Inference for Distributions
Comparing Two Means

Objectives (IPS Chapter 7.2)

Comparing two means

- Two-sample z statistic
- Two-samples t procedures
- Two-sample t significance test
- Two-sample t confidence interval
- Robustness
- Details of the two-sample t procedures
Comparing two samples

Which is it?

We often compare two treatments used on independent samples.

Is the difference between both treatments due only to variations from the random sampling (B), or does it reflect a true difference in population means (A)?

Independent samples: Subjects in one sample are completely unrelated to subjects in the other sample.

Two-Sample Procedures

- We have two independent samples, from two distinct populations (such as subjects given a treatment and those given a placebo).
- We want to compare the two population means, either by giving a confidence interval for \( \mu_1 - \mu_2 \) or by testing the hypothesis of no difference, \( H_0: \mu_1 = \mu_2 \).

- Distribution of the difference between two sample means –

\[
X_1 - X_2 \sim N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})
\]

Values of \( X_1 - X_2 \)
Two-sample \( z \) statistic

We have **two independent SRSs** (simple random samples) possibly coming from two distinct populations with \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\). We use \(\bar{x}_1\) and \(\bar{x}_2\) to estimate the unknown \(\mu_1\) and \(\mu_2\).

When both populations are normal, the sampling distribution of \((\bar{x}_1 - \bar{x}_2)\) is also normal, with standard deviation:

\[
\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
\]

Then the **two-sample \( z \) statistic** has the standard normal \(N(0, 1)\) sampling distribution.

\[
z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]

Two independent samples \( t \) distribution

We have **two independent SRSs** (simple random samples) possibly coming from two distinct populations with \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\) unknown. We use \((\bar{x}_1, s_1)\) and \((\bar{x}_2, s_2)\) to estimate \((\mu_1, \sigma_1)\) and \((\mu_2, \sigma_2)\), respectively.

To compare the means, both populations should be normally distributed. However, in practice, it is enough that the two distributions have similar shapes and that the sample data contain no strong outliers.
The two-sample \( t \) statistic follows approximately the \( t \) distribution with a standard error \( SE \) (spread) reflecting variation from both samples:

\[
SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
\]

Conservatively, the degrees of freedom is equal to the smallest of \((n_1 - 1, n_2 - 1)\).

**Two-sample \( t \) significance test**

The null hypothesis is that both population means \( \mu_1 \) and \( \mu_2 \) are equal, thus their difference is equal to zero.

\[
H_0: \mu_1 = \mu_2 \iff \mu_1 - \mu_2 = 0
\]

with either a one-sided or a two-sided alternative hypothesis.

We find how many standard errors (SE) away from \((\mu_1 - \mu_2)\) is \((\bar{x}_1 - \bar{x}_2)\) by standardizing with \( t \):

\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE}
\]

Because in a two-sample test \( H_0 \) poses \((\mu_1 - \mu_2) = 0\), we simply use

\[
t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

With \( df = \text{smallest}(n_1 - 1, n_2 - 1) \)
Does smoking damage the lungs of children exposed to parental smoking?

Forced vital capacity (FVC) is the volume (in milliliters) of air that an individual can exhale in 6 seconds.

FVC was obtained for a sample of children not exposed to parental smoking and a group of children exposed to parental smoking.

<table>
<thead>
<tr>
<th>Parental smoking</th>
<th>FVC $\bar{X}$</th>
<th>s</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>75.5</td>
<td>9.3</td>
<td>30</td>
</tr>
<tr>
<td>No</td>
<td>88.2</td>
<td>15.1</td>
<td>30</td>
</tr>
</tbody>
</table>

We want to know whether parental smoking decreases children's lung capacity as measured by the FVC test.

Is the mean FVC lower in the population of children exposed to parental smoking?

$H_0$: $\mu_{\text{smoke}} = \mu_{\text{no}}$  $=>$  $(\mu_{\text{smoke}} - \mu_{\text{no}}) = 0$

$H_a$: $\mu_{\text{smoke}} < \mu_{\text{no}}$  $=>$  $(\mu_{\text{smoke}} - \mu_{\text{no}}) < 0$ (one sided)

The difference in sample averages follows approximately the t distribution: $t \left( 0, \sqrt{\frac{s_{\text{smoke}}^2}{n_{\text{smoke}}} + \frac{s_{\text{no}}^2}{n_{\text{no}}}} \right)$, $df = 29$

We calculate the t statistic:

$t = \frac{\bar{x}_{\text{smoke}} - \bar{x}_{\text{no}}}{\sqrt{\frac{s_{\text{smoke}}^2}{n_{\text{smoke}}} + \frac{s_{\text{no}}^2}{n_{\text{no}}}}} = \frac{75.5 - 88.2}{\sqrt{\frac{9.3^2}{30} + \frac{15.1^2}{30}}} \approx -3.9$

In table D, for df 29 we find:

$|t| > 3.659 => p < 0.0005$ (one sided)

It’s a very significant difference, we reject $H_0$.

