Simulations as a Tool for Reasoning about Probability and Statistics
NCTM 2012

Hollylynne Lee	Tina Starling	Tyler Pulis
hollylynne@ncsu.edu	tinatstarling@yahoo.com	hpulis@ncsu.edu

North Carolina State University, Raleigh, NC

During this session, we will build probability simulations for the following tasks:

1. ***Most larger families have either more boys or more girls.***
   We will simplify the problem by looking at a family with four children and different ways to model this situation. Then, we will generate large sets of families of four and use graphical representations to visualize the frequency distributions as a means of making sense of whether "most families" have more boys or more girls in a large population. This approach can develop a view of probabilities as indicators of expectations over many trials

   Lee, H. S., & Lee, J. T. (2011). Simulations as a path for making sense of probability. In T. Dick & K. Hollebrands (Eds). Focus in High School Mathematics: Technology to Support Reasoning and Sense Making (pp. 69-88). Reston, VA: National Council of Teachers of Mathematics.

2. *Counties keep records of the number of live births that occur throughout the year. Consider the following data about live births recorded in two North Carolina counties in 2004:*
   A) Hyde County recorded 56 births and slightly less than 43% were male.
   B) Martin County recorded 314 births and slightly less than 43% were male.
   Data source: http://www.schs.state.nc.us/SCHS/data/births/bd.cfm
   Assuming that the probability of a male birth is 50%, is event A or B more likely to occur, or are the two events equally likely? Explain your reasoning.

   Lee, H. S., Hollebrands, K. F., & Wilson, P. H. (2010). *Preparing to Teach Mathematics with Technology: An Integrated Approach to Data Analysis and Probability (1st Edition).* Kendall Hunt Publishing.

The following pages include technology directions for conducting simulations with the graphing calculator, Microsoft *Excel* or GoogleDocs, *Fathom*, and *TinkerPlots*

Simulations as a Tool for Reasoning about Probability and Statistics
Hollylynne Lee, Tyler Pulis, and Tina Starling   NCTM 2012


TI-83/84 GRAPHING CALCULATOR

Using the randInt function on the TI-83/84
The random integer function, randInt(a, b, n), in which a is the first integer, b is the last integer, and n is the number of trials, produces n integers between a and b, inclusive. Directions for using randInt are below:

1. Press the MATH key.
2. Choose the PRB menu.
3. Select the randInt option (see the screenshot below).
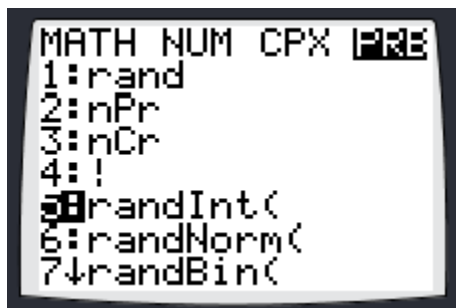
4. Enter the starting number, ending number, and number of integers to be generated (separated by comma). Suppose you want to select an integer that is either 0 (for a girl birth) or 1 (for a boy birth), ten times. The following screenshot shows the syntax for using the randInt function for this scenario as well as the results that appear once you press Enter. You will need to use the right arrow key to view all 10 outcomes.
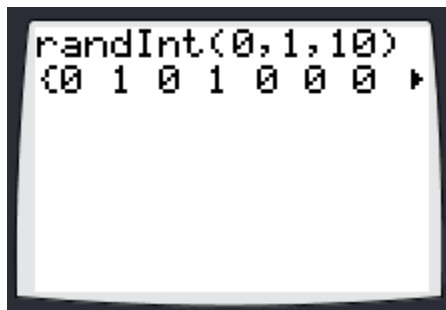

