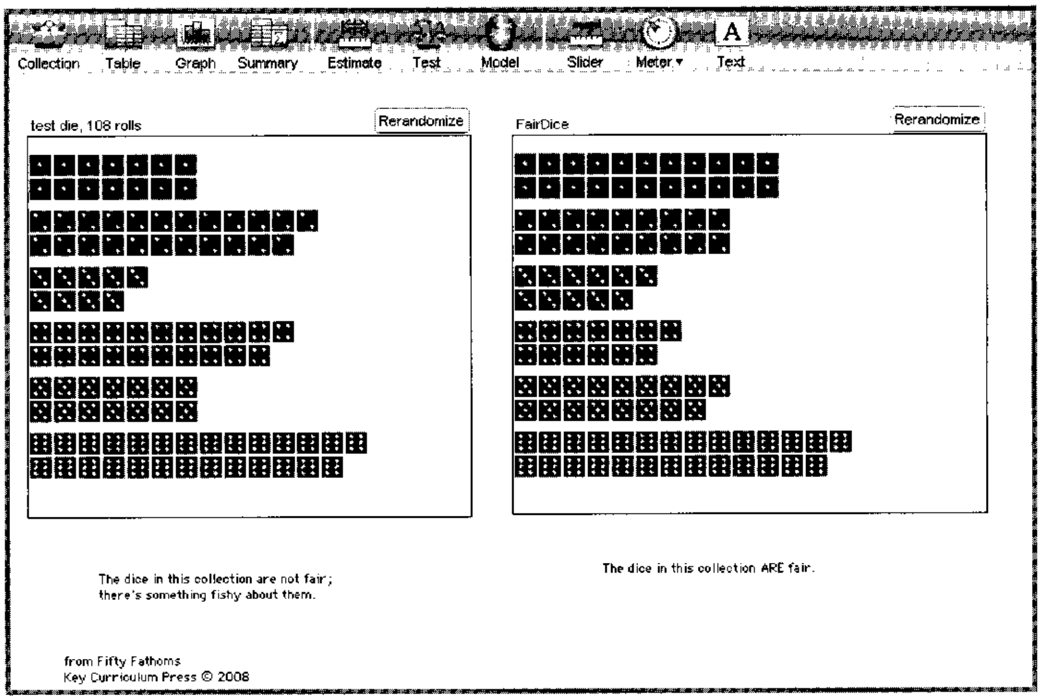
Demo 36: Fair and Unfair Dice

Creating a measure of "fairness" • Sampling distributions • Testing hypotheses empirically • The chi-square statistic

Suppose you have a test die, and you roll it many times, and you get suspicious. You think it might be unfair. How do you check? Compare it to dice that you know are fair. But even fair dice give uneven distributions, as you will see. The key is to figure out how to measure the unevenness of a distribution, and then to compare the unevenness of your test die's distribution to a sampling distribution of the same statistic for the fair dice.



What To Do

▹ Open **Fair and Unfair Dice.ftm.** It will look something like the illustration.

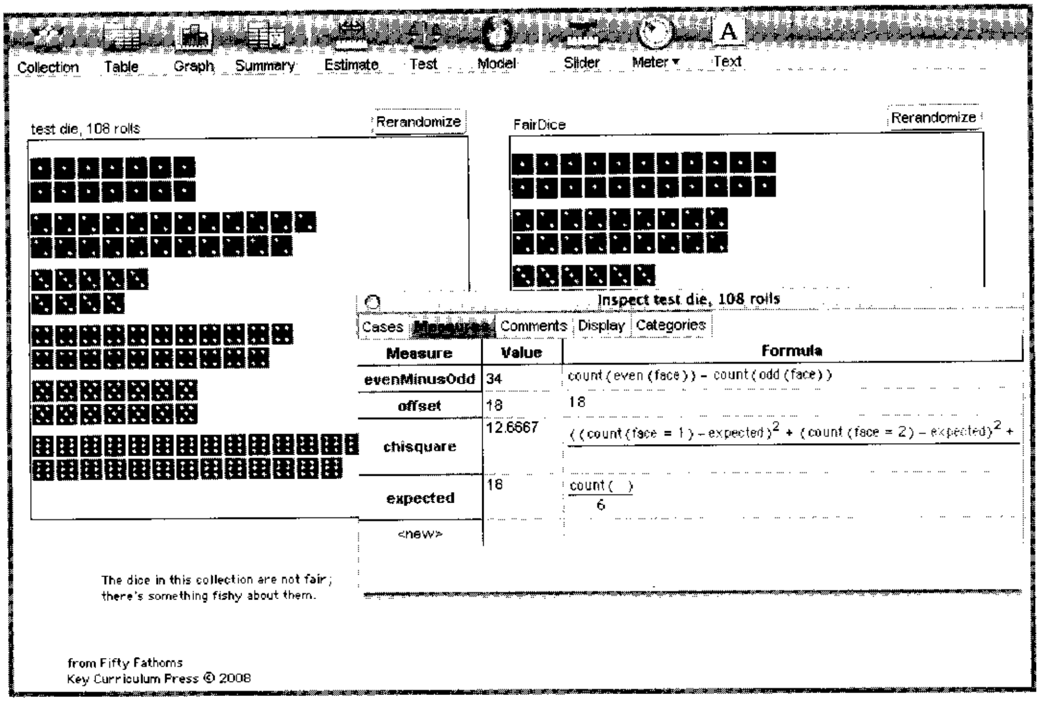
Here you see two collections of 108 cases each. The left one is a test die; the right one is fair. Note in advance that since there are 108 rolls, you would expect 18 of each number (since $6 \times 18 = 108$).