Lung capacity is significantly impaired in children of smoking parents.
Two-sample $t$ confidence interval

Because we have two independent samples we use the difference between both sample averages ($\bar{x}_1 - \bar{x}_2$) to estimate ($\mu_1 - \mu_2$).

**Practical use of $t$: $t^*$**

- $C$ is the area between $-t^*$ and $t^*$.
- We find $t^*$ in the line of Table D for $df =$ smallest $(n_1 - 1; n_2 - 1)$ and the column for confidence level C.
- The margin of error $m$ is:

$$m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = t^* SE$$

---

**Example**

Can directed reading activities in the classroom help improve reading ability? A class of 21 third-graders participates in these activities for 8 weeks while a control classroom of 23 third-graders follows the same curriculum without the activities. After 8 weeks, all children take a reading test (scores in table).

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>46</td>
</tr>
<tr>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>58</td>
<td>55</td>
</tr>
<tr>
<td>71</td>
<td>26</td>
</tr>
<tr>
<td>43</td>
<td>62</td>
</tr>
<tr>
<td>49</td>
<td>37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>$n$</th>
<th>$\bar{x}$</th>
<th>$s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>21</td>
<td>51.48</td>
<td>11.01</td>
</tr>
<tr>
<td>Control</td>
<td>23</td>
<td>41.52</td>
<td>17.15</td>
</tr>
</tbody>
</table>
Solution:

A 95% confidence interval for \((\mu_1 - \mu_2)\) is given by:

\[
CI : (\bar{x}_1 - \bar{x}_2) \pm m;
\]

\[
m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.086 * 4.31 \approx 8.99
\]

with df = 20 conservatively \( t^* = 2.086 \):

<table>
<thead>
<tr>
<th>df</th>
<th>2.000</th>
<th>2.201</th>
<th>2.353</th>
<th>2.447</th>
<th>2.576</th>
<th>2.681</th>
<th>2.781</th>
<th>2.841</th>
<th>2.959</th>
<th>3.091</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.050</td>
<td>0.100</td>
<td>0.150</td>
<td>0.200</td>
<td>0.250</td>
<td>0.300</td>
<td>0.350</td>
<td>0.400</td>
<td>0.450</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>80%</td>
<td>70%</td>
<td>60%</td>
<td>50%</td>
<td>40%</td>
<td>30%</td>
<td>20%</td>
<td>10%</td>
<td>5%</td>
</tr>
</tbody>
</table>

With 95% confidence, \((\mu_1 - \mu_2)\), falls within 9.96 ± 8.99 or 1.0 to 18.9. Third graders who receive directed reading score 1 to 19 points higher than kids who don’t.

Robustness

The two-sample \(t\) procedures are more robust than the one-sample \(t\) procedures. They are the most robust when both sample sizes are equal and both sample distributions are similar. But even when we deviate from this, two-sample tests tend to remain quite robust.

\[\Rightarrow\] When planning a two-sample study, choose equal sample sizes if you can.

As a guideline, a combined sample size \((n_1 + n_2)\) of 40 or more will allow you to work with even the most skewed distributions.
Details of the two sample $t$ procedures

The **true value of the degrees of freedom** for a two-sample $t$-distribution is quite lengthy to calculate. That's why we use an approximate value, $df = \text{smallest}(n_1 - 1, n_2 - 1)$, which errs on the conservative side (often smaller than the exact).

Computer software, though, gives the exact degrees of freedom—or the rounded value—for your sample data.

$$df = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left( \frac{s_1^2}{n_1} \right)^2 + \frac{1}{n_2 - 1} \left( \frac{s_2^2}{n_2} \right)^2}$$

Pooled two-sample procedures

There are two versions of the two-sample $t$-test: one **assuming equal variance** ("pooled 2-sample test") and one **not assuming equal variance** ("unequal" variance, as we have studied) for the two populations. They have slightly different formulas and degrees of freedom.

The pooled (equal variance) two-sample $t$-test was often used before computers because it has exactly the $t$ distribution for degrees of freedom $n_1 + n_2 - 2$.

