9.1 Hypothesis Testing

a. A **statistical hypothesis**, or simply a **hypothesis**, is an *assumption* about a population parameter.

b. **Hypothesis testing** is the procedure whereby we decide to “reject” or “fail to reject” a hypothesis.

c. **Null hypothesis** $H_0$: This is the hypothesis (assumption) under investigation or the statement being tested. The null hypothesis is a statement that “there is no effect,” “there is no difference,” or “there is no change.” The possible outcomes in testing a null hypothesis are ‘reject’ or ‘fail to reject.’

d. **Alternate hypothesis** $H_1$: This is a statement you will adopt if there is strong evidence (sample data) against the null hypothesis. A statistical test is designed to assess the strength of the evidence (data) against the null hypothesis.

e. **Fail to Reject** $H_0$: We never say we “accept $H_0$” - we can only say we “fail to reject” it. Failing to reject $H_0$ means there is NOT enough evidence in the data and in the test to justify rejecting $H_0$. So, we retain the $H_0$ knowing we have not proven it true beyond all doubt.

f. **Rejecting** $H_0$: This means there IS significant evidence in the data and in the test to justify rejecting $H_0$. When $H_0$ is rejected the data is said to be **statistically significant**. We adopt $H_1$ knowing we will occasionally be wrong.
Example 1
A car manufacturer advertises a car that gets 47 mpg. Let $\mu$ be the mean mileage for this model. You assume that the dealer will not underrate the mileage, but suspect he may overrate the mileage.

a. What can be used for $H_0$?

b. What can be used for $H_1$?

Guided Exercise 1A
A company that manufactures ball bearings claims the average diameter is 6 mm. To check that the average diameter is correct, the company decides to formulate a statistical test.

a. What can be used for $H_0$?

b. What can be used for $H_1$?

Guided Exercise 1B
A consumer group wants to test the truth in a package delivery company’s claim that it takes an average of 24 hours to deliver a package. Complaints have led the consumer group to suspect the delivery time is longer than 24 hours.

a. What can be used for $H_0$?

b. What can be used for $H_1$?
Types of Tests: Left-tailed, Right-Tailed, Two-Tailed

The null hypothesis generally states the parameter of interest equals a specific value; typically a historical value of a value of no change. For example, $H_0: \mu = k$. There are three types of statistical tests, which are determined by the alternate hypothesis as follows:

**Left-Tail Test**
- $H_0: \mu = k$
- $H_1: \mu < k$

**Right-Tail Test**
- $H_0: \mu = k$
- $H_1: \mu > k$

**Two-Tail Test**
- $H_0: \mu = k$
- $H_1: \mu \neq k$

**Level of Significance**
The level of significance $\alpha$ is the probability we are willing to risk rejecting $H_0$ when it is true; it is typically between 1% or 5%.

In the above pictures, think of $\alpha$ as the predetermined maximum area in the tail(s). Since $H_0: \mu = k$ is a statement of “no change,” and is assumed true, we reject $H_0$ only if we take a random sample and the sample mean $\bar{x}$ is so far away from the assumed mean ($H_0: \mu = k$) that it is statistically unlikely that the assumption $\mu = k$ can be true. That is, the area in the tail(s) must be less than or equal to the level of significance $\alpha$, to reject $H_0$. 
Example 2

Let $x$ be random variable that represents the heart rate in beats per minute of Rosie, and old sheep dog. From past experience the vet knows that $x$ is normally distributed with a mean of 115 bpm and standard deviation of $\sigma = 12$ bpm. Over the past several weeks Rosie’s heart rate (beats / min) was measured at

93 109 110 89 112 117

The sample mean is $\bar{x} = 105.0$. The vet is concerned that Rosie’s heart rate may be slowing. At a 5% level of significance, does the data indicate this is the case?

a. Establish the null hypothesis (i.e. nothing has changed) and the alternate hypothesis.

b. Draw the $\bar{x}$-distribution. Compute the probability of obtaining a sample mean of 105 bpm or less when the population mean is 115 bpm (by assumption). This area in the tail is called the $P$-value.

c. What can you conclude about Rosie’s heartbeat?
**P-value**

Assuming $H_0$ is true, the probability that the test statistic will take on values as extreme or more extreme than the observed test statistic is called the **P-value** of the test. The smaller the $P$-value computed from the sample data, the stronger the evidence against $H_0$. In the $\bar{x}$-distributions below, the $P$-value is the total area in the tail(s).

