\mu_x = 500$.

- Let’s assume that the standard deviation of the bulb is $\sigma = 42$ hours, so the standard deviation of $\bar{x}$ is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{49}} = 6$$
The sample mean of 476 is so far from the claimed mean we have reason to believe that the manufacture’s claim is probable not true. The consumer advocate would conclude that the population mean is likely some value less than 500 hours.
Illustrating Hypothesis Testing

- The sample mean of 476 hours is “too far” from the claimed mean, while 491 hours is “close enough” for us to believe that the population mean could be 500 hours.

- This raises an interesting question. How far from the presumed mean of 500 hours can our sample mean be before we reject the idea that the bulb lasts 500 hours?

  - To answer this question requires test of significance or hypothesis testing.
Hypothesis Testing

- **Significance Test or Hypothesis Testing** is a procedure, based on sample evidence and probability to test claims regarding a characteristic of one or more populations.
Two Types of Hypotheses

- **Null Hypothesis** $H_0$
  - It is the statement to be tested.
  - It is assumed true until evidence we have indicated otherwise
  - No effect or no change in the population

- **Alternative Hypothesis** $H_a$
  - It is the claim to be tested.
  - It is the conclusion that the researcher is trying to make.
  - Alternative to “no effect” or “no change”
Criminal Trial Analogy

- First, State the 2 hypothesis
  - \( H_0 \): Defendant is not guilty
  - \( H_a \): Defendant is guilty.

- Then, collect evidence such finger prints, blood spots, hair samples, etc.
  - In statistics, the **data** is the evidence

- Then, make initial assumption.
  - Defendant is innocent until proven guilty.
  - In statistics, we always **assume the null hypothesis is true**.
Criminal Trial Analogy

- Then, make a decision based on the available evidence.
  - If there is sufficient evidence “beyond a reasonable doubt”.
    - Behave as if defendant is guilty
    - (Reject the null hypothesis).
  - If there is not enough evidence.
    - Behave as if defendant is not guilty
    - (Fail to reject the null hypothesis)
Important Points

- Neither decision entails proving the null hypothesis or the alternative hypothesis. We merely state there is enough evidence to behave one way or the other.

- This is also always true in statistics! No matter what decision we make, there is always a chance we made an error.
Lesson 10-2, Part 2

Tests of Significance
Right-Tailed Test

\[ H_0 = _____ \]

\[ H_a > _____ \]

Fail to reject \( H_0 \)

Reject \( H_0 \)

Sign used in \( H_1: > \)
Left-Tailed Test

\[ H_0 = \____ \]
\[ H_a < \____ \]
Two-Tailed Test

\[
H_0 = \_
\]
\[
H_a \neq \_
\]

Sign used in \( H_1: \neq \)
Example – Page 564, #10.28

The census Bureau reports that households spend an average of 31% of their total spending on housing. A homebuilders association in Cleveland believes that this average is lower in their area. They interview a sample of 40 households in Cleveland metropolitan area to learn what percent of spending is devoted to housing among all Cleveland households. We want to test the hypothesis

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]

The population standard deviation is \( \sigma = 9.6\% \)
Example – Page 564, #10.28

\[ H_o : \mu = 31\% \quad n = 40 \]
\[ H_a : \mu < 31\% \quad \sigma = 9.6\% \]

A). What is the sampling distribution of the mean percent \( \bar{x} \) that the sample spends on housing if the null hypothesis is true? Sketch the density curve of the sampling distribution. (Hint: Sketch a normal curve first, then mark the axis using what you know about locating \( \mu \) and \( \sigma \) on a normal curve.)