▹ Press the **Rerandomize** button on the **FairDice** (right-hand) collection repeatedly. Note how the distribution is often far from "fair."

▹ Do the same on the **test die** (left-hand) collection. Again, watch the distribution. This time, see if you can figure out what it is about the distribution that makes it appear unfair. (It is not obvious from the picture above—you'll see it only when you see several examples.)

▹ Leave the **test die** collection in a "typical" state (the picture above is unusually even; pick one that shows the unfairness) and **Rerandomize** the right again, paying attention to what it is about the test distribution that is different from the fair. (Note: Don't **Rerandomize** the **test die** collection again—we'll be comparing fair dice to *this* distribution from here on.)

In general, the **test die** picture should suggest there are more even than odd numbers in the test collection. Now we need to develop a *measure* of how unfair it is. This measure is a *statistic*—a number—that applies to the entire collection. Ideally, it should be large if the distribution is extremely unfair, and small—or zero—if it is fair.



We have created a measure that fits these criteria: **evenMinusOdd,** which is the number of evens minus the number of odds. Let's see what that quantity is.

- ▹ Double-click in the left-hand (**test die**) collection to open its inspector. Click on the **Measures** tab to open that panel. It will resemble the inspector in the illustration.
- ▹ Note that our value for **evenMinusOdd** is 34 in the illustration. There are 34 more even-numbered dice than odd ones. If everything were flat, that number would be zero.
- ▹ Close the inspector. (You might recognize **chisquare** as another statistic that measures fairness. We'll come back to it later.)

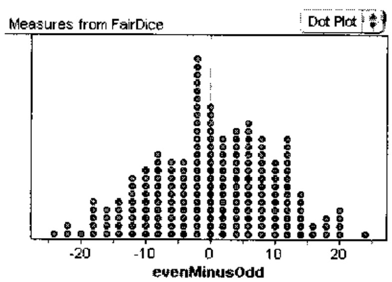
But is 34 (or whatever you have) an *unusual* value? To find out, we will calculate the value for the fair dice. It will not always be zero, because of random variation. Sometimes it could even be quite large. As big as 34? Let's see.

- ▹ Double-click the **FairDice** collection to open its inspector, and click on the **Measures** tab to go to that panel. There you can see the current value for **evenMinusOdd.** Note it informally (that is, is it bigger than 34?).
- ▹ Press the **Rerandomize** button on that collection repeatedly, watching the value of **evenMinusOdd** update in the foreground. Repeat until you have a good sense of whether 34 is unusual. (It is.)

Wouldn't it be great if Fathom would keep track of these numbers?

▹ Close the inspector to save screen space.

▹ Now drag the lower-right corner of the **FairDice** collection up and to the left, to make it so small that it turns into an icon. You will reveal the **Measures from FairDice** collection and a graph of 200 values of **evenMinusOdd,** each one representing a new resampling of 100 fair dice. It should look something like the illustration.



▹ Let's collect a different set of 200 measures. Select the **Measures from FairDice** collection and choose **Collect More Measures** from the **Collection** menu. Repeat as many times as you like.

At last we can answer the question "Is 34 (or whatever number you had) an unusual value?" Look at the distribution of **evenMinusOdd** values in the graph, and see where your test value (34 for us) lies. Chances are, if it's not outside the distribution entirely (as in the illustration above), there are only a few points at 34 or higher.

