

Simulating Probabilities

Purpose: In this lab you will study some "laws" of probability by using the computer to simulate random experiments. You'll learn

- a) how expected outcomes differ from actual outcomes, and how this difference is affected by repeating the experiment a large number of times.
- b) how to calculate "experimental" probabilities
- c) how the Central Limit Theorem behaves with coin-tossing experiments

Data: There are no data for this assignment. You will generate your own data.

Assignment: There is no "take home" section in this lab. You should instead turn in your answers to all of the questions below. You do not need to include graphs in your answers.

Before Class

You should be familiar with these terms:

Probability, event, expected value, random

You should bring paper on which to record your observations.

Before we begin...

You've learned some of Stata's computational abilities, but Stata also allows users to write fairly sophisticated programs. We're not going to ask you to do this, but instead you're going to use some programs that others have written in order to run simulations. These programs are not sold with Stata, but you can still install them. If you are working in the lab, you probably won't have to do this. If you are working from home, you will need to add these programs before you can complete the exercises. Instructions for downloading these programs are given below.

Start up Stata and issue the command:

heads, save

If the program "heads" is installed on your computer, a graph should appear. *This is a graph of the cumulative (total) proportion of tosses that lands "heads" for a fair coin, if you were to toss it 100 times.*

If "heads" is not working on your computer and you are either in the Statistics Lab or at home, follow these instructions:

1. Click on the word "HELP" on the menu bar
2. Click on the line that reads "STB and User-Written Programs"

3. Click on the colored letters which read: <http://www.stata.com>
4. Click on the colored letters which read users
5. Click on the colored letters which read ucla
6. Click on the colored letters which read teach
7. Click on the colored letters which read heads (you will need to scroll down a few pages to find it, it's in the far left hand column of words) Later, we will also download "dice" and "cards". You might want to do this as soon as you finish installing "heads".
8. Click on the colored letters which read (click here to install)
9. When Stata is finished installing, it will read something like "click here to return to the previous page", at this point, close the help window and return to your Stata commands.
10. DO NOT TRY THE PREVIOUS 8 INSTRUCTIONS IN THE CLICC LABS IT WILL NOT WORK.

If you did have the heads program installed, you're ready to proceed.

Useful Stata Commands

generate *newvar expression* -> the **generate** command creates a new variable with the name *newvar* based on the values of any *expression*.

display *expression* -> allows you to use Stata as a calculator by displaying the result of any expression. For example, **display 3+5** will show 8.

Binomial(n,k,p) -> when used with the display command, will calculate the probability of getting exactly *k* heads in *n* tosses of a coin that comes up heads with probability *p*.

tabulate *variable* -> shows values, frequencies, and cumulative frequencies of *variable*.

Activity 1: Expected Values

Although individual experiments vary and seldom come out exactly as expected, there is still long-term regularity to random phenomenon. This means that "in the long run" our expectations tend to be met. (Of course, it's important to expect the right outcome!)

1. Issue the command:

heads, save

It may not feel like it, but you have just tossed a fair coin (probability of heads = 1/2) 100 times. A dialog box has appeared and four variables have been created.

The dialog box lets you change the number of trials for your experiment, the probability of getting heads on each flip, and the number of coins you want to flip at each trial.

To interact with the data, you must first click on "**quit**" in the dialog box. Do that now.

The graph that appears "pretends" that after each toss of the coin, you recorded the percent of the tosses that were heads.

(a) Question: If I toss a coin 100 times, how many heads do I expect or what percentage do I expect to be heads?

Hopefully you wrote 50 or 50%. Now issue the command:

tabulate heads

For the variable *heads*, 1=heads 0=tails. Note how many heads you got (you will not get the same as your classmates). This value will be displayed on the graph where "Number of Trials" is 100.

(b) Question: How many heads did you get after 50 tosses? 75 tosses?

(c) After 100 tosses, how far off were you from what you expected (50)?

2. Toss the coin 100 times again. Repeat this 10 times. After each, record the number of heads, and how far off you were from the expected value of 50.

3. Toss the coin 10 times. Repeat 10 times. After each, record how far off you were from the expected value.

Hint: There's a faster way of repeating this experiment 10 times. Tossing a coin 10 times is the same thing as tossing 10 coins once. So to toss a coin 10 times, you could also set "number of coins" to 10 and "number of trials" to 1.

This means that if you want to toss a coin 10 times and then repeat this experiment 15 times, then you could set "number of coins" to 10 and "number of trials" to 15.

After you try this, type **list trials heads** to get a sense of how this works.

4.

Simulation v2.0

a) Which experiment had the biggest errors, the 100-tosses or the 10-tosses? (The "errors" are the differences between what you saw and what you expected to see.)

b) Now which experiment had the biggest percent-errors? The percent error is the error divided by the number of tosses. For example, if I toss a coin 10 times, and get 7 heads, the error is $7-5 = 2$. The percent error is $2/10 = .2$ -- an error of 20%.

5. Next we'll toss the coin 1000 times. You should expect 500 heads. BEFORE you do this, answer these questions:

a) Will the error generally be more or less than when you did 100 tosses? Why?

b) Will the percent error generally be more or less than when you did 100 tosses? Why?

Now toss 1000 coins, and repeat 10 times. Each time, record the number of heads, the error, and the percent error.

You'll see that at each "trial", 1000 fair coins were tossed, and you'll see (in the "heads" column) the number of heads that resulted. This was repeated 10 times.