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9.6}{\sqrt{40}} = 1.518\% 
\]
Example – Page 564, #10.28

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]

\[ N(31\%, 1.518\%) \]

\[ n = 40 \]
\[ \sigma = 9.6\% \]
\[ \sigma_x = 1.518\% \]
Example – Page 564, #10.28

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]

\[ n = 40 \]
\[ \sigma = 9.6\% \]
\[ \sigma_x = 1.518\% \]

B). Suppose that the study finds \( \bar{x} = 30.2\% \) for the 40 households in the sample. Mark this point on the axis of your sketch. Then suppose that the study result is \( \bar{x} = 27.6\% \). Mark this point on the axis of your sketch. Referring to your sketch, explain in simple language why one result is good evidence that average Cleveland spending on housing is less than 31%, whereas the other result is not.
Example – Page 564, #10.28

The lower percentage lies out in the low tail of the curve, while 30.2% is fairly close to the middle. Assume $H_0$ is true, observing a value like 30.2% would be surprising, but 27.6% is unlikely, and therefore provides evidence against $H_0$. 

\[ n = 40 \]
\[ \sigma = 9.6\% \]
\[ \sigma_x = 1.518\% \]
\[ \bar{x}_1 = 30.2\% \]
\[ \bar{x}_2 = 27.6\% \]
\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]
Example – Page 564, #10.28

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]

C). Shade the area under the curve that gives the P-value for the result \( \bar{x} = 30.2\% \) (Note that we are looking for evidence that spending less than the null hypothesis states.)

\[ n = 40 \]
\[ \sigma = 9.6\% \]
\[ \sigma_{\bar{x}} = 1.518\% \]
$n = 40$

$\sigma = 9.6\%$

$\sigma_x = 1.518\%$

$\bar{x}_1 = 30.2\%$

$\bar{x}_2 = 27.6\%$

$H_o : \mu = 31\%$

$H_a : \mu < 31\%$
What is a $P$-value?

We measure the strength of the evidence against the null hypothesis by the probability ($P$-value) under the normal curve.
P-Values or Probability Values

- The probability, is computed assuming that $H_0$ is true, The observed outcome would take a value as extreme or more extreme than that actually observed is called the $P$-value of the test.
- The $p$-value is the probability of getting a value that is extreme or more extreme than the one observed under the assumption that the null hypothesis is true.
- The $p$-value represents how likely we would be to observe such an extreme sample if the null hypothesis were true.
- A small $p$-value is evidence against the null hypothesis because it says that the observed result is unlikely to occur by just chance.
- A result with a small $p$-value (less than 0.05) is called statistically significant.
  - The chance of producing an extreme result is unusual.
Significance Level

- Is a fixed value on how much evidence against $H_0$ that we will require.
- We will compare the $P$-value to this fixed value.
- This fixed value is represented by the Greek letter alpha $\alpha$
  - Use $\alpha = 0.05$, we are requiring that the data give evidence against $H_0$ so strong that it would happen no more than 5% of the time (5 times in 100 samples in the long run) when $H_0$ is true.
  - Use $\alpha = 0.01$ we are insisting on stronger evidence against $H_0$, evidence so strong that it would appear only 1% of the time (1 time in 100 samples) if $H_0$ is in fact true.
Statistical Significance

If the $P$-value is small or smaller than alpha, we say that the data is **statistically significant** at level $\alpha$.

- $P$-value $\leq \alpha$ the result is statistically significant
- $P$-value $> \alpha$ the result is not statistically significant
Return to exercise 10.28, page 564.

The census Bureau reports that households spend an average of 31\% of their total spending on housing. A homebuilders association in Cleveland believes that this average is lower in their area. They interview a sample of 40 households in Cleveland metropolitan area to lean what percent of spending devoted to housing among all Cleveland households. We want to test the hypothesis

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]

The population standard deviation is \( \sigma = 9.6\% \)
A). Starting from the picture you drew there, calculate the p-value for both \( \bar{x} = 30.2\% \) and \( \bar{x} = 27.6\% \). The two P-values express in numbers the comparison you made informally in exercise 10.28.

\[
H_o : \mu = 31\%
\]

\[
H_a : \mu < 31\%
\]

\[\sigma = 9.6\%\]
Example – Page 569, #10.34

Finding the P-value

\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{9.6}{\sqrt{40}} = 1.51789 \]

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]
\[ \sigma = 9.6\% \]

\[ \text{normalcdf} (-E99, 30.2, 31, 1.51789) = 0.2991 \]
\[ \text{normalcdf} (-E99, 27.6, 31, 1.51789) = 0.0125 \]
B). Is the result \( \bar{x} = 27.6\% \) statistically significant at the \( \alpha = 0.05 \) level?