**Left-Tail Test**

$H_0$: $\mu = k$

$H_1$: $\mu < k$

**Right-Tail Test**

$H_0$: $\mu = k$

$H_1$: $\mu > k$

**Two-Tail Test**

$H_0$: $\mu = k$

$H_1$: $\mu \neq k$

Area = $P$-value

$\bar{x}$ $\mu = k$ $\mu > k$

Area = $\frac{P$-value}{2}$

$\bar{x}$ $\mu = k$ $\bar{x}$

**Type I and Type II Errors**

A **Type I error** occurs when we reject a true null hypothesis $H_0$. A **Type II error** occurs when we “fail to reject” a false null hypothesis $H_0$. For a given sample size reducing the probability of a type I error increases the probability of a type II error, and visa versa.

The probability of a type I error we are willing to accept in an application is called the **level of significance**, denoted $\alpha$ (alpha). Alpha is specified in advance.

$$\alpha = P(\text{making a type I error}) = P(\text{rejecting a true } H_0)$$

e.g. If $\alpha = 0.05$, then we say we are using a 5% level of significance. This means that in 100 similar situations $H_0$ will be rejected 5 times (on average) when it was true and should not have been.
Example 3
Reconsider Example 1 where
\[ H_0: \mu = 47 \text{ mpg} \quad \quad \quad H_1: \quad \mu < 47 \text{ mpg} \]

a. Suppose \( \alpha = 0.05 \). Describe a type I error and its probability.
A type I error is rejecting a true null hypothesis; in this case rejecting the dealer’s claim that \( \mu = 47 \text{ mpg} \) and concluding that \( \mu < 47 \text{ mpg} \) when in fact the average number of miles per gallon is 47 or higher. \( P(\text{type I error}) = 0.05 \).

b. Describe a type II error
A type II error is failing to reject a false null hypothesis. In this case we “fail to reject” the manufacturer’s claim that \( \mu = 47 \text{ mpg} \) when in fact \( \mu < 47 \text{ mpg} \).

Guided Exercise 2
Recall the ball-bearing example where \( H_0: \mu = 6 \text{ mm} \) and \( H_1: \mu \neq 6 \text{ mm} \). Suppose \( \alpha = 0.01 \).

a. Describe a type I error and its consequences and probability.
The probability of a type I error is 1%, the level of significance. A type I error would mean that we rejected the manufacturer’s claim the \( \mu = 6 \text{ mm} \) when in fact the average diameter was 6 mm. The consequence of a type I error in this application would be needless adjustment and delay in the manufacturing process.

b. Describe a type II error and its consequences
A type II error would mean that we “failed to reject” the manufacturer’s claim the \( \mu = 6 \text{ mm} \) when in fact \( \mu \neq 6 \text{ mm} \). The consequence of a type II error in this application would be the production of many bearings that do not meet specifications.
**Statistical Test Conclusions and Meanings**

For a given, preset level of significance $\alpha$, and a $P$-value computed from the sample data:

1. If $P$-value $\leq \alpha$, then $H_0$ is rejected. That is, there is enough evidence in the [sample] data to reject $H_0$. This means we chose the alternate hypothesis $H_1$ knowing we have not proven $H_1$ beyond all doubt.

2. If $P$-value $> \alpha$, then we fail to reject $H_0$. That is, there is not enough evidence in the [sample] data to reject $H_0$. This means we retain $H_0$ knowing we have not proven it beyond all doubt.

**Example 4**

A car manufacturer advertises a car that gets 47 mpg. Suppose that we sampled 40 cars and found a mean gas mileage of 46.26 mpg. The standard deviation is $\sigma = 2.22$ mpg. Test the manufacturer's claim at a 5% level of significance ($\alpha = 0.05$).

a. Establish the null and alternate hypotheses.

   $H_0$: $\mu = 47$ mpg $\quad H_1$: $\mu < 47$ mpg

b. Draw the normal $\bar{x}$-distribution and show the null hypothesis and sample statistic on the axis. Label the axis; include the units.

c. Compute the $p$-value.

   $p$-value = normalcdf(0, 46.26, 47, 2.22 / $\sqrt{40}$) = 0.0175
d. Conclude the test. Interpret its meaning in this application.

