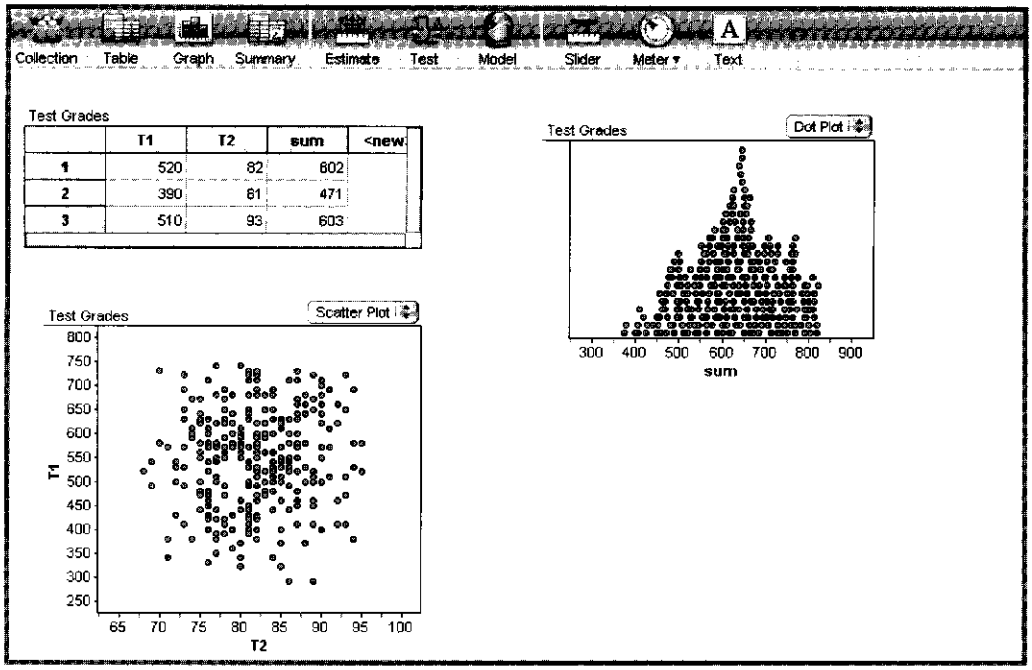
Demo 7: Standard Scores

Using standard scores to compare unlike scales • Making a scale in terms of standard deviations

Suppose we want to come up with a fair way of combining scores on two tests. Unfortunately, the scores are very different—they have different centers and spreads. The first score, like the SAT, can range only between 200 and 800; unlike the SAT, its scores pretty much cover that range. The second score can take any value between 0 and 100, though scores generally are above 70. Furthermore, since these tests are in completely different areas, they are uncorrelated. Nevertheless, we have to come up with a combined score. (Think of the biathlon in the Winter Olympics, where the contestants ski and shoot; there has to be some way of awarding the medals.)

Let's begin by trying the simplest solution—simply adding the two scores. We will see why that's a bad idea.



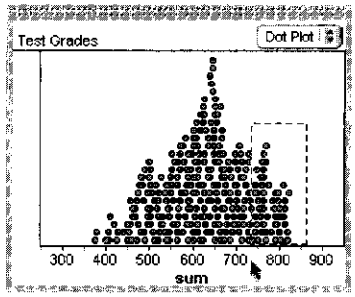
What To Do

▹ Open **Standard Scores.ftm.** It looks something like the illustration.

Here you can see the top of a table of test scores **T1** and **T2,** and their **sum.** The lower-left graph shows **T1** plotted against **T2;** the graph in the upper right shows the distribution of their sum.

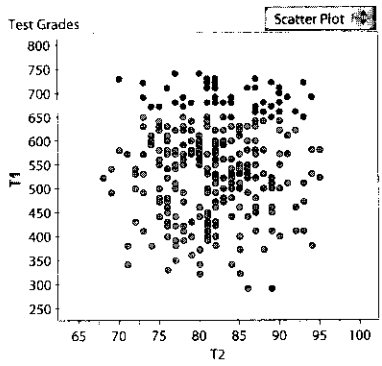
Let's see why simply adding is a bad way to combine the scores in this situation:

▹ With your mouse, drag a rectangle around the highest couple dozen scores in the **sum** graph, and release. The points you have selected turn red—and so do the *same* cases in the other graph.



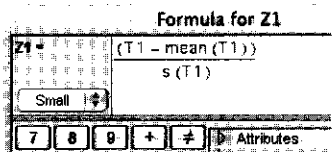
You should see a graph like the one in the next illustration. The band of selected points is essentially horizontal. Those points all have high scores on **T1,** *whatever their performance on T2.* That is, if we use only the sum, it's pretty much the same as ignoring **T2.**

© 2008 Key Curriculum Press

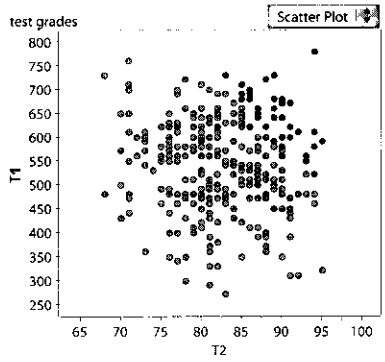


The problem is that the scales of the two scores are so different. The fact that the numbers are high for **T1** (and have such a range) doesn't mean it should dominate. One strategy is to transform the scores so that they have the same center and spread—and then add. This is called changing the data to *standard scores*. Let's try it:

- In the case table, click in the **<new>** box at the top of the blank column. Name the new attribute **Z1** and press **Return** or **Enter**. (*Z* is a traditional letter to use for a standard score.)
- Select the new attribute by clicking once, and choose **Edit Formula** from the **Edit** menu. The formula editor appears.
  - ⇨ The shortcut for **Edit Formula** is ⌘+**E** on the Mac or **Control+E** in Windows.



- Enter this formula: **(T1 – mean(T1)) / s(T1).** That is, take the difference of the score from its mean, and then divide the result by its standard deviation. That way, if your score on **T1** is one standard deviation above the mean in the group, your *standard score* for that test (**Z1**) will be +1.0. The formula will look like the illustration in the editor; when you have done this correctly, close the editor with **OK**.
  - ⇨ You can enter the formula many ways; one way is to type those keys exactly (including all parentheses).
- Make another new attribute, **Z2**, analogously. You may have to scroll the case table to the right to get another **<new>** column; the formula is **(T2 – mean(T2)) / s(T2).**
- Make a new graph by dragging one from the shelf. Put it in the lower right where you have some space.
- Drag **Z1** to the vertical axis of the new graph and **Z2** to the horizontal axis. Note how the graph looks just like the **T1-T2** graph, only with different scales.
- Finally, make a third new attribute: **Zsum.** Give it the formula **Z1 + Z2.** Drag its name to the horizontal axis of the top graph, replacing **sum.**
- Now select the top couple dozen points in **Zsum.** Now the **T1-T2** graph shows the top combined scores as a triangular region in the upper right—the people who did well on both tests, as you can see in the illustration.



Note: A *z*-score is *dimensionless,* similar to the way *t* is, as we will discuss in Demo 18, "The Road to Student's *t*."

Extensions

- Drag **sum**, and then **Zsum**, to the *middle* of either scatter plot.
- Edit the formulas for **T1** and **T2** to give them a different-shaped distribution (for example, exponential). Then use this machinery to see if you think this way of combining scores is fair for distributions that are not roughly bell-shaped.

© 2008 Key Curriculum Press