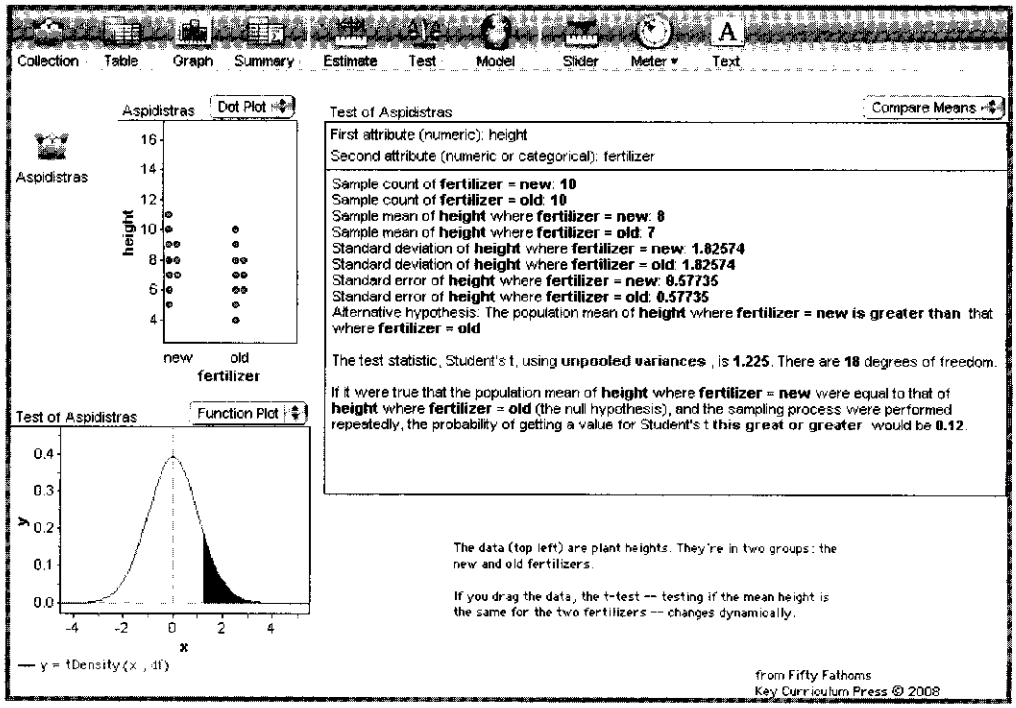
Demo 38: Using a *t*-Test to Compare Means

Comparing means with Student's t

In Demo 37, "Scrambling to Compare Means," we used randomization to compare the means of two groups. Now we'll compare the same two groups using the traditional *t*-test for difference of means.



What To Do

▷ Open **Compare Means Using t.ftm.** It will look like the illustration.

This shows the same data as before, but with Fathom's built-in *t*-test to help us decide if the two groups are different. The *P*-value (at the end of the test object) is 0.12—not usually small enough to justify rejecting the null hypothesis. Let's change the data.

▷ Play with the data to see how changing a single point affects the results of the test.

▷ Arrange the data, changing only one data point, so that the *P*-value is 0.05 or less.

Questions

1 How can you predict, before you drag a given point, which way the *P*-value will change? **Sol**

2 At the bottom of the window, what does the graph of the curve show?

Extension

▷ Return the data (through reopening the file or by multiple undo) to the way they were in the beginning.

▷ In the graph of the data, grab the lowest point among the **new** fertilizer data (**height = 5**) and slowly drag it upward. This will make the two groups more different.

▷ Watch the *P*-value as you drag. It goes down (to about 0.028) and then, eventually, starts back up.

Challenges

3 Explain how the *P*-value changes. That is, by dragging *up* a point from the higher group, you're making the two groups more different, aren't you? So it makes sense that *p* should decrease. But then why does it increase again? **Sol**

4 Investigate the two-tailed case: First, figure out how to change the test to a two-tailed test. Then describe how the results and displays are different. How much more (or less) do you have to change the data to get a *P*-value of 0.05 or less?

5 If you have done Demo 41, "Paired Versus Unpaired," (or even if you haven't), why do you suppose we didn't use a paired test here? Aren't paired tests generally more effective?

What You Should Take Away

- We often think of the test as comparing the means (in fact, it's called Compare Means, isn't it?). But the test really hinges on the *spreads* in the groups. Why? Because they set the scale by which the difference is measured: That difference is not measured in inches or centimeters—it's measured in standard errors.
- The results of a hypothesis test can be very sensitive to individual data points. You can see in this demo that one wacky point—or even a not-so-wacky one if your sample is small—can mean the difference between a highly significant *P*-value and a boring one. And it can go either way. Therefore, be skeptical: *Always* look at the data. Make the graph. Play "what-if" games with the points. If you just turn the crank on the test and cheer or weep depending on how *P* compares to 0.05, you're rushing things.