5. To view a histogram of results, first store the results in L1 by pressing STO and typing 2nd 1. Pressing ENTER will display the results again. Then, press 2nd y= to access the StatPlot menu. Select On to activate Plot1. Select the Histogram icon for Type of plot (see screenshot below). Once the histogram option is selected, the Xlist will display L_1 and Freq will display 1 by default. This means that data for the plot is stored in L_1 and each value is counted once. Since our data is stored in L_1, we do not have to change the Xlist.
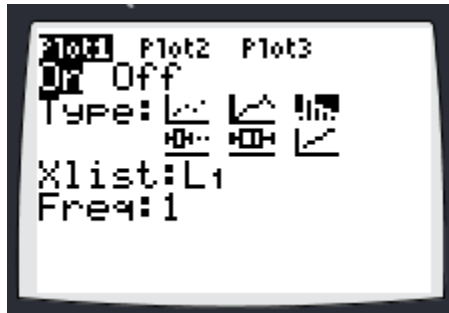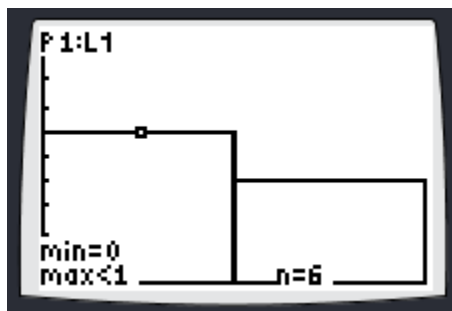
6. Press the WINDOW key and change the settings (Xmin=0, Xmax=2).
7. Press the GRAPH key to display the new histogram. Press the TRACE key (and right/left arrow keys) to show the frequencies of outcomes (see example below).
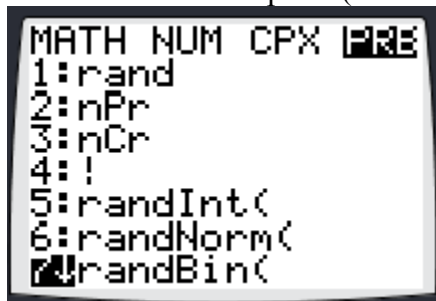

Using the randBin function on the TI-83/84
The random binomial function, randBin(*sample size, probability of success, number of samples*), generates the specified number of random integers from a specified binomial distribution. Directions for using randBin are below:

1. Press the MATH key.
2. Choose the PRB menu.
3. Select the randBin option (see the screenshot below).

4. Now, suppose you want to simulate 56 births (to represent Hyde county) with a 0.51 probability of male birth. To collect 10 samples and store the sample proportions, enter randBin(56, 0.51, 10)/56. Then press enter. You will need to use the right arrow key to view all 10 outcomes.

5. Store the results in L1 by pressing STO and typing 2nd 1. The screenshot below shows an example of the first seven sample statistics. Use the down arrow key to view all 10.
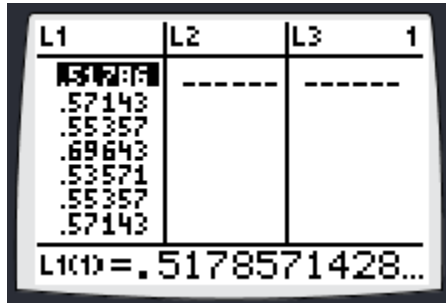
6. To view a histogram, you will likely need to change your window settings. Press WINDOW and set your bin width by changing xScl. One possible window configuration and the resulting histogram are shown below.
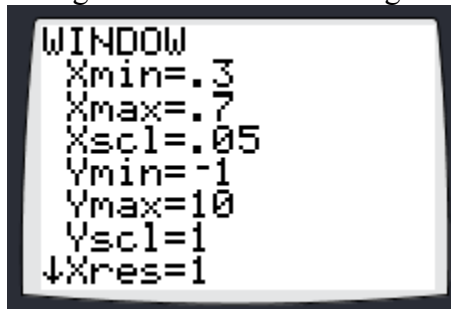
MICROSOFT *EXCEL* OR GOOGLEDOCS

Always remember to enter an = before any function in a spreadsheet!

Using the randBetween function:
The randBetween(*a,b*) function generates a single integer between *a* and *b*, inclusive. To generate more than one, simply grab the bottom right corner of the cell with the formula and drag down to fill in as many cells as you wish.

