

Monte Carlo Modeling  
MAT 135/223, Spring 2016

**Instructions:** Student will work in small groups of 2-4 on this project. Groups should work together and not race ahead of other members of the group who are slower. Make sure everyone in the group is at the same place before moving on to the next problem. **When you reach a point where you are asked to consult with your instructor, please do so, even if your instructor is moving around the room.** These points are designed for us to have a conversation, to make sure everyone is on the same page before moving on. We will work on this project for several days in class, so be sure to bring it back with you the next day until the project is completed.

Choose a packet of dice from those provided by your instructor. Dice come in many varieties, some of which are shown below.



Tetrahedral (D4, four-sided) dice



Standard cubic (D6, six-sided) dice



Octohedral (D8, eight-sided) dice



Decahedral (D10, ten-sided) dice; come in two varieties



Dodecahedral (D12, twelve-sided) dice



Icosahedral (D20, twenty-sided) dice



Tetraicosahedron (D24, twenty-four-sided) dice



Tricontahedron (D30, thirty-sided) dice



Hexacontahedron (D60, sixty-sided) dice



Hecatohedron (D100, one-hundred-sided) dice



# of dice rolled	# of successes in the rule

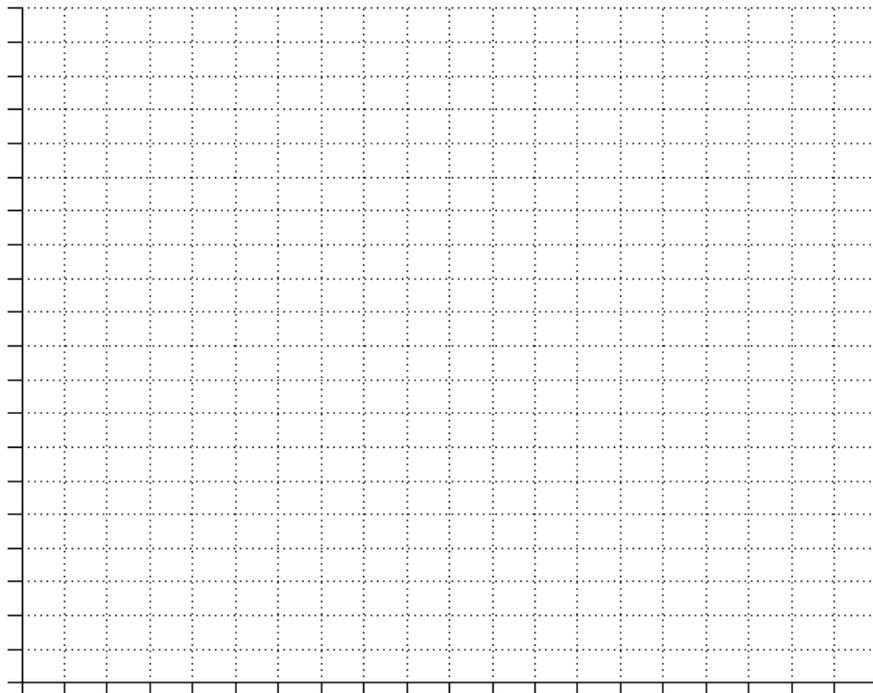
# of dice rolled	# of successes in the rule

# of dice rolled	# of successes in the rule

- E. Now, increase the number of dice you are rolling and repeat the experiment five more times with that number of dice. Keep doing this until you have filled the table(s) with data. (You should end up with six sets of 5.) Your last data point should use all the dice you have available.

***Check with instructor.***

- F. Enter the data into your calculator with the first column (total number of dice) as  $L_1$ , and the second column (# of successes) as  $L_2$ . Use the calculator to create a scatterplot and reproduce the graph below. Label the axes appropriately.