Is it significant at the \( \alpha = 0.01 \) level?

Since \( 0.0125 \leq 0.05 \) the result is significant at the 0.05 level

Since \( 0.0125 > 0.01 \) the result is not significant at the 0.01 level

\[ H_o : \mu = 31\% \]
\[ H_a : \mu < 31\% \]
\[ \sigma = 9.6\% \]
\[ P-value = 0.0125 \]
The Logic of Hypothesis Testing

- In order to test hypothesis regarding the population mean, we need certain requirements to be satisfied.
  - A simple random sample is obtained
  - The population from which the sample is drawn is normally distributed or the sample size is large ($n \geq 30$);
  - The population standard deviation, $\sigma$, is known

- If these requirements are met then the distribution of $\bar{x}$ is normal with mean $\mu$ and standard deviation

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]
Decision to Reject or Fail to Reject

- **Fail to Reject** $H_0$
  - If the test statistic does not fall within the critical region.
  - If the p-value > $\alpha$ (where $\alpha$ is the significance level).

- **Reject** $H_0$
  - If the test statistic falls within the critical region.
  - If the p-value $\leq \alpha$ (where $\alpha$ is the significance level).
A pharmaceutical manufacturer forms tablets by compressing a granular material that contains the active ingredient and various fillers. The hardness of a sample from each lot of tablets produced is measured in order to control the compression process. The target values for the hardness are $\mu = 11.5$ and $\sigma = 0.2$. The hardness data for a sample of 20 tablets are


Is there significant evidence at the 5% level that the mean hardness of the tablets is different from the target value? Use the inference toolbox
Step 1 – Identify the population of interest and the parameter you want to draw conclusions about. Identify the null and alternative hypothesis in words and symbols.

\[ M = \text{the mean tablet hardness in this company.} \]

\[ H_0: \mu = 11.5 \]
\[ H_a: \mu \neq 11.5 \]
Step 2 – Choose the appropriate inference procedure. Verify the conditions for using the selected procedure.

one-sample z test

Conditions:

- Assuming that the data came from an SRS
- Sampling distribution is approximately normal since the normal probability plot shows a linear trend with no outliers.
Using Your Calculator

STAT/TESTS/Z-TEST

EDIT CALC TEST
1: Z-Test...
2: T-Test...
3: 2-SampZTest...
4: 2-SampTTest...
5: 1-PropZTest...
6: 2-PropZTest...
7: ZInterval...

Z-Test
Inpt: UBF Stats
μ₀: 11.5
σ: 2
List: L1
Freq: 1
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw

μ₀: 11.5
z = 0.3667151483
p = 0.7138316189
x = 11.5164
Sx = 0.0949550586
n = 20
Example – Page 576, #10.38

Calculate the test statistic

\[ z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{11.5164 - 11.5}{0.2 / \sqrt{20}} = 0.3667 \]

STAT/TESTS/Z-TEST

EDIT CALC Z-TEST
1:Z-Test...
2: T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-ProfZTest...
6:2-ProfZTest...
7:2Interval...

Z-Test
Inpt: \bar{x} = 11.5164, \sigma = 11.5, \sigma = 0.2, n = 20
List: L1
Freq: 1
Statistic: z = 0.3667

P = 0.7138316189
X = 11.5164
Sx = 0.0949550586
n = 20
Step 4 – Interpret your results in the context of the problem

There is sufficient evidence to fail to reject $H_0$, since the p-value $> \alpha$ (0.714 $> 0.05$) and conclude mean tablet hardness is equal to 11.5.