The $p$-value is 0.018. Since the $p-value = 0.018 \leq \alpha = 0.05$, we reject $H_0$, which means at a 5% level of significance the sample data is significant and supports that the mean car mileage is less than 47 mpg.

e. Repeat part d, but test the manufacturers claim at a 1% level of significance ($\alpha = 0.01$).

The $p$-value is 0.018. Since the $p-value = 0.018 > \alpha = 0.01$, we fail to reject $H_0$, which means at a 5% level of significance the sample data is not strong enough to say the mean car mileage is less than 47 mpg.
9.1 Homework
1. Do problems 1-8 all.

2. On problems 9-14 follow these steps:
   
   (a) Write the null and alternate hypotheses. Include units.

   (b) Compute the standard error \( \sigma_x = \sigma / \sqrt{n} \). Then sketch the normal curve and the area under the curve that represents the \( p \)-value. Label the axis to include the assumption in the null hypotheses and 3 standard deviation on both sides. Include units.

   (c) Compute the \( p \)-value (without using the ZTest function).

   (d) Conclude the test. That is, if the \( P \)-value \( \leq \alpha \), then reject \( H_0 \), otherwise do not reject \( H_0 \).

   (e) Summarize the results.

**Example**

\[ H_0 : \mu = 47 \text{ mpg} \]
\[ H_1 : \mu < 47 \text{ mpg} \]

\[
p-value = \text{area in the tail(s)}
= \text{normalcdf}(0, 46.26, 47, \frac{2.22}{\sqrt{40}})
= 0.0175
\]

\( p-value = 0.0175 < \alpha = 0.05 \)
Reject \( H_0 \)

At a 5% l.o.s. the sample data is significant and supports that the mean car mileage is less than 47 mpg.
9.2 Testing the Mean $\mu$

Example 3 Testing the Mean $\mu$ when $\sigma$ is Known

Some scientists believe sunspot activity is related to drought duration. Let $x$ by a random variable representing the number of sunspots observed in a four-week period. A random sample of 40 such periods in Spanish colonial times gave the following data:

| 12.5 | 14.1 | 37.6 | 48.3 | 67.3 | 70.0 | 43.8 | 56.5 | 59.7 | 24.0 |
| 12.0 | 27.4 | 53.5 | 73.9 | 104.0 | 54.6 | 4.4 | 177.3 | 70.1 | 54.0 |
| 28.0 | 13.0 | 6.5 | 134.7 | 114.0 | 72.7 | 81.2 | 24.1 | 20.4 | 13.3 |
| 9.4 | 25.7 | 47.8 | 50.0 | 45.3 | 61.0 | 39.0 | 12.0 | 7.2 | 11.3 |

The sample mean is $\bar{x} \approx 47.0$. Previous studies indicate that $\sigma = 35$. It is thought that for thousands of years, the mean number of sunspots per four-week period was about $\mu = 41$. Do the data indicate, at a 5% level of significance, that the sunspot activity during the Spanish colonial period was higher than 41?

a. Establish the hypotheses.

b. What does a 5% level of significance mean in this application?

We are willing to tolerate at most a 5% probability of rejecting a true null hypothesis. That is, assuming $H_0$: $\mu = 41$ is true, to reject $H_0$ means the probability that a sample $\bar{x}$ is as extreme or more extreme than our observed sample statistic ($\bar{x} \approx 47.0$) must be less than $\alpha = 0.05$.

c. Explain the meaning of the $P$-value in this application.

Assuming $H_0$: $\mu = 41$ is true, the $P$-value is the probability that a sample $\bar{x}$ is as extreme or more extreme than our observed sample statistic ($\bar{x} \approx 47.0$).
d. Draw the $\bar{x}$-distribution. Place the null hypothesis and the observed $\bar{x}$ on the axis. Then compute the $P$-value.

![Graph of a normal distribution]


e. Conclude the test. That is, if the $P$-value $\leq \alpha$, then reject $H_0$, otherwise do not reject $H_0$.

f. Interpret your results.
10.2 Exercises #1-16: Steps to Test the Mean $\mu$

1. Establish $H_0$ and $H_1$:

<table>
<thead>
<tr>
<th>Left-Tailed Test</th>
<th>Right-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$: $\mu = k$</td>
<td>$H_0$: $\mu = k$</td>
<td>$H_0$: $\mu = k$</td>
</tr>
<tr>
<td>$H_1$: $\mu &lt; k$</td>
<td>$H_1$: $\mu &gt; k$</td>
<td>$H_1$: $\mu \neq k$</td>
</tr>
</tbody>
</table>

2. Indicate which test you are using. The output for either test is the $P$-value.
   a. If $\sigma$ is known, then the convention is to compute the $P$-value with a normal distribution. The Z-Test uses a normal distribution (STAT / TESTS / 1: Z-Test).

   b. If $\sigma$ is NOT known, then the convention is to compute the $P$-value with the more conservative Student’s $t$-Distribution (STAT / TESTS / 2: T-Test).

