Consider a measurement $X$ on two distinct populations: $\Omega_1$ where $X$ has mean $\mu_1$ and variance $\sigma_1^2$, and $\Omega_2$ where $X$ has mean $\mu_2$ and variance $\sigma_2^2$. We wish to test hypotheses about the difference in means $\mu_1 - \mu_2$:

1. $H_0: \mu_1 - \mu_2 = M$, with a one-sided alternative $H_a: \mu_1 - \mu_2 < M$.
2. $H_0: \mu_1 - \mu_2 = M$, with a one-sided alternative $H_a: \mu_1 - \mu_2 > M$.
3. $H_0: \mu_1 - \mu_2 = M$, with a two-sided alternative $H_a: \mu_1 - \mu_2 \neq M$.

The decision to reject a null hypothesis is based on the difference in sample means $\bar{x}_1 - \bar{x}_2$ from independent random samples, of sizes $n_1$ and $n_2$ respectively, conducted on the populations.

**Arbitrary Measurements**

When there are no assumptions on the measurements, we define the $z$ test statistic by

$$z = \frac{\bar{x}_1 - \bar{x}_2 - M}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

or

$$\frac{\bar{x}_1 - \bar{x}_2 - M}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

for large samples when $\sigma_1^2$ and $\sigma_2^2$ are unknown.

The test statistic follows an approximate standard normal distribution $Z \sim N(0, 1)$. Therefore, we can compute the left or right tail probability values created by the test statistic, $P(Z \leq z)$ and $P(Z \geq z)$, which give the $P$-values for the one-sided alternatives.

**Example 1.** A consumer report states that the salary gap is shrinking between full-time workers with a college degree and those without a college degree. The report states that in many areas the average gap is only $3000.

A local economist conducts a random survey of 250 full-time workers with a college degree and obtains a sample average income of $\bar{x}_1 = 48,560.75$ with a sample deviation of $4350.50$. An independent random survey of 700 workers without a college degree yields an average income of $\bar{x}_2 = 44,920.20$ with a sample deviation of $3870.64$.

Can the economist refute the report’s claim for his locale?
Solution. Note that $\bar{x}_1 - \bar{x}_2 = 3640.55 > 3000$. Let $\mu_1$ denote the average salary among all college-degreed full-time workers and let $\mu_2$ denote the average salary among all full-time workers without a degree. We shall test the null hypotheses

$$H_0: \mu_1 - \mu_2 = 3000 \text{ versus } H_a: \mu_1 - \mu_2 > 3000.$$  

The test statistic is

$$z = \frac{(48,560.75 - 44,920.20) - 3000}{\sqrt{\frac{4350.50^2}{250} + \frac{3870.64^2}{700}}} \approx 2.0555.$$  

Then $P(Z \geq 2.0555) \approx 0.0199$. This low $p$-value gives statistical evidence to reject $H_0$ in favor of the alternative. If $\mu_1 - \mu_2 = 3000$ were true, then there would be less than a 2% chance of obtaining an $\bar{x}_1 - \bar{x}_2$ of $3640.55$ or larger with samples of these sizes.

Alternate Solution. For this one-sided alternative, we also could use a pre-assigned level of significance, say $\alpha = 0.05$. Then for the right-tail probability to be 0.05 on the standard normal curve, the required $z$-score is 1.645. Because our $z$-statistic 2.0555 is beyond 1.645, then $\bar{x}_1 - \bar{x}_2 = 3640.55$ is too big. So we can reject $\mu_1 - \mu_2 = 3000$.

Calculator Command

The 2-SampZTest screen from the STAT TESTS menu can be used to compute this type of hypothesis test. However this command is designed to test the hypothesis that the difference of means is 0; i.e., that the means are equal:

$$H_0: \mu_1 - \mu_2 = 0 \text{ which is equivalent to } H_0: \mu_1 = \mu_2.$$  

The alternatives then are $\mu_1 < \mu_2$, $\mu_1 > \mu_2$, or $\mu_1 \neq \mu_2$, which are equivalent to $\mu_1 - \mu_2 < 0$, $\mu_1 - \mu_2 > 0$, and $\mu_1 - \mu_2 \neq 0$.

But our desired test is $H_0: \mu_1 - \mu_2 = M$ which is equivalent to $H_0: \mu_1 = \mu_2 + M$. To make the adjustment on the calculator and create the correct test statistic, simply add the value $M$ to the value of $\bar{x}_2$ as demonstrated below.
**Example 2.** The Survey of Study Habits and Attitudes was given to first-year students at a private college. The tables below show a random sample of the scores.