Using the countIf function:
The countif(*cellrange,criteria*) function will count the number of cells in a given range meeting a certain criteria. For example, if you have generated random integers between 1 and 100 in cells A1 to A56 using =randBetween(1,100), to model a probabilistic situation with two outcomes having probability 51% and 49%, use =countif(A1:A56, "<52") to count the number of outcomes in that range that are numbers from 1 to 51 inclusive. (You could also use "<=51".)

If desired, you can also assign Male to the numbers from 1 to 51 and Female to numbers 52 to 100 using an **IF(logical_test,value_if_true,**value_if_false**)** statement for each cell, for example, =if(A1<52, "Male", "Female") You can then use =countif(A1:A56,"Male") to count the number of Males.

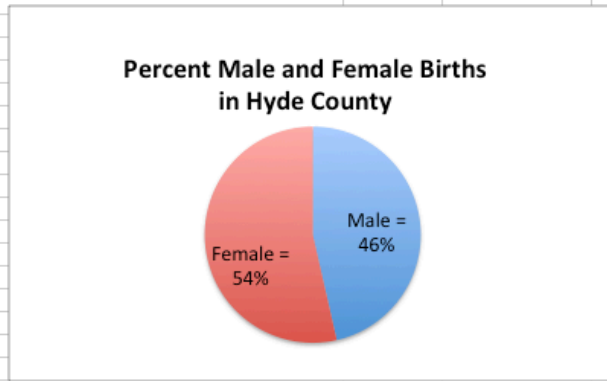# Simulations as a Tool for Reasoning about Probability and Statistics
Hollylynne Lee, Tyler Pulis, and Tina Starling   NCTM 2012

Note: To rerandomize on a PC, press the F9 key. On a Mac, press Command + = , and on GoogleDocs, press Ctrl + R.
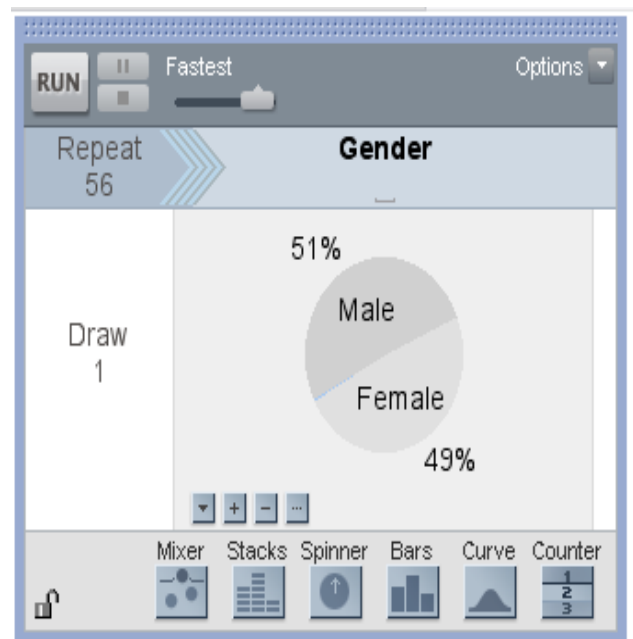
	A	B	C	F	G	H	I	J
2	10	Male	Let 1-51 represent a Male birth.	Male =		26	Percent Male =	46.4285714	
3	31	Male	Let 52-100 represent a Female birth.	Female =		30	Percent Female =	53.5714286	
4	32	Male		Total =		56			
5	88	Female							
6	88	Female							
7	71	Female							
8	89	Female							
9	75	Female							
10	31	Male							
11	70	Female							
12	94	Female							
13	56	Female							
14	90	Female							
15	30	Male							
16	27	Male							
17	28	Male							
18	34	Male							
19	80	Female							
20	23	Male							
21	25	Male							
22	76	Female							
23	80	Female	To show functions:	To rerandomize:					
24	91	Female	Control" + '~'	Mac:	Commmand' + '='				
25	1	Male		PC :	F9				
26	76	Female							
27	91	Female							