A mean score as high as 11.52 would occur about 71 times in a 100 samples if the population mean were 11.5.
Lesson 10-2, Part 3

Tests of Significance
Confidence Interval and Two Sided Tests

- Recall the level of confidence in a confidence interval is a probability that represents the percentage of intervals that will contain $\mu$ if repeated samples are obtained.
- The level of confidence is denoted by $(1 - \alpha)100\%$.
  - Example: If $\alpha = 0.05$ the level of confidence is 95%.
- We can use confidence intervals to test $H_0: \mu = \mu_o$ versus $H_a: \mu \neq \mu_o$.
When testing $H_0: \mu = \mu_o$ versus $H_a: \mu \neq \mu_o$

- If a $(1 - \alpha)100\%$ confidence interval contains $\mu_o$, then we do not reject the null hypothesis.

- If the confidence interval does not contain $\mu_o$, then we have evidence that supports the claim stated in the alternative hypothesis and conclude $\mu \neq \mu_o$ at the level of significance, $\alpha$. 

Confidence Interval and Two Sided Tests
Example – Page 582, #10.44

Here are the IQ test scores of 31 seventh-grade girls in a Midwest school district:

<table>
<thead>
<tr>
<th>114</th>
<th>100</th>
<th>104</th>
<th>89</th>
<th>102</th>
<th>91</th>
<th>114</th>
<th>114</th>
<th>103</th>
<th>105</th>
<th>108</th>
<th>130</th>
<th>120</th>
<th>132</th>
<th>111</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>118</td>
<td>119</td>
<td>86</td>
<td>72</td>
<td>111</td>
<td>103</td>
<td>74</td>
<td>112</td>
<td>107</td>
<td>103</td>
<td>98</td>
<td>96</td>
<td>112</td>
<td>112</td>
<td>93</td>
<td></td>
</tr>
</tbody>
</table>

Treat the 31 girls as an SRS of all seventh-grade girls in the school district. Suppose that the standard deviation of IQ scores in this population is known to be $\sigma = 15$. 
Example – Page 582, #10.44

A). Give a 95% confidence interval for the mean IQ score \( \mu \) in the population

\[
\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 105.84 \pm 1.96 \frac{15}{\sqrt{31}} = 105.84 \pm 5.28
\]

We are 95% confident that this interval captures the population mean, \( \mu \).
Example – Page 582, #10.44

B). Is there significant evidence at the 5% level that the mean IQ score in the population differs from 100. Give appropriate statistical evidence to support your conclusion.

\[
H_o : \mu = 100 \\
H_a : \mu \neq 100
\]

Check conditions:
1. Given SRS
2. \( n \geq 30 \)
3. \( \sigma \) known – use a Z - Test

\[
n = 31 \\
\sigma = 15
\]
Example – Page 582, #10.44

$H_o : \mu = 100$

Reject $H_o$ since the mean does not fall within the 95% confidence interval
C). In fact, the scores are those of all seventh-grade girls in one of the several schools in the district. Explain carefully why your result from (a) and (b) cannot be trusted.

Since the scores all came from a single school, the sample may not be representative of the school district as a whole.
Example – Page 583, #10.48

Cobra Cheese Company buys milk from several suppliers. Cobra suspects that some producers are adding water to their milk to increase their profits. Excess water can be detected by measuring the freezing point of milk. The freezing temperature of natural milk varies normally, with mean \( \mu = -0.545 ^\circ \text{C} \) and standard deviation \( \sigma = -0.008 ^\circ \text{C} \). Added water raises the freezing temperature toward \( 0 ^\circ \text{C} \), the freezing point of water. Cobra’s laboratory manager measures the freezing temperature of five consecutive lots of milk from one producer. The mean measurement is \( -0.538 ^\circ \text{C} \). Is this good evidence that the producer is adding water to the milk?
Example – Page 583, #10.48

Step 1 – Identify the population parameter of interest. State your null hypothesis and alternative hypothesis.