However, the assumption of equal variance is hard to check, and thus the unequal variance test is safer.
When both population have the same standard deviation, the pooled estimator of $\sigma^2$ is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

The sampling distribution for $(x_1 - x_2)$ has exactly the $t$ distribution with $(n_1 + n_2 - 2)$ degrees of freedom.

A level $C$ confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t_C s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

To test the hypothesis $H_0: \mu_1 = \mu_2$ against a one-sided or a two-sided alternative, compute the pooled two-sample $t$ statistic for the $t(n_1 + n_2 - 2)$ distribution.

### 7.2 Comparing Two Means

#### Using the TI-83

We next consider confidence intervals and significance tests for the difference of means $\mu_1 - \mu_2$ given two normal populations that have unknown standard deviations. The results are based on independent random samples of sizes $n_1$ and $n_2$. For the most accurate results, we can use the [Z-SampInt] and [Z-SampTest] features from the STAT TESTS menu.

These features require that we specify whether or not we wish to use the pooled sample variance $s_p^2$. We should specify “Yes” only when we assume that the two populations have the same (unknown) variance. In this case, the critical values $t^*$ are obtained from the $t(n_1 + n_2 - 2)$ distribution and the standard error is $s_p \sqrt{1/n_1 + 1/n_2}$. When we specify “No” for the pooled variance, then the standard error is $\sqrt{s_1^2/n_1 + s_2^2/n_2}$ and the degrees of freedom $r$ are given by

$$r = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1} \frac{s_1^2}{n_1} + \frac{1}{n_2-1} \frac{s_2^2}{n_2}}$$

But if the true population standard deviations $\sigma_1$ and $\sigma_2$ are known, then we should use the [Z-SampInt] and [Z-SampTest] features for our calculations.
Comparing two Means: Using the TI-83

- STAT > TESTS > 0: 2-SampTInt

```
2-SampTInt
Inpt:Data Stats
List1:L1
List2:L2
Freq1:1
Freq2:1
C-Level:.95
Pooled: NO Yes

2-SampTInt
(-12.28,77.083)
df=4.649534043
\bar{x}_1=123.8
\bar{x}_2=116.4
5\bar{x}_1=4.6
5\bar{x}_2=16.09
```

Comparing two Means: Using the TI-83

- STAT > TESTS > 4: 2-SampTTest

```
2-SampTTest
Inpt: Data Stats
List1:L1
List2:L2
Freq1:1
Freq2:1
\mu_1=\mu_2 < \mu_2 > \mu_2
Pooled: NO Yes

2-SampTTest
t=9.887816684	p=0.047
\mu_1=123.8
\mu_2=116.4
```
Example 7.8 Does Polyester Decay? How quickly do synthetic fabrics decay in landfills? A researcher buried polyester strips for different lengths of time, then dug up the strips and measured the force required to break them. Breaking strength is easy to measure and a good indicator of decay. One set of strips was dug up after 2 weeks and the other was dug up after 16 weeks. We suspect that decay increases over time. The results are given in the table below. Test the hypothesis $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 > \mu_2$, then give a 95% confidence interval for $\mu_1 - \mu_2$.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>$\bar{x}$</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 weeks</td>
<td>5</td>
<td>123.80</td>
<td>4.60</td>
</tr>
<tr>
<td>16 weeks</td>
<td>5</td>
<td>116.40</td>
<td>16.09</td>
</tr>
</tbody>
</table>

Solution. Use the 2-SampTTest feature from the STAT TESTS menu, and set Inpt to Stats. Enter the given statistics, set the alternative, and enter No for Pooled. Then press ENTER on Calculate.

We obtain a p-value of 0.1857 from a test statistic of 0.8888 with 4.6495 degrees of freedom. If the true mean decay rates were equal, then there would be a good chance of our means being as different as they are (the difference is 7.4) with samples of these sizes. We therefore fail to reject $H_0$ and conclude that there is not a significant difference in decay of polyester with the extra 14 weeks.
To calculate a confidence interval for \( \mu_1 - \mu_2 \), we use 2-SampTInt. Our data is still there, so one can simply move the arrow down to enter the desired confidence level and then calculate. We obtain the interval \((-12.28, 27.08)\). These were small samples, which is one reason for the wide interval. We also note that 0 is included in the interval, which supports our conclusion of no significant increase in polyester decay with the additional 14 weeks of burial.