3. Conclude the Test: If $P$-value $\leq \alpha$, then the sample data is **significant** and we reject $H_0$, otherwise we do not reject $H_0$.

4. State your conclusions in the context of the application.
Example 3
A zoo wishes to obtain eggs from a rare river turtle so they can be hatched and raised to preserve the species. Carol, a staff biologist, finds a nest of 36 eggs she suspects to be from the rare turtle species. Research has shown that the size of rare turtle eggs are normally distributed with a population mean of $\mu = 7.50$ cm. Furthermore, the mean length of the eggs of the other (common) turtle species is known to be longer than 7.50 cm. For the sample, the mean length of the 36 eggs is $\bar{x} = 7.74$ cm. The standard deviation of all turtle eggs is $\sigma = 1.5$ cm. So, Carol is concerned that the eggs may have come from a common turtle species. Do the data indicate that the eggs from the rare river turtle at a 5% level of significance.

1. Establish $H_0$ and $H_1$.
   
   $H_0$: $\mu = 7.50$ cm
   $H_1$: $\mu > 7.50$ cm

2. State the possible conclusions and their interpretations in this application.

<table>
<thead>
<tr>
<th>Test Conclusion</th>
<th>Interpretation of the Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail to reject $H_0$</td>
<td>At a 5% level of significance the sample data is not strong enough to reject $H_0$. That is, the sample evidence is not strong enough to say the eggs are from the common turtle.</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>At a 5% level of significance the sample data is statistically significant and is sufficient to reject $H_0$, which suggests the eggs are from the common turtle. We will be wrong at most $\alpha = 5%$ of the time.</td>
</tr>
</tbody>
</table>
3. Explain a 5% level of significance in this application. Explain how serious a type I error is in this application?

A 5% level of significance means we are taking a 5% risk of a type 1 error – a 5% risk of rejecting a true H₀. In this application we are only willing to take a 5% chance of rejecting that the eggs are from the rare river turtle.

5. Find the probability that our assumed mean in the null hypothesis (H₀: μ = 7.50 cm) is at or further away than the test statistic (x̄). That is, find the P-value.

6. Conclude the test.

7. Interpret the results.
Example 5
The drug 6-mP (6-mercaptopurine) is used to treat leukemia. The following data represent the remission times (in weeks) for a random sample of 21 patients using 6-mP.

<table>
<thead>
<tr>
<th>10</th>
<th>7</th>
<th>32</th>
<th>23</th>
<th>22</th>
<th>6</th>
<th>16</th>
<th>34</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>11</td>
<td>20</td>
<td>19</td>
<td>6</td>
<td>17</td>
<td>35</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sample mean is 17.1 weeks with a sample standard deviation of 10.0 weeks. Let $x$ be a random variable representing the remission times (in weeks) for all patients. Assume the $x$-distribution is mound-shaped and symmetric. A previous drug treatment had a remission time of 12.5 weeks. At a 1% level of significance do the data indicate the mean remission time for 6-mP is different (either way)?

1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test. Show your work and/or indicate the test used on your calculator to compute the $P$-value.

3. Interpret the results.
Example 6
Archeologists become excited when they find an anomaly in a newly discovered artifact. The anomaly may or may not indicate a new trading region or a new method of craftsmanship. Suppose the lengths of arrowheads at a certain site have a mean length of $\mu = 2.6$ cm. A random sample of 61 recently discovered arrowheads in an adjacent cliff dwelling had a sample mean length of 2.92 cm. The standard deviation is $\sigma = 0.85$ cm. Do these data indicate that the mean length of arrowheads in the adjacent cliff dwelling is longer than 2.6 cm? Use a 1% level of significance.

1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test.
   Show your work and/or indicate the test used on your calculator to compute the $P$-value.