<table>
<thead>
<tr>
<th>Women’s scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>154 109 137 115 152 140 154 178 101</td>
</tr>
<tr>
<td>103 126 126 137 165 165 129 200 148</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Men’s scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>108 140 114 91 180 115 126 92 169 146</td>
</tr>
<tr>
<td>109 132 75 88 113 151 70 115 187 104</td>
</tr>
</tbody>
</table>

From years of previous tests, the standard deviation among women’s scores is assumed to be 25 and the standard deviation among men’s scores is assumed to be 32.

(a) Test the supposition that the mean scores are equal among all women and all men that are first-year students at this college.

(b) Test the supposition that the mean score for all women is 20 points higher than the mean score for all men.

**Solution.** In the **STAT Edit** screen, enter the women’s scores into list L1 and the men’s scores into list L2. Then compute **1-VarStats L1** and **1-VarStats L2**.

We see that $\bar{x}_1 \approx 141.056$ and $\bar{x}_2 = 121.25$, and then $\bar{x}_1 - \bar{x}_2 \approx 141.056 - 121.25 \approx 19.806$.

(a) Now let $\mu_1$ be the mean score among all first-year women and let $\mu_2$ be the mean score among all first-year men. We shall test the hypothesis $H_0: \mu_1 = \mu_2$ versus the alternative $H_a: \mu_1 > \mu_2$ because $\bar{x}_1 > \bar{x}_2$.

Call up the **2-SampZTest** feature, set the **Inpt** to **Data**, enter the specified standard deviations and the desired lists L1 and L2, set the alternative to $> \mu_2$, and calculate:

We obtain a $z$-statistic of 2.1366 and a $P$-value of 0.0163. If $\mu_1 = \mu_2$ were true, then there would be only a 1.63% chance of $\bar{x}_1 - \bar{x}_2$ being 19.806 or higher with samples of these sizes. This low $P$-value gives us evidence to reject $H_0$ and conclude that $\mu_1 > \mu_2$. That is, the mean score for all women is higher than the mean score for all men among first-year students at this college.
(b) Again note that \( \bar{x}_1 - \bar{x}_2 = 141.056 - 121.25 = 19.806 \). So now we shall test the hypothesis \( H_0: \mu_1 - \mu_2 = 20 \) with a one-sided alternative \( H_a: \mu_1 - \mu_2 < 20 \).

Call up the **2-SampZTest** feature, but now set the **Inpt** to **Stats**, in order to make the adjustment of adding 20 to \( \bar{x}_2 \). Set the alternative to \( < \mu_2 \), and calculate:

\[
\begin{array}{c|c}
\text{2-SampZTest} & \text{2-SampZTest} \\
\hline
\sigma_1: 2.5 & \mu_1 < \mu_2 \\
\sigma_2: 2.0 & z = 0.029769715 \\
\bar{x}_1: 141.055555... & p = 0.4963137085 \\
n_1: 18 & \bar{x}_2: 141.25 \\
x_2: 121.25 + 20 & n_2: 18 \\
\end{array}
\]

We obtain a \( z \)-statistic of about \(-0.021\) and a \( P \)-value of 0.49. If \( \mu_1 - \mu_2 = 20 \) were true, then there would be about a 49% chance of \( \bar{x}_1 - \bar{x}_2 \) being 19.806 or lower with samples of these sizes. The extremely high \( P \)-value means we cannot reject \( H_0 \).

**Normally Distributed Measurements**

As with confidence intervals, when the population measurements are known to be *normally distributed*, then we can be more accurate in the analysis and computations.

**I. Assume Both Normal Measurements Have the Same Standard Deviation**

In this case, the common deviation \( \sigma \) is approximated by the “pooled deviation”

\[
S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}},
\]

where \( S_1 \) and \( S_2 \) are the respective sample deviations. We then define the test statistic by

\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - M}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},
\]

which follows an exact \( t \)-distribution with \( n_1 + n_2 - 2 \) degrees of freedom. The \( P \)-values are then computed with this \( t \)-distribution, rather than with a standard normal distribution.

**II. The Normal Measurements Have Different Standard Deviations**

When we do not pool the deviation, then we define the test statistic as with arbitrary populations. However in this case with normally distributed measurements, the test statistic \( t = \frac{(\bar{x}_1 - \bar{x}_2) - M}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \) follows an approximate \( t \)-distribution with \( k \) degrees of freedom, where \( k \) is largest integer satisfying
As above, we use this $t$-distribution to compute the tail values.