At each trial we expected 500 heads. The error can be found by typing:

generate error = heads - 500

And the percent error can be found by typing:

generate pcterror = error/1000

Now you can display or graph or summarize the error and pct error.

6. Of the three experiments (10 tosses, 100 tosses, and 1000 tosses), which had the biggest errors and which had the biggest percent errors?

Activity 2: Experimental Probabilities

Simulations can be used to estimate theoretical probabilities which, after all, are long-run frequencies.

The theoretical probability of getting exactly x heads in 10 tosses of a fair coin is

x	P(get exactly x heads)
0	.0010
1	.0098
2	.0439
3	.1172
4	.2051
5	.2461
6	.2051
7	.1172
8	.0439
9	.0098
10	.0010

(The back of your text book has tables for similar theoretical probabilities. These can also be calculated by a well-known formula called the binomial probability distribution.)

According to this table, the theoretical probability of getting 6 heads in 10 tosses is 20.51%. This means that if you repeat this experiment many times, in about 20.51% of them, you'll get 6 heads.

1. Toss 10 fair coins 10 times. (Set: "number of coins" to 10, set "trials" to 100). You can use the list command to see the number of heads at each of the 100 trials. And you can use the graph command to see a histogram.

Type

tabulate heads

to see the outcomes and the frequency with which they occurred. What is your empirical probability of getting exactly 0 heads? 1 head? 2 heads? (Note that many of you will get different empirical probabilities.) Compare your table of empirical probabilities to the theoretical table above. How do they compare.

2. The gap between the theoretical probabilities and your empirical probability will narrow as the number of repetitions on which the empirical probability is based increases.

a) Toss 10 fair coins 100 times. Compare the empirical probability table with the theoretical table.

b) Toss 10 fair coins 1000 times. Compare the empirical probability table to the theoretical probability table. Of the three tables you've generated, which is closest to the theoretical? Which outcomes had their empirical probabilities change the most? Which were the most stable?

3. We can do the same experiment for "unfair" coins. When you type heads, save

the dialog box gives you the opportunity to generate outcomes from a coin in which the probability of heads is not 1/2.

a) Calculate the theoretical probability of getting exactly 5 heads in 10 tosses when the probability of a head is 0.20. To do this, type

display Binomial(10,5,.2)

b) Use simulations to find an empirical probability for the probability of getting exactly 5 heads in 10 tosses of an unfair coin in which the probability of heads is 0.20. How many repetitions of this experiment did you use for your empirical probability?

Activity 3: Central Limit Theorem

In the last section you learned how to find experimental probabilities for a coin-toss experiment. We'll use this knowledge to discover an interesting property about averages.

Throughout all of these examples, we'll use an "unfair" coin with probability of heads equal to 0.20.

1. Toss the unfair coin 10 times. Repeat this experiment 100 times. Make a histogram of the number of heads, and notice that the experimental distribution is somewhat skewed to the right. Because we repeated the experiment 100 times, the experimental distribution is a fairly good approximation of the theoretical distribution.

Reminder: You can do this most efficiently by setting "number of tosses" to 100 and "number of trials" to 10. Make sure "P(head)" is set to 0.20.

Note that we should expect about 2 heads, but of course we don't get exactly two in each trial.

2. Now toss the unfair coin 20 times. Repeat this experiment 100 times. Is the histogram more or less skewed. How many heads should you expect to see?

3. Toss the unfair coin 100 times, and repeat 100 times. Make a histogram. (Note: you'll get a better picture if you set the number of bins to 10: **graph heads, bin(10)**)

What's the expected value? Is the histogram centered close to this value? How would you describe the histogram.

4. Repeat the last question, but toss the unfair coin 1000 times.

Not only is the histogram getting more symmetric as the number of coins tossed increases, but in fact the distribution is getting closer to a normal distribution.

How can we see this? One tool that statisticians use to judge the "normality" of a set of a data is called the q-q plot. This plots plots the observed values in the sample against what the observed values should have been if, in fact, the sample had come from a population whose values followed the normal distribution.

The command

qnorm x

creates such a plot for variable x. If the distribution is close to normal, the points will lie close to the line that Stata superimposes on the graph.

5. Toss the unfair coin 10 times, repeating this experiment 100 times. Create the qnorm plot. Now toss the unfair coin 100 times, repeating 100 times. Compare this qnorm plot to the previous one. What's the difference? Now toss the unfair coin 1000 times and compare to the previous qnorm plots.

Activity 4: An Application

One measure of the success of a portfolio manager is the "monkey and dartboard measure." Here's a greatly simplified explanation. Portfolio managers put together combinations of stocks and bonds which their clients buy as a package. The idea is to make more money than you would by saving your money in a bank. As a client, you hope the portfolio manager knows better than you which stocks will do well. And you certainly hope that he could do better than a monkey could, if the monkey threw darts at a wall that had the names of stocks written on it. In other words, suppose we chose 10 stocks at random. Would they outperform the 10 stocks the portfolio manager chose?

Let's ignore the amount that stocks change, and just classify all stocks as either "up" or "down". In a good market, at the end of 3 months suppose that 60% of all stocks have gone up. Now suppose you selected 10 stocks at random. A portfolio manager claims to have used his expertise to select 10 stocks very carefully. And 7 of his 10 stocks are "up".

Using simulations, decide whether you think the portfolio manager's performance is very good. Approximately what's the probability that 10 randomly selected stocks would do as well or better than the portfolio manager's?