3. Interpret the results.
Example 7
By taking thousands of practice shots at driving ranges, Pam knows her mean distance using a #1 wood is 225 yards with a standard deviation $\sigma = 25$ yards. Taking 100 shots with a new ball, Pam found her sample mean distance was 230 yards. At a 5% level of significance, determine if Pam improved her driving distance using the new ball?
1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test. Show your work and/or indicate the test used on your calculator to compute the $P$-value.

3. Interpret the results.
Example 8
A large company with offices around the world occasionally must move their employees from one city to another. From long experience, the company knows its employees move on average once every 8.50 years with a standard deviation of 3.62 years. Recent trends have led some to believe a change might have occurred. A sample of 48 employees were asked the number of years since the company last moved them. The mean time was 7.91 years. Has the mean time between moves significantly changed? Use $\alpha = 0.05$.

1. Establish the hypotheses.

2. Without using the ZTest, find the $P$-value of the test statistic and conclude the test. Show your work and/or indicate the test used on your calculator to compute the $P$-value.

3. Interpret the results.
Guided Exercise 5
Production records show that a machine that makes bottle caps makes caps with a mean diameter of 1.85 cm and a standard deviation of 0.05 cm. An inspector measured a random sample of 64 caps and found a mean diameter of 1.87 cm. At a 1% level of significance, determine if the machine slipped out of adjustment?

1. Establish the hypotheses.

2. Without using the ZTest, find the $P$-value of the test statistic and conclude the test. Show your work and/or indicate the test used on your calculator to compute the $P$-value.

3. Interpret the results.
Confidence Interval versus Two-tailed Hypothesis Test

Suppose a two-tailed hypothesis test has a level of significance $\alpha$ and null hypothesis $H_0: \mu = \mu_0$. Let $c$ be the confidence level for the mean $\mu$ based on the sample data. Then $c = 1 - \alpha$ and

1. $H_0$ is not rejected whenever $\mu_0$ falls inside the $c$ confidence interval for the mean $\mu$.
2. $H_0$ is rejected whenever $\mu_0$ falls outside the $c$ confidence interval for the mean $\mu$.

Exercise 19, Section 9.2

Consider a two-tailed hypothesis test with $\alpha = 0.01$ and

$$H_0: \mu = 20 \quad H_1: \mu \neq 20$$

A random sample of size 36 has a sample mean of 22. It is known the standard deviation $\sigma = 4$. Use $\alpha = 0.03$.

a. Use hypothesis testing to see if there is sufficient evidence to reject $H_0$.

b. Solve using a confidence interval.
   i. What is the confidence level corresponding to a level of significance of 0.03? Find the ____% confidence interval for the mean $\bar{x}$.

   We are ____% confident that the population mean $\mu$ is between _________ and _________.

   ii. Do we reject or fail to reject $H_0$ based on the 97% confidence interval.
9.3 Testing a Proportion \( p \)

Setup and Assumptions
1. Let \( r \) be the binomial random variable representing the number of successes out of \( n \) trials.
2. The sample size \( n \) is large so that it can be approximated by a normal distribution. That is, \( np > 5 \) and \( nq > 5 \).
3. For the probability of success use \( \hat{p} = r / n \) for the point estimate of the population parameter \( p \).
4. The possible sets of hypotheses are:

<table>
<thead>
<tr>
<th>Left-Tailed Test</th>
<th>Right-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: p = k )</td>
<td>( H_0: p = k )</td>
<td>( H_0: p = k )</td>
</tr>
<tr>
<td>( H_1: p &lt; k )</td>
<td>( H_1: p &gt; k )</td>
<td>( H_1: p \neq k )</td>
</tr>
</tbody>
</table>

5. **TI-84: STAT / TESTS / 1-PropZTest**

   Input: 
   - \( p_0 \): from the \( H_0 \)
   - \( x \): the number of successes (the \( r \)-value)
   - \( n \): number of trials
   - \( < p_0, > p_0, \neq p_0 \) depending on \( H_1 \)

   Output: the \( P \)-value

6. Conclude the Test: If \( P \)-value \( \leq \alpha \), then the sample data is significant and we reject \( H_0 \), otherwise we conclude the sample data is not strong enough reject \( H_0 \).

7. Summarize your conclusion in the specific situation.
Example 9
A team of eye surgeons has developed a new technique for a risky eye operation to restore the sight of people blinded from a certain disease. Under the old method, only 30% of the patients recovered their eyesight. Surgeons have performed the new technique 225 times and 88 of those patients have recovered their sight. Can we justify the claim that the new technique is better than the old one at a 1% level of significance?
1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test.
   Show the test used on your calculator to compute the $P$-value.

