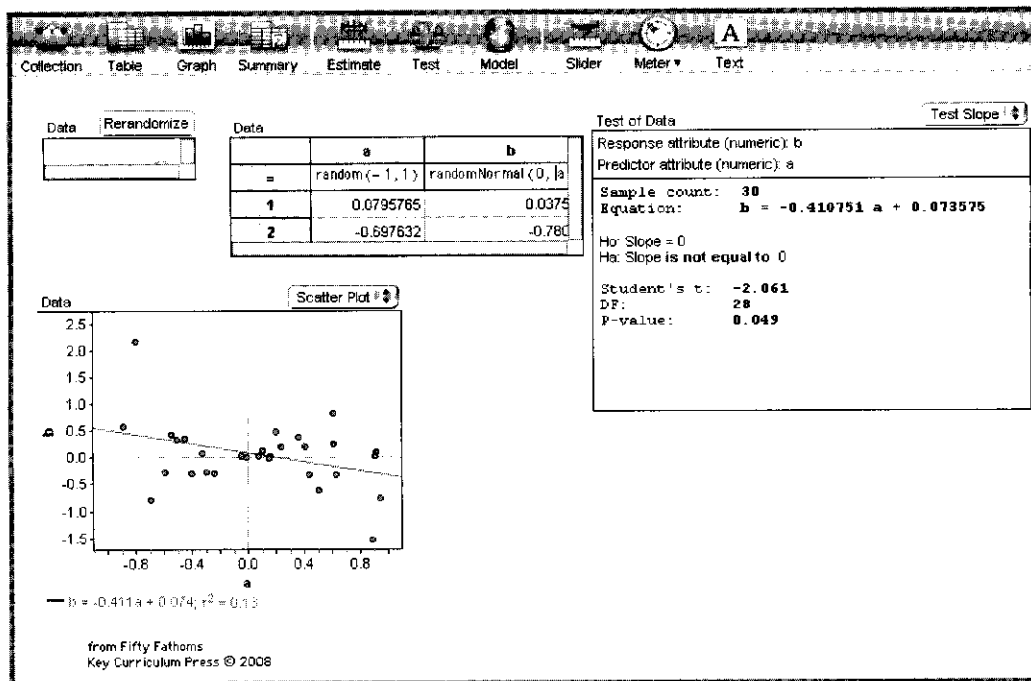


Demo 46: Heteroscedasticity and Its Consequences

*Homoscedasticity—an assumption behind many statistical calculations •
What happens when that assumption is not met*

When you calculate a least-squares line, you get the slope and intercept. Suppose you get a slope of 0.1. Is that significantly different from 0? To find out, you do a test of that slope (or estimate it with a confidence interval). That test has behind it an assumption of *homoscedasticity*—that the variance of the y variable is the same for any value (or infinitesimally thin swath) of x .

First, though, why is this demo in the section on power? Because to understand it, you need to have delved deeply into the meaning of a statistical test. In particular, you need to know that, if the assumptions are met and the null hypothesis is true, the distribution of P -values for repeated tests will be uniform. We learned that in Demo 43, “The Distribution of P -Values.” Without knowing that, it’s hard to explain why homoscedasticity is important. After all, the procedure for getting the slope will give you the right answer on the average if the data are heteroscedastic. But the inferential statements you make, as we shall see, will be incorrect.



What To Do

- Open **Heteroscedasticity.ftm**. It looks something like the illustration.

We have a collection, **Data**, with two attributes, **a** and **b**, plotted against each other. Both are determined randomly. You can see their formulas in the “=” row of the case table. The attribute **a** is uniform in the range $(-1, 1)$; **b** is normally distributed with a mean of zero—whatever the value of **a**—but its standard deviation gets larger the farther it is from the axis. That

But How Do You Pronounce It?

Paul Velleman defends the hard “c,” with “sc” as in “scatter”: ho-mo or he-te-ro-ske-das-ti-si-ty. And he has a Greek scholar to back him up. Yet Random House goes for the soft (or silent) “c.”

is, its true distribution looks like butterfly wings or a bow tie, and the true slope of **b** on **a** is zero—it must be because its mean value is zero whatever the value of **a**. The test, upper right, is testing whether the slope is not equal to zero; you can see the *P*-value of 0.3 in the illustration—your file will be different. That value suggests that we should *not* reject the null hypothesis (which is good, because it's true).

- ▷ Press **Rerandomize** (in the data collection) repeatedly. The graph and test will change.
- ▷ Look informally at the slopes. For example, count how many slopes are positive and how many are negative in ten tries. You should be able to detect no obvious bias.
- ▷ Do the same for *P*-values. You will probably judge that they are spread pretty well over the range [0, 1].

But there *is* nonuniformity; we just need a way to detect it.