- G. Choose two points from your data set or the graph and draw a line between them. (We will be doing this more than once. Your instructor has colored pencils so that you can tell them apart.) Use the point-slope equation of a line to find the equation of the line that connects the points. Rewrite the equation in slope-intercept form. Recall:  $y - y_1 = m(x - x_1)$        $m = \frac{y_2 - y_1}{x_2 - x_1}$

Enter the equation in to  $Y_1$  on your calculator and add the line to your scatterplot. How well does the equation appear to fit the data in the graph?

- H. Repeat the above step with a different pair of points. What is the equation you came up with? Enter your equation in  $Y_2$  in your calculator and add it to your scatterplot. (Change the color of the line, and label each clearly.)
- I. Did this line do any better approximating the data in your scatterplot? Or is it worse? Discuss with your group some strategies for choosing points that might lead to a better fit for the data. Record your ideas below.

***Check with your instructor.***

- J. We are now going to use the calculator to construct a *linear regression* equation. This line is sometimes called the “least-squares (regression) line” or the “line of best-fit”. This line minimizes the sum of the squared distances between the points on the graph, and the equation of the line at each value of  $x$ . Deriving the formula from basic principles requires calculus, but

the calculator can use the formula derived from that process to compute the line for us. Follow the calculator steps below to compute the linear regression line.

1. Turn on Diagnostics:

a. On newer TI-84s, select **MODE** and scroll down toward the bottom on the list of commands (onto the second page). You will see a line for the Diagnostics. Select ON to be highlighted.

b. On older machines (TI-83s, and some older 84s), press **2nd** **0** for **CATALOG**. Scroll down through the alphabetical list to the line that reads Diagnostics

On. Select **ENTER** when you get to that line and then press **ENTER** again to execute the command.

(You should only have to do this once unless someone clears the memory.)

2. Select **STAT** and then scroll over the **CALC** menu. Scroll down to **4: LinReg (ax+b)**. This is the linear regression function. (8 is also a linear regression function with a+bx. It doesn't

matter which you choose, but don't switch back and forth.) Select **ENTER** or **4**. On newer TI-84s, your screen will look like this:

```

LinReg(ax+b)
Xlist: 1
Ylist: L2
FreqList:
Store RegEQ:
Calculate
    
```

Accept the defaults for the List locations unless you used a different list. Leave the frequency line blank. On the Store RegEQ line type:

**VAR** then select **Y-VARS**, then **FUNCTION**, then **Y3**. Then scroll down to Calculate and

press **ENTER**.

On older machines your screen will look like this:

```

LinReg(ax+b)
LinReg(ax+b) L1,L2,Y3
    
```

Before hitting anything else, follow the steps above to select **Y3** regression equation location. If you need to change the default lists, list them first separated by commas until your screen looks like this:

Then press **ENTER** to execute the command.

Your screen will look something like what is shown here, though, with different numbers.

The first two numbers are the slope and intercept for your line. The equation they belong to is shown at the top. The two bottom values show up only when the Diagnostic setting is ON. The closer these values are to 1 (in absolute terms), the better the fit to the data. ( $r$  is the correlation value: it can be between  $-1 \leq r \leq 1$ , having the same sign as the slope;  $r^2$  is the coefficient of

**LinReg**  
 $y=ax+b$   
 $a=.926504585$   
 $b=2.00913154$   
 $r^2=.9175188413$   
 $r=.9578720381$

determination: it measures the proportion of the variation explained by the regression line).

- K. After completing these steps for calculating the regression line, give the equation of the line below, along with the correlation value. Add your line to the graph above in a different color.
- L. Visually inspect the line relative to your data: does it appear to fit the data better than your previous lines? Why or why not?

***Check with your instructor.***

- M. Do you think it's reasonable to include the point (0,0) in your data set? If so, add the point to your lists and recalculate the regression line. Store this equation in Y4, and add it to your graph above. Does the point improve (or change) the line much? If not, explain why it should not be included.