Percent Male and Female Births in Hyde County

Male = 46%
Female = 54%

TINKERPLOTS

Stage 1: Setting up the Simulation for One Sample of 56 births
1. Pull down a Sampler from the toolbar.
2. For our example of births, we used a spinner to represent the two outcomes, male and female. Drag the spinner icon into the Sampler window.
3. Label the spinner Male and Female.
4. Enter 1 for the number of Draws.
5. Name the attribute Gender.
6. Display the percentages for the spinner by selecting the small down arrow tab under the spinner and then choosing Show Percent.
7. Adjust the percent by grabbing and dragging part of the spinner or by clicking on the percent and typing in 51 for male.
8. Type 56 for the number of Runs.
9. Click Run. (A new table should appear called Results of Sampler 1 with the attribute Gender and 56 results labeled Male or Female.)
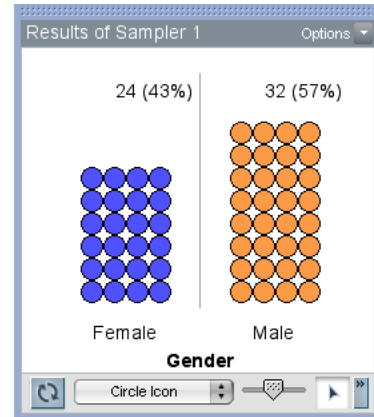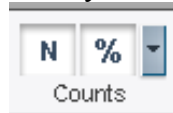
RUN   Fastest   Options ▾
Repeat   **Gender**
56
                    51%
                        Male
Draw
1   Female
                            49%

Mixer Stacks Spinner Bars Curve Counter

Stage 2: Plot the Outcome of the Simulation
1. Pull down a plot window.
2. Drag the Gender attribute from your collection to the x-axis.
3. Display the count and percentage for the two bins displaying Female and Male by selecting the Counts on the toolbar.

Stage 3: Collecting the Percent Statistic for your Sample
1. Right-click on the percentage displayed in the Male bin and select Collect Statistic. (A new table should appear called History of Results of Sampler 1.)
2. Pull down a plot and drag the percent_Gender_male attribute to the x-axis.
3. Click RUN on the Sampler to generate new samples and collect the sample's percent male statistic. As you generate more points on your percentage plot, select a point and drag it to the right to create a continuous number line on the x-axis. You can also stack the dots to make a neater graph. One plot displaying the percent male statistic for 30 samples is shown below.
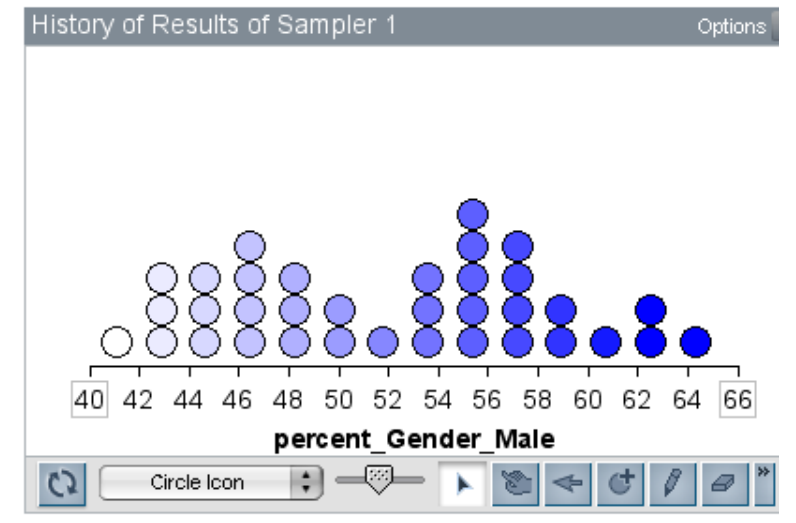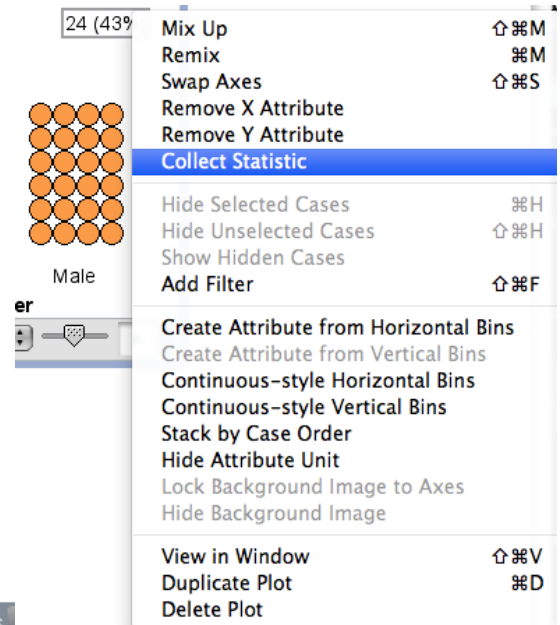