Since there are 200 points in the graph, you can compute the *P*-value empirically—the probability that a fair die will give a value of **evenMinusOdd** that large *or larger* in a sample of 108 rolls. Generally, if $P < 0.05$ (that is, fewer than 10 as-extreme points in 200 samples), people consider that significant. That is, a fair die will rarely show behavior like your test die. We can infer that the test die is unfair (and risk being wrong at the 5% level).

Questions

1 Values for **evenMinusOdd** are always even. Why is that so?

2 What is the largest possible value you could get for **evenMinusOdd**? Explain why. **Sol**

3 What is your *P*-value? Choose **Collect More Measures** to compute several *P*-values.

Extension: Chi-Square

The measure **evenMinusOdd** is elegant, but useful only if we observe that the evens outnumber the odds, so that their difference is relevant. There is another statistic, chi-square (written χ^2), that measures how far the distribution is from "even" no matter how it differs. Its formula is

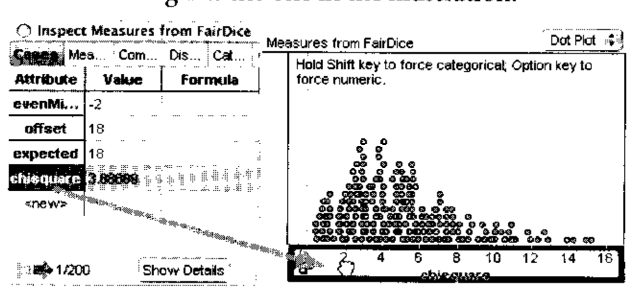
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where the sum is over all possible outcomes—in this case, the die faces 1 through 6. O is the number of cases *observed* to have that outcome, and E is the number *expected* to have that outcome. In our case, E is 18—1/6 of the 108 dice—for each of the six outcomes.

▹ Open the inspector for the test die again, and look at the **Measures** panel. Note the value for **chisquare.** In our example above, it was 11.6667. Close the inspector to open up screen space.

▹ Open the inspector for the **Measures from FairDice** collection (double-click the box of balls). Click on the **Cases** tab to bring that panel to the front.

▹ Drag the name of the attribute **chisquare** to the horizontal axis of the graph, replacing **evenMinusOdd.** Your graph should look something like the one in the illustration.



Note: The attribute **chisquare** also appears as a measure in the **FairDice** collection itself. If you drag it to a graph, nothing will happen. Be sure you're looking at the **Cases** panel of the **Measures from FairDice** collection.

- ▹ Count (approximately) how many cases are equal to or greater than your test value (in our case, where $\chi^2 = 11.6667$, about 8 cases). Compute the *P*-value (for us it is 8/200 = 0.04).
- ▹ Collect new sets of measures as before: Select the **Measures from FairDice** collection and then choose **Collect More Measures** from the **Collection** menu.

As before, you can calculate *P*-values repeatedly. In each graph, look at how many sets of 108 fair dice produce results that are at least as extreme as your test distribution.

More Questions

4 In general, does it seem that the test die results are consistent with its being a fair die based on the chi-square distribution?

5 Does the **chisquare** statistic seem more unusual than, less unusual than, or about the same as the **evenMinusOdd** statistic? **Sol**

6 When you compute *P*, why do you divide by 200, even though there are only 108 fair dice in the collection?

Challenges

7 Suppose all of the rolls in 108 die rolls came up *one*. What would chi-square be then?

8 Explain why, in the equation for χ^2, the measure will be small when the results are "fair" and large when they are grossly unfair.

9 Explain why χ^2 can never be negative.

10 Why do you suppose (as is likely) there were more cases of fair dice having an extreme **chisquare** than there were of fair dice having an extreme **evenMinusOdd**—that is, a value for the statistic greater than that of the test die?

11 Suppose we had a test value for **evenMinusOdd** of 30. When we look at the sampling distribution, and count how many of the **FairDice** samples are greater than or equal to 30, should we also add in the samples that have an **evenMinusOdd** less than or equal to −30? Explain.

12 Change the unfairness of the test die. You can find it in the formula for **face** in the **test die** collection. Then see how well you can detect the unfairness of the die.

13 Make a wholly new measure of unfairness and see if it works. You'll need to create a measure for it in both the **test die** (to compute the test statistic) and **FairDice** collections. Then **Collect More Measures** to get a graph of its distribution. Finally, compare the test statistic to the distribution to see how extreme your test die is. One example of such a measure might be "eighteen minus the number of ones."

14 Explore and explain the advantages and disadvantages of using the traditional measure, χ^2, as opposed to these measures we're making up ourselves. You can get one insight by doubling the number of cases in the test collection, and then also doubling the number in **FairDice.** You will see that for some statistics, the distribution of sample statistics will change (location, probably not shape) whereas for χ^2, it will not.