3. Interpret the results.
Example 10
A botanist has produced a new variety of hybrid wheat that is better able to withstand drought than other varieties. He knows that 80% of the seeds from the parent plants germinate. He claims the hybrid has the same germination rate. To test this claim, 400 seeds from the hybrid plant are tested and 312 germinated. Test the botanist claim at a 5% level of significance.

1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test. Show the test used on your calculator to compute the $P$-value.

3. Interpret the results.
9.4 Tests Involving Paired Differences
(Dependent Samples)

Dependent Samples
Dependant samples have data that are naturally paired.
Dependent samples occur naturally in many applications, such as “before and after” situations – where the same object is measured before and after a treatment. In such cases the difference in the two measures is tested.

Examples of Dependent Samples
a. A shoe manufacturer claims that among adults in the United States, the left foot is longer than the right foot.

b. A weekend refresher math course is administered to new students. An exam is administered to each student before and after the course.
Testing the Difference, \( d \), of Paired Data

a. It is assumed the paired data are such that the difference \( d \) between the first and second members of each pair are approximately normally distributed with a population mean \( \mu_d \).

b. A random sample of \( n \) data pairs with sample mean \( \overline{d} \) and sample standard deviation \( s_d \) follow a Student’s \( t \) distribution and can be tested with STAT / TESTS / 2: T-Test.

c. The possible sets of hypotheses to be tested are:

<table>
<thead>
<tr>
<th>Left-Tailed Test</th>
<th>Right-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: \mu_d = 0 )</td>
<td>( H_0: \mu_d = 0 )</td>
<td>( H_0: \mu_d = 0 )</td>
</tr>
<tr>
<td>( H_1: \mu_d &lt; 0 )</td>
<td>( H_1: \mu_d &gt; 0 )</td>
<td>( H_1: \mu_d \neq 0 )</td>
</tr>
</tbody>
</table>

4. TI-83: STAT / TESTS / 2: T-Test
   Input: \( \mu_0 \): from the \( H_0 \)
   \( \overline{x} \): the mean of the differences \( \overline{d} \)
   \( s_x \): standard deviation of \( \overline{d} \), \( s_d \)
   \( n \): number of pairs in the sample
   \( \mu \): \( < \mu_0 \), \( > \mu_0 \), \( \neq \mu_0 \) depending on \( H_1 \)
   Output: the \( P \)-value

5. Conclude the Test: If \( P \)-value \( \leq \alpha \), then the sample data is significant and we reject \( H_0 \), otherwise we conclude the sample data is not strong enough reject \( H_0 \).

6. Interpret the results (specific to application).
Example 10
Heart surgeons know that many patients who undergo heart surgery have a dangerous buildup of anxiety before the operation. Psychiatric counseling may relieve some of that anxiety. The data shown are the anxiety scores of patients before and after counseling. Higher scores mean higher levels of anxiety. Can we conclude that counseling reduces anxiety? Use \( \alpha = 0.01 \).

1. Establish the hypotheses.

<table>
<thead>
<tr>
<th>Patient</th>
<th>( B ) Score before counseling</th>
<th>( A ) Score after counseling</th>
<th>( d = A - B ) Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>121</td>
<td>76</td>
<td>-45</td>
</tr>
<tr>
<td>B</td>
<td>93</td>
<td>93</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>105</td>
<td>64</td>
<td>-41</td>
</tr>
<tr>
<td>D</td>
<td>115</td>
<td>117</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>130</td>
<td>82</td>
<td>-48</td>
</tr>
<tr>
<td>F</td>
<td>98</td>
<td>80</td>
<td>-18</td>
</tr>
<tr>
<td>G</td>
<td>142</td>
<td>79</td>
<td>-63</td>
</tr>
<tr>
<td>H</td>
<td>118</td>
<td>67</td>
<td>-51</td>
</tr>
<tr>
<td>I</td>
<td>125</td>
<td>89</td>
<td>-36</td>
</tr>
</tbody>
</table>

2. Find the \( P \)-value of the test statistic and conclude the test. Show the test used on your calculator to compute the \( P \)-value.

3. Interpret the results [specific to the context of the application].
Example 11
To test the quality of two brands of tires, one tire of each brand was placed on six test cars. After 6 months the amount of wear on each tire was measured in thousandths of inches.

Can we conclude the two tire brands show unequal wear at a 2% level of significance?