FATHOM

---

1. Open a new collection in *Fathom*.
2. Double click on the collection to view the Inspect Collection window. Name the attribute Birth.
3. Then, click on Edit and choose Edit Formula (or you may double click the Formula to add or edit a formula). Type randomBinomial(1,0.51) to generate either a value of 0 or 1, with 1 having a probability of 0.51.
4. Notice that, at this point, there are no cases. To collect a sample of 56 births (to simulate Hyde county), right click on the case arrows at the bottom left corner of the Inspect Collection window. Select New Cases. Type 56 to add 56 cases to this collection. Then click OK.
5. Repeat these steps with 314 cases to create a separate collection for Martin county within the same *Fathom* document.
    *Note, the *Fathom* file, NCCountyBirthSim.ftm contains two collections, one for each of Hyde and Martin counties.
6. Create a histogram to view the frequencies of 0s and 1s in each collection. To do this, drag down an empty graph from the object shelf. Drag the Birth attribute from a collection to the y-axis. Then, use the drop-down menu in the graph to change the dot plot to histogram. Adjust the scales on the x- and y-axis accordingly.
7. Press Ctrl+Y (on a PC) or (Command+Y) on a Mac to rerandomize and create new samples of Hyde and Martin counties.
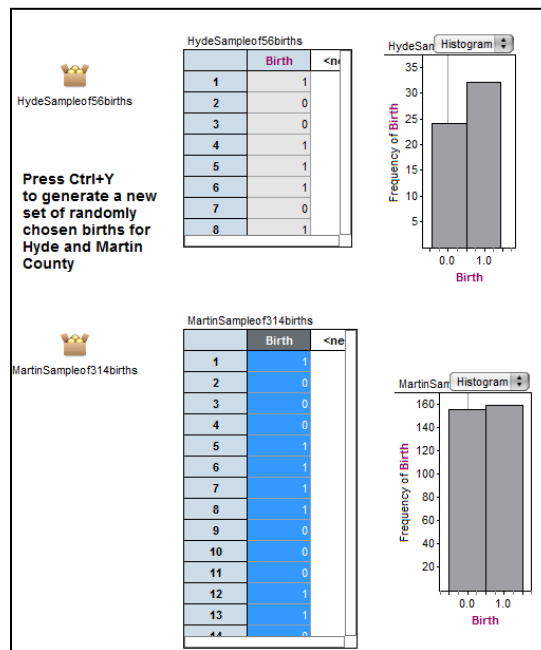
8. For each of the two collections in the NCCountyBirthSim file, a measure has been defined to compute the proportion of males. Click on a collection (e.g. HydeSampleof56Births). Under the Object menu, choose Inspect Collection. The inspect window will appear (see screenshot below). The Cases tab will be selected by default.
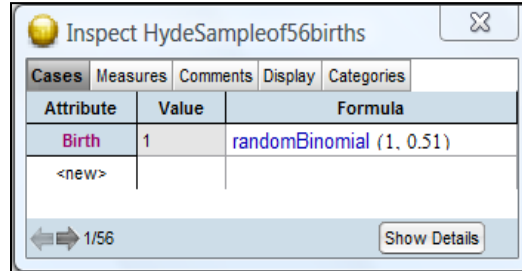
9. Click on the Measures tab. Notice the measure "ProportionofBoys" has been defined with a formula (see screenshot below). You could also double-click on <new> and create and define a new measure of interest. Each time the collection is rerandomized (Ctrl+Y), the measures will be recomputed based on the new set of data.
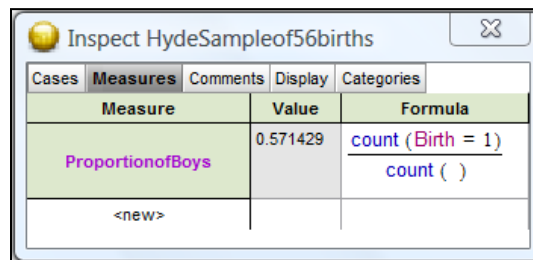
10. To conduct repeated samples and automatically create a new collection containing the measures (ProportionofBoys) for each sample, begin by clicking on a collection to select it. Under the collection menu, choose collect measures. A new collection, named "Measures from <Name of Collection>," will appear and will already contain 5 cases (see screenshots below).
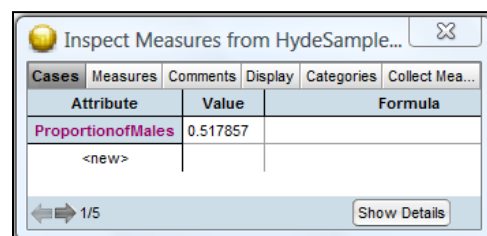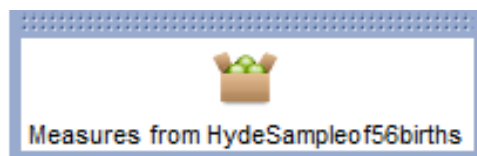