***Check with your instructor.***

- N. What is the slope of (each of) your lines? For the line of best fit (either Y3 or Y4), how similar is this to the value of proportion you obtained in C? Which one is the closest? What does this tell you about the line?
- O. What do you think this value means in the context of the rolling dice and your rule for success?
- P. What is the y-intercept of your best-fit equation? What does this value mean in the context of the example? If the correlation was perfect, what should this value be?
- Q. Use your regression equation to predict the number of successes in your rule that would come up if you rolled 150 dice? What about 500 dice?
- R. Do you think your prediction is reasonable in each case? Why or why not?

***Check with your instructor.***



```

MATH NUM CMLPX PROB FRAC
1: rand
2: nPr
3: nCr
4: !
5: randInt(
6: randNorm(
7: randBin(

```

The syntax for this function is **randBin(# of trials, probability of success)**. You can add a third piece to the syntax for number of repetitions. Newer calculators will give you a “wizard” screen to remind you what each piece means.

Enter the number of trials, the proportion you found in C, and since we want 5 repetitions of each sample size, enter 5 as the last value. Then select paste. Your screen will appear as shown below.

```

randBin(500,.5,5)
.....{258 258 240 258 236}.....

```

```

randBin
n:500
P:.5
repetitions:5
Paste

```

The numbers shown are the number of successes the resulted from a simulation, similar to the results we

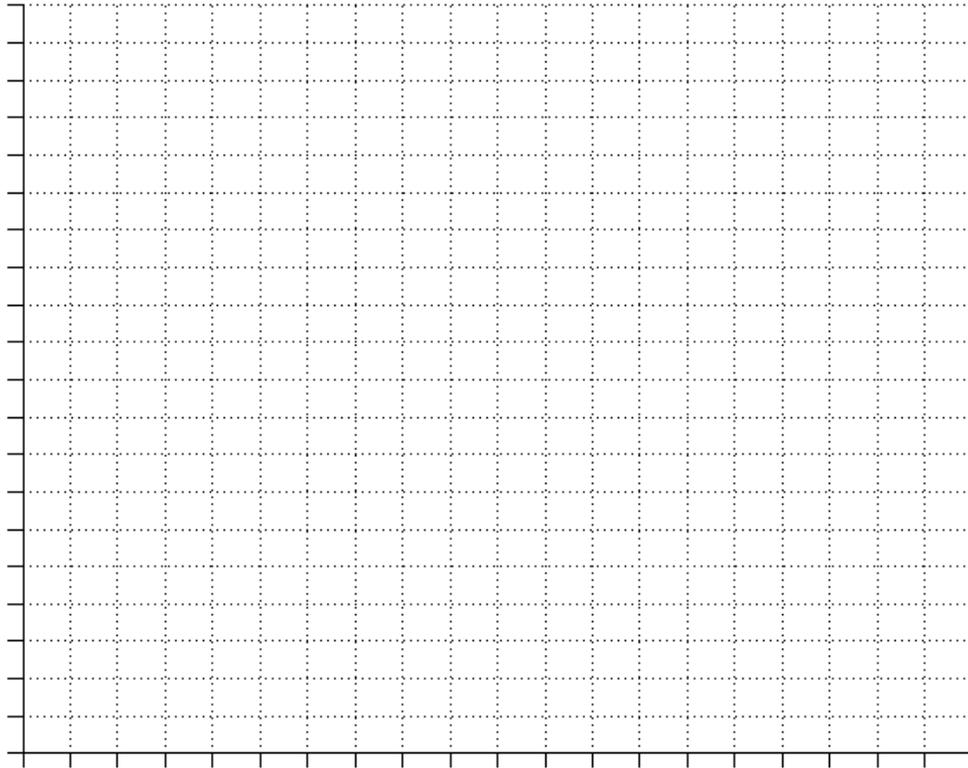
might obtain using the random number generator. Note: your calculator will take some time to generate the list of successes, and longer for larger sample sizes. Once you’ve generated your large samples, add them to the data tables.