1. Establish the hypotheses.

<table>
<thead>
<tr>
<th>Car</th>
<th>Soapstone</th>
<th>Bigyear</th>
<th>Difference d = S - B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>132</td>
<td>140</td>
<td>-8</td>
</tr>
<tr>
<td>2</td>
<td>71</td>
<td>74</td>
<td>-3</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>110</td>
<td>-20</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>93</td>
<td>105</td>
<td>-12</td>
</tr>
<tr>
<td>6</td>
<td>107</td>
<td>119</td>
<td>-12</td>
</tr>
</tbody>
</table>

2. Find the P-value of the test statistic and conclude the test.
Show the test used on your calculator to compute the P-value.

3. Interpret the results [specific to the context of the application].
9.5 Testing $\mu_1 - \mu_2$ and $p_1 - p_2$ (Independent Samples)

Samples are independent if there is no relationship whatsoever between specific values of the two distributions.

Example 12
A teacher wishes to compare the effectiveness of two teaching methods. Students are randomly divided into two groups: The first group is taught by method 1 and the second group by method 2. At the end of the course, a comprehensive exam is given to all students. The mean scores, $\bar{x}_1$ and $\bar{x}_2$, of the two groups are compared. Are the samples independent or dependent?

Example 13
A shoe manufacturer claims that for U.S. adults the average length of the left foot is longer than the average length of the right foot. A random sample of 60 adults is drawn and the length of both their left and right feet are measured and averaged as $\bar{x}_1$ and $\bar{x}_2$, respectively. Are the samples independent or dependent?

Theorem 9.2
Let $x_1$ have a normal distribution with mean $\mu_1$ and standard deviation $\sigma_1$. Let $x_2$ have a normal distribution with mean $\mu_2$ and standard deviation $\sigma_2$. Suppose random sample of size $n_1$ and $n_2$ are taken from the respective distributions. Then the variable $\bar{x}_1 - \bar{x}_2$ has
1. A normal distribution.
2. Mean $\mu_1 - \mu_2$
3. Standard deviation $\sqrt{\sigma_1^2/n_1 - \sigma_2^2/n_2}$
Steps for Section 9.5 Problems

1. Establish $H_0$ and $H_1$.

<table>
<thead>
<tr>
<th>Left-Tailed Test</th>
<th>Right-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$: $\mu_1 = \mu_2$</td>
<td>$H_0$: $\mu_1 = \mu_2$</td>
<td>$H_0$: $\mu_1 = \mu_2$</td>
</tr>
<tr>
<td>$H_1$: $\mu_1 &lt; \mu_2$</td>
<td>$H_1$: $\mu_1 &gt; \mu_2$</td>
<td>$H_1$: $\mu_1 \neq \mu_2$</td>
</tr>
</tbody>
</table>

2. Indicate which test you are using.
   
   a. **If $\sigma_1$ and $\sigma_2$ are known**, then the convention is to compute the $P$-value with a normal distribution. The 2-SampZTest uses a normal distribution (STAT / TESTS / 3: 2-SampZTest).

   b. **If $\sigma_1$ and $\sigma_2$ are not known**, then the convention is to compute the $P$-value with the more conservative Student’s $t$-Distribution (STAT / TESTS / 4: 2-SampTTest). Input the sample standard deviation $s$.

3. Conclude the Test: If $P$-value $\leq \alpha$, then the sample **data is significant** and we reject $H_0$, otherwise we conclude the sample data is not strong enough to reject $H_0$.

4. Interpret the results [specific to the context of the application].
Example 14
A consumer group measures the heating capacity of camp stoves by measuring the time it takes the stove to boil 2 quarts of water from 50°F. Two competing models were tested:

Model 1: \( \bar{x}_1 = 11.4 \text{ min} \quad \sigma_1 = 2.5 \text{ min} \quad n_1 = 10 \)

Model 2: \( \bar{x}_2 = 9.9 \text{ min} \quad \sigma_2 = 2.5 \text{ min} \quad n_2 = 12 \)

Is there a difference in the performance of the two models at a 5% level of significance?

1. Establish the hypotheses.

2. Find the \( P \)-value of the test statistic and conclude the test.
   Show the test used on your calculator to compute the \( P \)-value.