**Calculator Command**

The 2-SampTTest screen will compute the test statistic and $P$-value (for either the pooled or non-pooled deviation) for the hypothesis $H_0: \mu_1 = \mu_2$ (i.e., $\mu_1 - \mu_2 = 0$). As before, to test $H_0: \mu_1 - \mu_2 = M$, add the value $M$ to the value of $\bar{x}_2$ in the calculation screen.

**Example 3.** Havoline is testing the effectiveness of a fuel additive by comparing the mpg of a various similar models of automobiles with and without the additive. Under everyday use, a control group of 20 midsize cars without the additive averaged 24.2 mpg with a sample deviation of 1.3. Under similar use, a test group of 18 midsize cars with the additive averaged 25.4 mpg with a sample deviation of 1.2.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Sample Mean</th>
<th>Sample Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>With additive</td>
<td>18</td>
<td>25.4</td>
</tr>
<tr>
<td>Without</td>
<td>20</td>
<td>24.2</td>
</tr>
</tbody>
</table>

Assume that the actual mpg of all such midsize models is normally distributed.

(a) Is there evidence to state that “on average” the additive increases the auto’s mpg?

(b) Test whether the additive increases the average mpg by more than 1 mpg.

**Solution.** Let $\mu_1$ be the average mpg of all similar midsize cars that use the additive and let $\mu_2$ be the average mpg of all midsizes that do not. Note that $\bar{x}_1 - \bar{x}_2 = 1.2$.

(a) We shall test $H_0: \mu_1 - \mu_2 = 0$ with a one-sided alternative $H_a: \mu_1 - \mu_2 > 0$. Due to the closeness of the sample deviations, we shall assume that deviations in mpg of all models under consideration are the same with or without the additive. Thus we shall pool the deviation.

![Calculator Screen Images]

Note that the pooled deviation is also displayed as $S_{xp} = 1.253772$.

The test statistic is $t = 2.946$ which gives a low $P$-value of 0.0028. If $\mu_1 - \mu_2 = 0$ were true, then there would be almost no chance of obtaining $\bar{x}_1 - \bar{x}_2$ of 1.2 or higher with samples of these sizes. We can reject the claim that $\mu_1 = \mu_2$. 

(b) Now test the claim \( H_0: \mu_1 - \mu_2 = 1 \) with the alternative \( H_a: \mu_1 - \mu_2 > 1 \).

(Add 1 to \( \bar{x}_2 \).)

The test stat is now \( t \approx 0.491 \) and the \( P \)-value is about 0.3132. If \( \mu_1 - \mu_2 = 1 \) were true, then there would still be a 31.32% chance of obtaining \( \bar{x}_1 - \bar{x}_2 \) of 1.2 or higher with samples of these sizes. This high \( P \)-value means we do not have evidence to state conclusively that the additive increases the average mpg by more than 1 mpg. Thus, we can accept \( H_0 \).

**Example 4.** Verbal SAT scores are usually found to be normally distributed with girls having more variance in score than boys. But a school district claims that “on average” boys score about 15 points less than girls.

The ETS provides random scores of boys and girls from one large district resulting in the following data:

<table>
<thead>
<tr>
<th>Group</th>
<th>( n )</th>
<th>( \bar{x} )</th>
<th>( S )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>252</td>
<td>507.14</td>
<td>90.28</td>
</tr>
<tr>
<td>Boys</td>
<td>208</td>
<td>495.48</td>
<td>65.68</td>
</tr>
</tbody>
</table>

Use the district as a sample to test the hypothesis that girls average 15 points higher than boys versus a suitable alternative.

**Solution.** Let \( \mu_1 \) be the girls’ average throughout the district and let \( \mu_2 \) be the boys’ average. Then we must test the hypothesis \( H_0: \mu_1 - \mu_2 = 15 \) (girls average 15 points higher). Note that \( \bar{x}_1 - \bar{x}_2 = 507.14 - 495.48 = 11.66 < 15 \).

We now test \( H_0: \mu_1 - \mu_2 = 15 \) with the alternative \( H_a: \mu_1 - \mu_2 < 15 \) using a non-pooled deviation.

The \( P \)-value for the one-sided test is 0.3247 which does not provide enough evidence to reject \( H_0 \). If \( \mu_1 - \mu_2 = 15 \) were true, then there would still be an 32.47% chance of \( \bar{x}_1 - \bar{x}_2 \) being 11.66 or lower with samples of these sizes. Cannot reject \( H_0 \).
Exercises

1. A school district is concerned about the gender gap in high school GPA among college bound students. A random sample of 360 college bound females yields a sample mean GPA of 3.363 with a sample deviation of 0.347. An independent sample of 300 college bound males yields an average GPA of 3.106 with a sample deviation of 0.672.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>$\bar{x}$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>360</td>
<td>3.363</td>
<td>0.347</td>
</tr>
<tr>
<td>Males</td>
<td>300</td>
<td>3.106</td>
<td>0.672</td>
</tr>
</tbody>
</table>

Let $\mu_1$ be the average GPA among all college-bound females in the district and let $\mu_2$ be the average GPA among all college-bound males in the district.