- U. Add this new data to the end of your lists from the smaller sample sets. Rerun your regression equation with this new data and store it in Y5 and report the regression equation and correlation here. Extend your scatterplot in your calculator to see all the lines along with the new data points.
- V. Did the new slope get closer to the proportion you found in C? Did the diagnostics get close to 1 or further?

***Check with your instructor.***



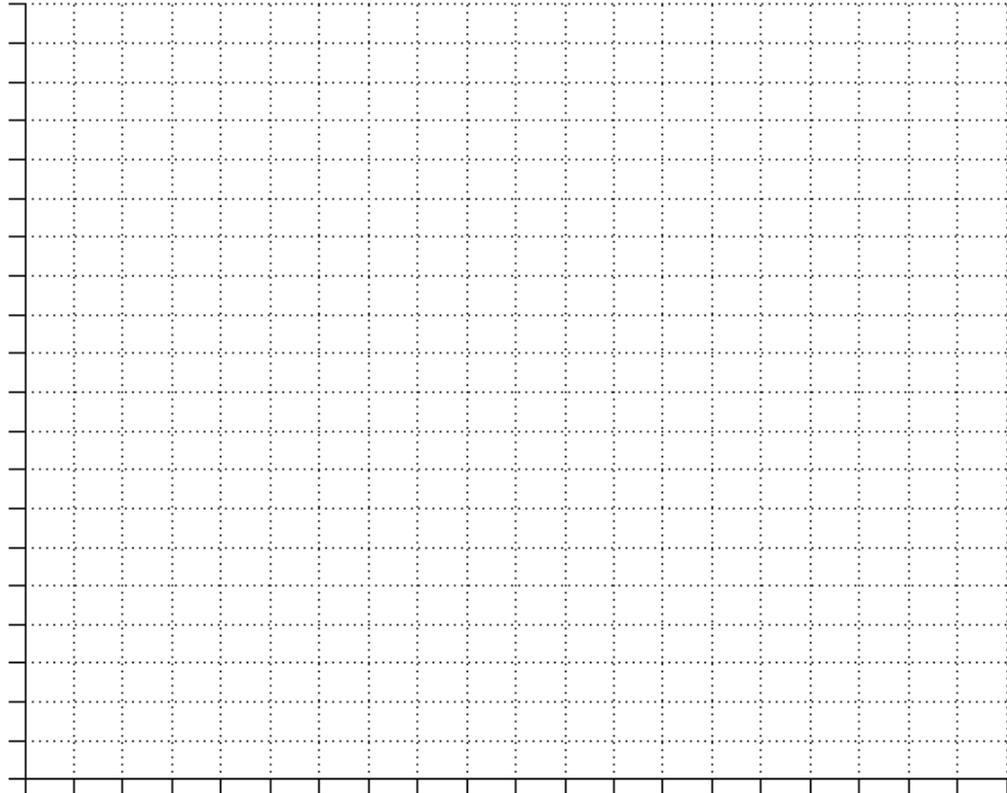
- X. Clear out all your equations from the **Y=** screen. Enter the proportion value from C into Y1. Then run your **LinReg** function on L<sub>1</sub> and L<sub>3</sub>. Use **ZOOM** **9** **ZoomStat** to adjust the window to look at the resulting scatterplot. Sketch the graph below. Label axes appropriately.



What do you notice about the line and the points? What is the correlation value now? Has it changed from our previous value when we were working with successes rather than proportions? What is the equation of the line? Record both here. (Note: Pay attention to the scientific notation!)

- Y. The Law of Large Numbers states that the larger the sample size, the closer the proportion of the successes a trial will have to the theoretical value. Explain how the graph above illustrates (or fails to illustrate) this law.





FF. The Central Limit theorem says that the spread of results of a sample decreases proportional to  $\frac{1}{\sqrt{n}}$ . (As a power, this is  $n^{-\frac{1}{2}}$ .) Does this agree with the results you obtained above? Why or why not?

***Check with your instructor.***