Example 7.9 College Study Habits. 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>

(a) Examine each sample graphically to determine if the use of a \( t \) procedure is acceptable.
(b) Test the supposition that the mean score for all men is lower than the mean score for all women among first-year students at this college.
(c) Give a 90% confidence interval for the mean difference between the SSHA scores of male and female first-year students at this college.
Solution. (a) We shall make normal quantile plots of these data. Enter the women’s scores into list L1 and the men’s scores into list L2. In the STAT PLOT screen, adjust the Type settings for both Plot1 to the last type for a normal quantile plot, then graph each list separately. Press ZOOM[9] to display the plots. The resulting plots appear close enough to linear to warrant use of \( t \) procedures.

\[
\begin{array}{ccc}
101 & 103 & 148 \\
102 & 146 & 192 \\
106 & 74 & 188 \\
129 & 88 & 113 \\
\text{L3(3) =} & & \\
\end{array}
\]

(b) Next, let \( \mu_1 \) be the mean SSHA 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 \( H_A: \mu_1 > \mu_2 \). On the 2-SampTTest screen, set Inpt to Data, enter the desired lists L1 and L2, set the alternative to > \( \mu_2 \), enter No for Pooled, and press ENTER on Calculate or Draw.

We obtain a \( p \)-value of 0.02358. If the true means were equal, then there would be only a 2.358% chance of \( \bar{x}_1 \) being so much larger than \( \bar{x}_2 \) with samples of these sizes. The relatively low \( p \)-value gives us evidence to reject \( H_0 \) and conclude that \( \mu_1 > \mu_2 \). That is, the mean score for all men is lower than the mean score for all women among first-year students at this college.
(c) Adjust the settings in the 2-SampTInt screen and calculate. We obtain (3.5377, 36.073). That is, the mean score of female first-year students should be from about 3.5377 points higher to about 36.073 points higher than the mean score of male first-year students at this college.

---

### Pooled Two-Sample t Procedures

**Example 7.10 Infants’ Hemoglobin.** A study of iron deficiency in infants compared samples of infants following different feeding regimens. Here are the summary results on hemoglobin levels at 12 months of age for two samples of infants:

<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) Is there significant evidence that the mean hemoglobin level is higher among breast-fed babies? State $H_0$ and $H_A$, and carry out a $t$ test.

(b) Give a 95% confidence interval for the difference in mean hemoglobin levels between the two populations of infants.
Solution. Let $\mu_1$ be the mean hemoglobin level for all breast-fed babies and let $\mu_2$ be the mean level for all formula-fed babies. Because the sample deviations are so close, it appears that true standard deviations among the two groups could be equal; thus, we may use the pooled two-sample $t$ procedures. For part (a), we will test $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 > \mu_2$.

Call up the 2-SampTTest feature, and set Inpt to Stats. Enter the given statistics, set the alternative, enter Yes for Pooled, and calculate.

With a $p$-value of 0.052, we conclude that there is not statistical evidence (at the 5% level) to reject $H_0$. If the true means were equal, then there is greater than a 5% chance of $\bar{x}_1$ being 0.9 higher than $\bar{x}_2$ with samples of these sizes.

(b) Next, calculate a 95% confidence interval with the 2-SampTInt feature set to Yes on Pooled. We see that $-0.1938 < \mu_1 - \mu_2 < 1.9938$. That is, the mean hemoglobin level of breast-fed babies could be from 0.1938 lower to 1.9938 higher than the mean level of formula-fed babies. Since 0 is in the interval, we are further convinced that there is not a significant difference between breast-fed and formula-fed babies’ hemoglobin levels, on average.
Inference for Distributions
Optional Topics in Comparing Distributions

IPS Chapter 7.3

Objectives (IPS Chapter 7.3)

Optional topics in comparing distributions

- Inference for population spread
- The F test
- Power of the two-sample t-test
Inference for population spread

It is also possible to compare two population standard deviations $\sigma_1$ and $\sigma_2$ by comparing the standard deviations of two SRSs. However, these procedures are not robust at all against deviations from normality.

When $s_1^2$ and $s_2^2$ are sample variances from independent SRSs of sizes $n_1$ and $n_2$ drawn from normal populations, the $F$ statistic

$$F = \frac{s_1^2}{s_2^2}$$

has the $F$ distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom when $H_0: \sigma_1 = \sigma_2$ is true.

The $F$ distributions are right-skewed and cannot take negative values.

- The peak of the $F$ density curve is near 1 when both population standard deviations are equal.
- Values of $F$ far from 1 in either direction provide evidence against the hypothesis of equal standard deviations.