\[ \mu = \text{mean freezing point of milk.} \]

\[ H_o : \mu = -0.545 \]

\[ H_a : \mu > -0.545 \]

\[ \mu = -0.545 \]

\[ \sigma = 0.008 \]

\[ \bar{x} = -0.538 \]

\[ n = 5 \]
Example – Page 583, #10.48

Step 2 – Identify the statistical procedure you should use. Then verify the conditions for using this procedure

One-sample Z test

Conditions

1. Not sure if its SRS, we may not be able to generalize our results
2. Sampling distribution is normally distributed since the population is normally distributed

\[ \mu = -0.545 \]
\[ \sigma = 0.008 \]
\[ \bar{x} = -0.538 \]
\[ n = 5 \]
Step 3 – Calculate the test statistic and the \( p \)-value. Illustrate with a graph.

\[
Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{-0.538 - (-0.545)}{0.008} = 1.957
\]

\[
z = 1.96
\]

\[
P - value = 0.0252
\]
Step 4 – State your conclusions clearly in complete sentences.

There is sufficient evidence to reject $H_0$, since the p-value $> \alpha$ ($0.0252 \leq 0.05$) and conclude mean freezing point of milk is greater than $-0.545$. 
Lesson 10-3

Making sense of statistical significance
Purpose of a Test of Significance

- Is to give a clear statement of the degree of evidence by the sample against the null hypothesis.
  - Make some decision or take some action if your evidence reaches a certain standard.
  - A level of significance $\alpha$ sets such standard.
Choosing a Level of Significance

- Choose $\alpha$ by asking how much evidence is required to reject $H_0$.
  - How plausible is $H_0$?
    - If $H_0$ represents an assumption that the people you must convince have believed for years, strong evidence (small $\alpha$) will be needed to persuade them.
  - What are the consequences of rejecting $H_0$?
    - If rejecting $H_0$ in favor of $H_a$ means making an expensive changeover from a type of product packaging to another, you need strong evidence that the new packaging will boost sales.
Choosing a Level of Significance

- Since different people may want to use different significant levels, it is better to report the $P$-value, which allows each of us to decide individually if the evidence is sufficiently strong.

- There is no practical distinction between the $P$-values 0.049 and 0.051.
A null hypothesis ("no effect" or "no difference") can be rejected at the usual levels, $\alpha = 0.05$ or $\alpha = 0.01$, there is good evidence that an effect is present. But the effect may be very small.

When large samples are available, even the tiny deviations from the null hypothesis will be significant.

Statistical significance is not the same thing as practical significance.
Let us suppose that SAT Math scores in the absence of coaching vary normally with mean $\mu = 475$ and $\sigma = 100$. Suppose also that coaching may change $\mu$ but does change $\sigma$. An increase in the SAT Math score from 475 to 478 is of no importance seeking admission to college, but this unimportant change can be statistically very significant. To see this, calculate the P-value for the test of

$$H_o : \mu = 475$$

$$H_a : \mu > 475$$

in each of the following situations
A). A coaching service coaches 100 students. Their SAT Math scores average $\bar{x} = 478$.

Use a Z-test

$p-value = 0.3821$
Example – Page 589, #10.58

B). By next year, the service has coached 1000 students. Their SAT Math scores average is still $\bar{x} = 478$.

$Ho : \mu = 475$

$Ha : \mu > 475$

$\mu = 475$

$\sigma = 100$

$\bar{x} = 478$

$n = 1000$

Use a Z-test

$p-value = 0.1714$
C). An advertising campaign brings the number of students coached to 10,000. There average score is still \( \bar{x} = 478 \).

Use a Z-test

\[ p-value = 0.0014 \]
Which of the following questions does a test of significance answer?

A). Is the sample or experiment properly designed?

B). Is the observed effect due to chance?