3. Interpret the results [specific to the context of the application].
**Example 15**

Two competing headache remedies claim to give fast-acting relief. An experiment was performed to compare the mean lengths of time required for bodily adsorption of brand A and brand B:

- **Brand A:** \( \bar{x}_1 = 21.8 \text{ min} \quad s_1 = 8.7 \text{ min} \quad n_1 = 12 \\
- **Brand B:** \( \bar{x}_2 = 18.9 \text{ min} \quad s_2 = 7.5 \text{ min} \quad n_2 = 12 \)

Assuming both distributions are approximately normal, test the claim that there is no difference in the mean time required for bodily absorption.

1. Establish the hypotheses.

2. Find the *P*-value of the test statistic and conclude the test.
   Show the test used on your calculator to compute the *P*-value.

3. Interpret the results [specific to the context of the application].
Testing Two Proportions $p_1$ & $p_2$

STAT / TESTS / 6: 2-PropZTest

Example 16
The Macek County Clerk wishes to improve voter registration. One method under consideration is to send reminders in the mail to all citizens in the county who are eligible to register. A random sample of 1250 potential register voters was taken.

Group 1: There were 625 people in this group. No reminders to register were sent to them. The number of potential voters from this group who registered was 295.

Group 2: There were 625 people in this group. Reminders to register were sent to them. The number of potential voters from this group who registered was 350.

At a 5% level of significance, did reminders improve voter registration?

1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test. Show the test used on your calculator to compute the $P$-value.

3. Interpret the results [specific to the context of the application].
Guided Exercise 11
The Macek County Clerk wishes to improve voter registration. One method under consideration is to send reminders in the mail to all citizens in the county who are eligible to register. A random sample of 1100 potential register voters was taken.

Group 1: There were 500 people in this group. No reminders to register were sent to them. The number of potential voters from this group who registered was 248.

Group 2: There were 600 people in this group. Reminders to register were sent to them. The number of potential voters from this group who registered was 332.

At a 1% level of significance, did reminders improve voter registration?

1. Establish the hypotheses.

2. Find the $P$-value of the test statistic and conclude the test.
   Show the test used on your calculator to compute the $P$-value.

3. Interpret the results [specific to the context of the application].
<table>
<thead>
<tr>
<th><strong>TI-83/84</strong> STAT / TESTS menu</th>
<th><strong>Section</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Z-Test</td>
<td>9.2</td>
<td>Testing the mean $\mu$ when $\sigma$ is known. Be able to do these problems without using the Z-Test function. That is, sketch the distribution and compute the $p$-value using the normalcdf function.</td>
</tr>
<tr>
<td>2: T-Test</td>
<td>9.2, 9.4</td>
<td>Testing the mean $\mu$ when $\sigma$ is not known, or testing dependent paired data $\mu_d = 0$.</td>
</tr>
<tr>
<td>3: 2-SampZTest</td>
<td>9.5</td>
<td>Testing two mean $\mu_1 - \mu_2$ when $\sigma_1$ and $\sigma_2$ are known.</td>
</tr>
<tr>
<td>4: 2-SampTTest</td>
<td>9.5</td>
<td>Testing two mean $\mu_1 - \mu_2$ when $\sigma_1$ and $\sigma_2$ are not known.</td>
</tr>
<tr>
<td>5: 1-PropZTest</td>
<td>9.3</td>
<td>Testing a proportion $p$.</td>
</tr>
<tr>
<td>6: 2-PropZTest</td>
<td>9.5</td>
<td>Testing two proportions.</td>
</tr>
<tr>
<td>7: ZInterval</td>
<td>8.1</td>
<td>Estimating $\mu$ when $\sigma$ is known. Be able to do these problems without using the ZInterval function. That is, sketch the distribution and compute the interval using the invNorm function.</td>
</tr>
<tr>
<td>8: TInterval</td>
<td>8.2</td>
<td>Estimating $\mu$ when $\sigma$ is not known.</td>
</tr>
<tr>
<td>9: 2-SampZInt</td>
<td>8.5</td>
<td>Estimating $\mu_1 - \mu_2$ when $\sigma_1$ and $\sigma_2$ are known.</td>
</tr>
<tr>
<td>0: 2-SampTInt</td>
<td>8.5</td>
<td>Estimating $\mu_1 - \mu_2$ when $\sigma_1$ and $\sigma_2$ are known.</td>
</tr>
<tr>
<td>B: 2-PropZInt</td>
<td>8.5</td>
<td>Estimating $p_1 - p_2$</td>
</tr>
</tbody>
</table>