11. To collect additional measures, click on the last tab labeled Collect Measures in the Inspect window. Adjust the properties to reflect how many measures you wish to collect, then click on the Collect More Measures. For a large number of measures, it may be helpful to unselect the Animation On box (see screenshot below).

Simulations as a Tool for Reasoning about Probability and Statistics
Hollylynne Lee, Tyler Pulis, and Tina Starling   NCTM 2012
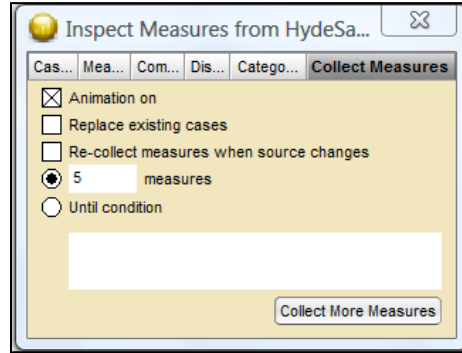
12. Create a dot plot with the attribute ProportionofMales on the x-axis. Leave the dot plot open, return to the Inspect window of the Collection of Measures. On the Collect Measures tab, leave Animation on, change the number of measures to 95 (add to the existing 5 for total of 100), and press Collect More Measures.
13. To repeatedly collect measures from *n* samples, open the Inspect window for a collection of measures. Deselect the Animation option, and select the option to Replace existing cases. Enter the number *n* of samples to collect and click on Collect More Measures. Repeat this process to collect measures from *n* new samples.