Table E in the back of the book gives critical $F$-values for upper $p$ of 0.10, 0.05, 0.025, 0.01, and 0.001. We compare the $F$ statistic calculated from our data set with these critical values for a one-side alternative; the $p$-value is doubled for a two-sided alternative.
### Table E  
*F* distribution critical values

<table>
<thead>
<tr>
<th>p</th>
<th>df&lt;sub&gt;num&lt;/sub&gt; = n&lt;sub&gt;1&lt;/sub&gt; - 1</th>
<th>df&lt;sub&gt;den&lt;/sub&gt; = n&lt;sub&gt;2&lt;/sub&gt; - 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.10</td>
<td>39.86</td>
<td>49.50</td>
</tr>
<tr>
<td>0.05</td>
<td>61.45</td>
<td>125.11</td>
</tr>
<tr>
<td>0.025</td>
<td>647.70</td>
<td>864.16</td>
</tr>
<tr>
<td>0.010</td>
<td>4052.2</td>
<td>5624.6</td>
</tr>
<tr>
<td>0.001</td>
<td>40,028.4</td>
<td>56,000</td>
</tr>
</tbody>
</table>

|     | 3                                 | 4                                  |
| 0.10 | 8.55                              | 10.16                              |
| 0.05 | 18.51                             | 19.26                              |
| 0.025| 38.51                             | 39.30                              |
| 0.010| 98.50                             | 99.30                              |
| 0.001| 998.50                            | 999.30                             |

|     | 5                                 | 6                                  | 7                                  |
| 0.10 | 10.13                             | 11.20                              | 11.31                             |
| 0.05 | 17.44                             | 18.44                              | 18.51                             |
| 0.025| 34.12                             | 35.94                              | 37.50                             |
| 0.010| 107.03                            | 118.03                             | 118.10                            |
| 0.001| 784.14                            | 948.64                             | 949.44                            |

#### 7.3 Optional Topics in Comparing Distributions

We now demonstrate a test for determining whether or not two normal populations have the same variance. If so, then we would be justified in using the pooled two-sample *t* procedures for confidence intervals and significance tests about the difference in means. For the test, we will need the 2-SampFTest feature from the STAT TESTS menu.

One word of caution about this test is in order, however. This is extremely sensitive to any departures from normality, and the lack of robustness does not improve substantially with larger sample sizes. As such, it should be used with wariness; it is difficult to tell whether a significant *p*-value is due to a difference in the standard deviations, or to a lack of normality.
The F Ratio Test

Example 7.11 More on SSHA. Consider again the data from Example 7.9 regarding the SSHA scores of first-year students at a private college. Test whether the women’s scores are less variable than the men’s.

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

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

Solution. Let \( \sigma_1 \) be the standard deviation of all women’s scores and let \( \sigma_2 \) be the standard deviation for all men’s scores. We shall test \( H_0: \sigma_1 = \sigma_2 \) versus \( H_A: \sigma_1 < \sigma_2 \). To do so, first enter the data sets into lists, say L1 and L2. Next, bring up the 2-SampFTest screen from the STAT TESTS menu, set Inpt to Data, enter the appropriate lists names and alternative, and calculate.

With a \( p \)-value of 0.18624, we do not have strong evidence to reject \( H_0 \). If \( \sigma_1 \) were equal to \( \sigma_2 \), then there would be an 18.6% chance of the women’s sample deviation of \( Sx1 = 26.4363 \) being so much lower than the men’s sample deviation of \( Sx2 = 32.8519 \). Thus, we cannot assert strongly that the women’s scores are less variable.
Example 7.12  Calcium and Blood Pressure. Here are the summary statistics for the drop in blood pressure from two sample groups of patients undergoing treatment. Test to see if the groups in general have the same standard deviation.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>$\bar{x}$</th>
<th>s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium</td>
<td>10</td>
<td>5.000</td>
<td>8.743</td>
</tr>
<tr>
<td>Placebo</td>
<td>11</td>
<td>-0.273</td>
<td>5.901</td>
</tr>
</tbody>
</table>

Solution. Let $\sigma_1$ be the standard deviation of all possible patients in the calcium group, and let $\sigma_2$ be the standard deviation of all possible patients in the placebo group. We will test the hypothesis $H_0$: $\sigma_1 = \sigma_2$ versus $H_A$: $\sigma_1 \neq \sigma_2$. Bring up the 2-SampFTest screen and set Inpt to Stats. Enter the summary statistics and alternative, then calculate.

We obtain an $F$-statistic of 2.195 and a $p$-value of 0.2365. If $\sigma_1$ were equal to $\sigma_2$, then there would be about a 23.6% chance of $Sx1$ and $Sx2$ being so far apart with samples of these sizes. This evidence may not be significant enough to reject $H_0$ in favor of the alternative.