(a) Use a one-sided test to see if there is statistical evidence to say that the female average GPA is higher than the male average. State the null hypothesis, alternative, and explain the conclusion.

(b) Is there sufficient evidence to reject a claim that $\mu_1 - \mu_2 = 0.3$? Explain.

2. The hemoglobin levels of babies are found to be normally distributed regardless of the type of feeding. A pediatrician wants to see if there is any apparent difference in the average hemoglobin levels for breast-fed babies versus formula-fed babies. Below is random data acquired by the doctor.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>$\bar{x}$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast-fed</td>
<td>23</td>
<td>13.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Formula</td>
<td>19</td>
<td>12.4</td>
<td>1.8</td>
</tr>
</tbody>
</table>

(a) Test to see if there is statistical evidence to say that the average hemoglobin levels are different.

(b) Test to see if there is statistical evidence to say that the difference in averages is lower than 2.

3. The following table gives data on the birth weights (in grams) for a sample babies whose mothers had tested positive for cocaine vs. those whose mothers were drug-free. Birth weights from both groups are found to be normally distributed.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>$\bar{x}$</th>
<th>$S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug-free</td>
<td>200</td>
<td>3118</td>
<td>472</td>
</tr>
<tr>
<td>Positive test</td>
<td>134</td>
<td>2733</td>
<td>399</td>
</tr>
</tbody>
</table>

Let $\mu_1$ be the average weight among all babies born to drug-free mothers and let $\mu_2$ be the average weight among all babies born to “positive-test” mothers. We wish to test the claim that $\mu_1$ is only 300 grams higher than $\mu_2$. State an appropriate null hypothesis and alternative hypothesis, then test the hypothesis and explain the conclusion.
Solutions

1. (a) Test $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 > \mu_2$ (2–SampZTest). Note that $\bar{x}_1 - \bar{x}_2 = 0.257$.

The $P$-value is about $1.04 \times 10^{-9}$. If $\mu_1 - \mu_2 = 0$ were true, then there would be no chance of obtaining $\bar{x}_1 - \bar{x}_2$ of 0.257 or higher with these sample sizes. Can reject $H_0$.

(b) Now test $H_0: \mu_1 - \mu_2 = 0.3$ versus $H_a: \mu_1 - \mu_2 < 0.3$. Again note that $\bar{x}_1 - \bar{x}_2 = 0.257$. (Add 0.3 to $\bar{x}_2$ in the 2–SampZTest screen.)

The $P$-value is now 0.158. If $\mu_1 - \mu_2 = 0.3$ were true, then there would be a 15.8% chance of obtaining $\bar{x}_1 - \bar{x}_2$ of 0.257 or lower with these sample sizes. Cannot reject $H_0$.

2. (a) Test $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 > \mu_2$. Due to closeness of the sample deviations, assume the true standard deviations are equal. Also note that $\bar{x}_1 - \bar{x}_2 = 0.9$.

The $P$-value is about 0.052. If $\mu_1 = \mu_2$ were true, then there would be about a 5.2% chance of $\bar{x}_1 - \bar{x}_2$ being 0.9 or higher with samples of these sizes. This may be enough evidence to reject $H_0$.

(b) Now test $H_0: \mu_1 - \mu_2 = 2$, with a one-sided alternative $H_a: \mu_1 - \mu_2 < 2$.

The $P$-value is now 0.024. If $\mu_1 - \mu_2 = 2$ were true, then there would be only a 2.4% chance of obtaining $\bar{x}_1 - \bar{x}_2$ of 0.9 or lower with samples of these sizes. Can reject $H_0$.

3. Note that $\bar{x}_1 - \bar{x}_2 = 385$. We shall test $H_0: \mu_1 - \mu_2 = 300$, with a one-sided alternative $H_a: \mu_1 - \mu_2 > 300$ using a non-pooled 2–SampTest.

The $P$-value is 0.0387. If $\mu_1 - \mu_2 = 300$ were true, then there would be only a 3.87% chance of obtaining $\bar{x}_1 - \bar{x}_2$ of 385 or higher with these sample sizes. Reject $H_0$. 