The figure below shows the dotplots produced in Fathom by applying the above process and collecting 100 samples of birth data for both Hyde County and Martin County.
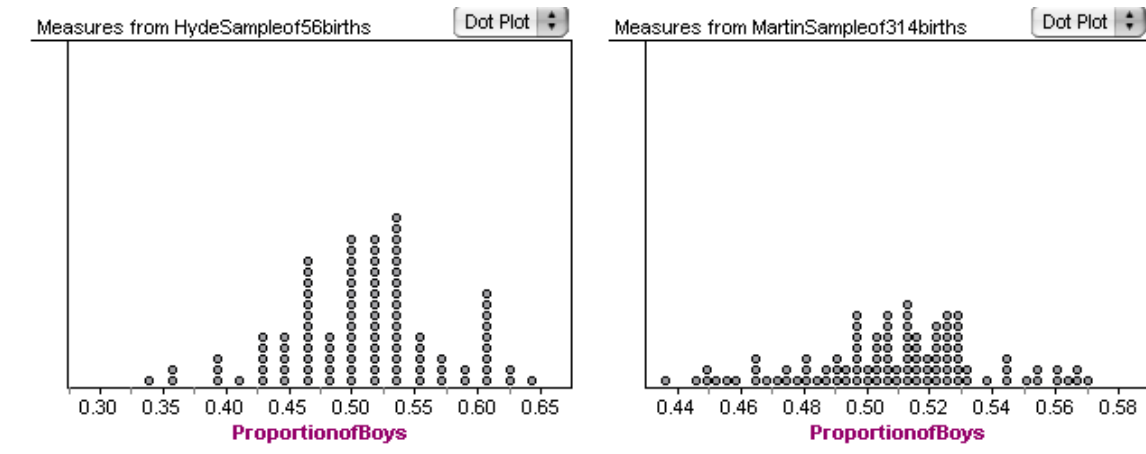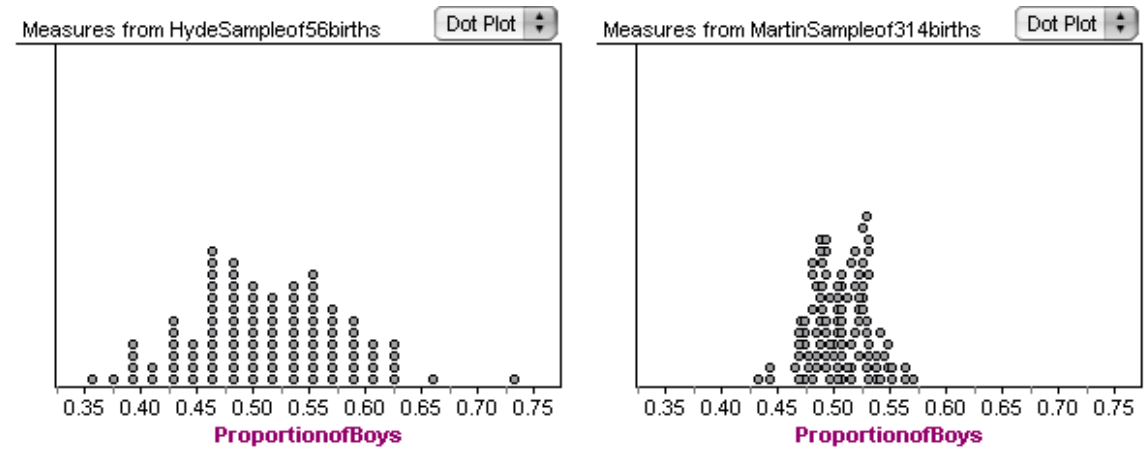

Note: To compare the two distributions, it is helpful to adjust the scales on the axes as shown below. To adjust the scale, double-click on the scale and type in values for xLower and xUpper.

Simulations as a Tool for Reasoning about Probability and Statistics
Hollylynne Lee, Tyler Pulis, and Tina Starling   NCTM 2012

You can access the technology files used in this presentation using the following links:

Task 1:
http://dl.dropbox.com/u/39398075/NCTM12files/fourFamilies.xls

http://dl.dropbox.com/u/39398075/NCTM12files/FathomFamilies.ftm

Task 2:
http://dl.dropbox.com/u/39398075/NCTM12files/2004_NCBirthData_Lat_Long.ftm

http://dl.dropbox.com/u/39398075/NCTM12files/NC_BirthData_2004_Long_Lat.tp

http://dl.dropbox.com/u/39398075/NCTM12files/HydeCounty_Excel.xlsx

http://dl.dropbox.com/u/39398075/NCTM12files/NCCountyBirthSim.ftm

http://dl.dropbox.com/u/39398075/NCTM12files/BirthSimulator.tp