C) Is the observed effect important?
A local television station announces a question for a call-in opinion poll on the six o’clock news, and then gives the response on the eleven o’clock news. Today’s question concerns a proposed gun-control ordinance. Of the 2372 calls received, 1921 oppose the new law. The station, following standard statistical practice, makes a confidence statement: “81% of the Channel 13 Pulse Poll sample opposed gun control. We can be 95% confident that the proportion of all viewers who opposed the law is within 1.6% of the sample result.” Is the station’s conclusion justified? Explain your answer.
No, the percentage was based on a voluntary response sample, and so cannot be assumed to be a fair representation of the population. Such a poll is likely to draw a higher-than-actual proportion of people with a strong opinion, especially a strong negative opinion.
Lesson 10-4

1. Type I and Type II Errors
2. Error Probabilities
Type I and Type II Errors

- Nobody’s perfect. Even with lots of evidence, we can still make the wrong decision. In fact, when we perform a hypothesis test, we can make mistakes in two ways:
  - The null hypothesis is true, but we mistakenly reject it.
    - This is called a Type I Error
  - The null hypothesis is false, but we fail to reject it.
    - This is called a Type II Error
## Errors in Criminal Trials

<table>
<thead>
<tr>
<th>Jury Decision</th>
<th>Truth</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Guilty (H₀)</td>
<td>Guilty (Hₐ)</td>
</tr>
<tr>
<td>Guilty</td>
<td><strong>Type I Error</strong></td>
<td><img src="true" alt="Yes" /></td>
</tr>
<tr>
<td></td>
<td>Putting an innocent person in jail.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reject H₀ when H₀ is true</td>
<td></td>
</tr>
<tr>
<td>Not Guilty</td>
<td><img src="true" alt="Yes" /></td>
<td><strong>Type II Error</strong></td>
</tr>
<tr>
<td></td>
<td>Letting a guilty person go free</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fail to reject H₀ when Hₐ is true</td>
<td></td>
</tr>
</tbody>
</table>
Example – Type I and Type II Errors

- In medical disease testing, the null hypothesis is usually the assumption that a person is healthy.
- The alternative is that he or she has the disease we’re testing for.
- So a Type I error is a **false positive**.
  - A healthy person is diagnosed with the disease.
- A Type II error is a **false negative**.
  - An infected person is diagnosed as disease free.
- These errors have other names, depending on the particular discipline and context.
## Illustration of the Situations

<table>
<thead>
<tr>
<th>My Decision</th>
<th>Truth About the Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_o$ is true</td>
</tr>
<tr>
<td>Reject $H_o$</td>
<td>Wrong Decision</td>
</tr>
<tr>
<td></td>
<td>Type I Error</td>
</tr>
<tr>
<td></td>
<td>$\alpha$</td>
</tr>
<tr>
<td>Retain $H_o$</td>
<td>Correct Decision</td>
</tr>
<tr>
<td></td>
<td>$1 - \alpha$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$H_o$ is true</th>
<th>$H_a$ is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Decision</td>
<td>Power</td>
</tr>
<tr>
<td>$1 - \beta$</td>
<td></td>
</tr>
</tbody>
</table>

Wrong Decision | Type II Error |
|               | $\beta$      |
Your company markets a computerized medical diagnostic program. The program scans the results of routine medical tests (pulse rate, blood tests, etc.) and either clears the patient or refers the case to a doctor. The program is used to screen thousands of people who do no have specific medical complaints. The program makes a decision about each person.
A). What are the two hypothesis and the two types of error that the program can make? Describe the two types of errors in terms of “false positive” and “false negative” test results.

\( H_0 \): the patient is ill (or “the patient should see a doctor”)
\( H_a \): the patient is healthy (or “the patient should not see a doctor”)

Type I Error: clearing a patient who should be referred to a doctor – false negative

Type II Error: sending a healthy patient to see a doctor – false positive
B). The program can be adjusted to decrease one error probability at the cost of an increase in the other probability. Which one error probability would you choose to make smaller, and why? (This is a matter of judgment. There is no single correct answer.)