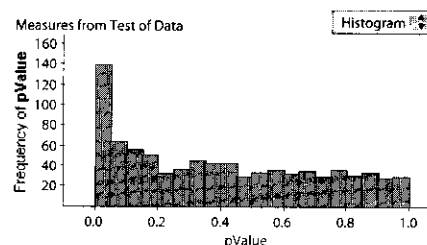
- ▷ Choose **Show Hidden Objects** from the **Object** menu. An empty collection, **Measures from Test of Data**, and a graph, appear.
- ▷ Press the **Collect More Measures** button in the collection. Twenty points appear in the graph. They'll be way down at the bottom. They are *P*-values from the 20 tests Fathom performed on 20 new samples.

Since the null hypothesis is true, this distribution of *P*-values should be uniform. But is it?

- ▷ Do it again, a few times, watching the distribution begin to grow.
- ▷ Finally, double-click that measures collection to open its inspector. It should open to the **Collect Measures** panel; if it does not, click on the tab to take you there. Change the panel to turn off animation and to collect 200 points at a time. Close the inspector.

⇒ For extra speed, iconify the test object in the upper right, or even the scatter plot. (Do so by dragging a corner until they're small.) Then Fathom does not have to redraw them.

- ▷ Again press **Collect More Measures** a few times, and build up the distribution so that you see a clear nonuniformity. Your graph will eventually look something like the illustration.



That is, it seems as if the test gives a low *P*-value, rejecting the null hypothesis more often than it would if the distribution were uniform. How much more often? Let's see:

- ▷ Select the histogram by clicking on it once. Then choose **Scale | Relative Frequency** from the **Graph** menu. In our graph, it showed that the left-hand bar—the one showing **pValue** < 0.05—had a relative frequency of 0.16.

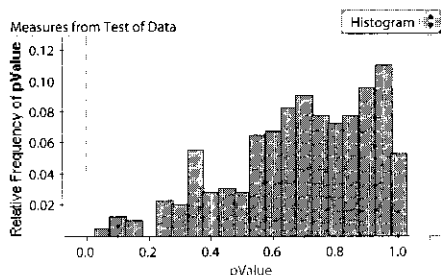
What does this mean? Suppose we said we would reject the null hypothesis with a *P*-value of less than 0.05. Even though the null hypothesis is true—**b** does not depend on **a**, so the slope relating them is in fact zero—we would reject it 16% of the time. That is, *we would get three times as many Type I errors as we should*. That means that the whole logic of the hypothesis test is wrong; if you get a low *P*, you shouldn't be as confident that the slope is not zero.

The graph can tilt the other way, too:

- Open **Heteroscedasticity 2.ftm**. It will look a lot like the first one except that the **Measures from Test of Data** collection is already showing.

The distribution of **b**, instead of looking like a butterfly, looks like a diamond: It has a wider variance in the middle and a narrower variance at the ends. See the formulas to understand how it's done.

- Press **Rerandomize** on the **Data** collection a few times to see the scatter plot and its least-squares line update, as before. Note the values in the test, too.
- Press **Collect More Measures** in the **Measures from Test of Data** collection. Twenty more data points join the collection, and the bars grow in the histogram. Repeat a few times to watch the distribution of *P*-values begin to develop.
- As before, open the inspector, turn off animation, and increase the number of measures to 200. Press **Collect More Measures** to build up the distribution.
- Select the histogram and choose **Scale | Relative Frequency** from the **Graph** menu. The graph will look something like the illustration.



This time, the distribution of *P* is tilted the other way. That is, a Type I error is extremely rare. That also means that, if the true slope were *not* zero, it would be hard to reject the null hypothesis; that is, the power of this test is smaller than it would be if the data were homoscedastic.

Challenges

- 1 Explain qualitatively why the slopes in the “butterfly” distribution would give you too many significant results, and why the slopes in the “diamond” distribution would give you too few.

Sol

- 2 Suppose we wanted to correct the test for the heteroscedasticity. In this latest, “diamond” distribution case, what value of *P* had about 5% of the cases below it? (It looks like about 0.2; see if you agree.) So then could we say that we get a significant result (at the 0.05 level) if $P < 0.2$?
- 3 Do the same for the “butterfly” case. What is the critical value for *P* if you want a “true” alpha of 0.05?
- 4 In the “diamond” case, alter the simulation so that **b** really does have a slope. That is, make a slider **m**, and make **b**’s formula

$$m \cdot a + \text{randomNormal}(0, 1 - |a|)$$

How does the distribution of *P*-values change as **m** increases? Explain what you find.

- 5 We commented that the power of the test in the diamond case was smaller than it would have been if the distribution were homoscedastic. Does that mean that heteroscedasticity made the test in the butterfly case *more* powerful? Explain this; that is, if you think it’s more powerful, wouldn’t we always want to encourage butterfly heteroscedasticity?
- 6 Chris Olsen, on reviewing this demo, suggested that, while the diamond and butterfly examples were good, the more common way you get heteroscedasticity in practice is in proportion to the measure. That is, the distribution looks more like a trumpet. Predict first, and then simulate: What difference would that kind of distribution make in the distribution of *P*? (Simulation hint: Change the formula for **a** to be positive only, then look at the formula for **b** to see if you need to change it as well.)