One might wish to lower the probability of a false negative so that most ill patients are treated. On the other hand, if money is an issue, or there is concern about sending too many patients to see the doctor, lowering the probability of false positives might be desirable.
How often will a Type I error occur?

- It happens when the null hypothesis is true.
  - But we’ve had the bad luck to draw an unusual sample.
- To reject $H_o$, the $P$-value must fall below $\alpha$.
  - When you choose level $\alpha$, you’re setting the probability of a Type I error to $\alpha$.
- What if $H_o$ is not true?
  - Then we can’t possibly make a Type I error.
    - You can’t get a false positive from a sick person.
- A Type I error can happen only when $H_o$ is true.
Power

- When $H_0$ is false and we reject it, we have done the right thing.
- Our ability to detect a false hypothesis is called the **power** of the test.
- In jury trial, power is the ability of the criminal justice system to convict people who are guilty.
Type II Error

- When $H_0$ is false and we fail to reject, we have made a Type II error.
- We assign the letter $\beta$ to the probability of this mistake.
- What is the value of $\beta$?
  - That’s harder to assess than $\alpha$ because we don’t know how false $H_0$ is.
- When $H_0$ is true, we have single parameter. But when $H_0$ is false, we don’t have one; we have many.
- We can compute the probability $\beta$ for any parameter value in $H_a$.
- But which one should we choose?
There is no single value for $\beta$.

- We can think of a whole collection of $\beta$’s and display them as curve, plotting each $\beta$ against the corresponding parameter value in $H_a$.
- One way to focus our attention is by thinking about the effect size.
  - That is, ask “How big a difference would matter?”
- Suppose a charity wants to test whether personalized address labels in the envelope along with a request for a donation increases the response rate above the baseline of 5%.
- If the minimum response that would pay for the address labels is 6%, we would calculate $\beta$ for the alternative $p = 0.06$. 
Type I Versus Type II Error

- We can reduce $\beta$ for all values in the alternative by increasing $\alpha$.
- If we make it easier to reject the null, then we’re more likely to reject it whether it’s true or not.
- So we’d reduce $\beta$, but we’d make more Type I errors.
- In the political arena, think of the ongoing debate between those who favor provisions to reduce:
  - Type I errors in the courts (Miranda rights, warrants require for wire taps, legal representation paid by the state if needed, etc.) and those who are concerned that
  - Type II errors have become too common (admitting into evidence confessions make when no lawyer is presents, eavesdropping on conferences with lawyers, restricting paths of appeals, etc.
- The only way to reduce both types of errors is to collect more evidence or in statistical terms, to collect more data. Otherwise, we just wind up trading off one kind of error for the other.
Power of a Test

- Remember, we can never prove a null hypothesis true.
  - We can only fail to reject it.
- When we fail to reject the null hypothesis, it’s natural to wonder whether we looked hard enough.
  - Might the null hypothesis actually be false and our test is too weak to tell?
- When the null hypothesis actually is false, we hope our test is strong enough to reject it.
  - We’d like to know how likely we are to succeed.
  - The power of the test gives us a way to think about it.
- The **power** of a test is the probability that it correctly rejects a false null hypothesis.
Power of a Test

- When the power is high, we can be confident that we’ve looked hard enough.
- We know that $\beta$ is the probability that a test fails to reject a false null hypothesis.
  - So the power of the test is the complement, $1 - \beta$.
- When a study fails to reject the null hypothesis, the test’s power comes into question.
  - Was the sample size big enough to detect an effect had there been one?
  - Might we have missed an effect of an interesting size just because we failed to gather sufficient data or because there was too much variability in the data we could gather?
Calculating Power

- When we calculate power, we have to imagine that the null hypothesis is false.
- The value of the power depends on how far the truth lies from the null hypothesis value.
- The farther the null hypothesis value is from the true value, the greater the power.
- Obtaining a larger sample size decreases the probability of a Type I error and Type II error, while the power of test will increase.