Monte Carlo Modeling
MAT 223, Fall 2015

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

Tetraicosahdehron (D24, twenty-four-sided) dice

Tricontahedron (D30, thirty-sided) dice

Hexacontahedron (D60, sixty-sided) dice

Hecatohedron (D100, one-hundred-sided) dice

This is far from an exhaustive list, but this list covers the most common ones.

A. Which kind of dice did you choose? How many sides does it have and how are the faces labeled?

Next we are going to roll small sets of these dice.

B. Choose a subset of those faces. After each roll, these are the sides of the dice we will count as a “success”. Try to make your rule (set) simple to follow. For instance, you might choose all the numbers divisible by 3, or all the odd numbered sides, etc. Write your rule out in words (or symbols), and then list the faces of the dice that this rule corresponds to. (Note: if you use the same rule on different dice, your lists of the faces will be different.)

Check with your instructor.

C. What proportion of the total does your rule correspond to? i.e. $\frac{\#\ of\ sides\ in\ your\ rule}{total\ \#\ of\ sides\ on\ each\ die}$

D. Begin with 5 dice. Roll the dice 5 times and after each roll record the number of faces come up that correspond to your rule (subset). Record them in the table.

# of dice rolled	# of successes in the rule

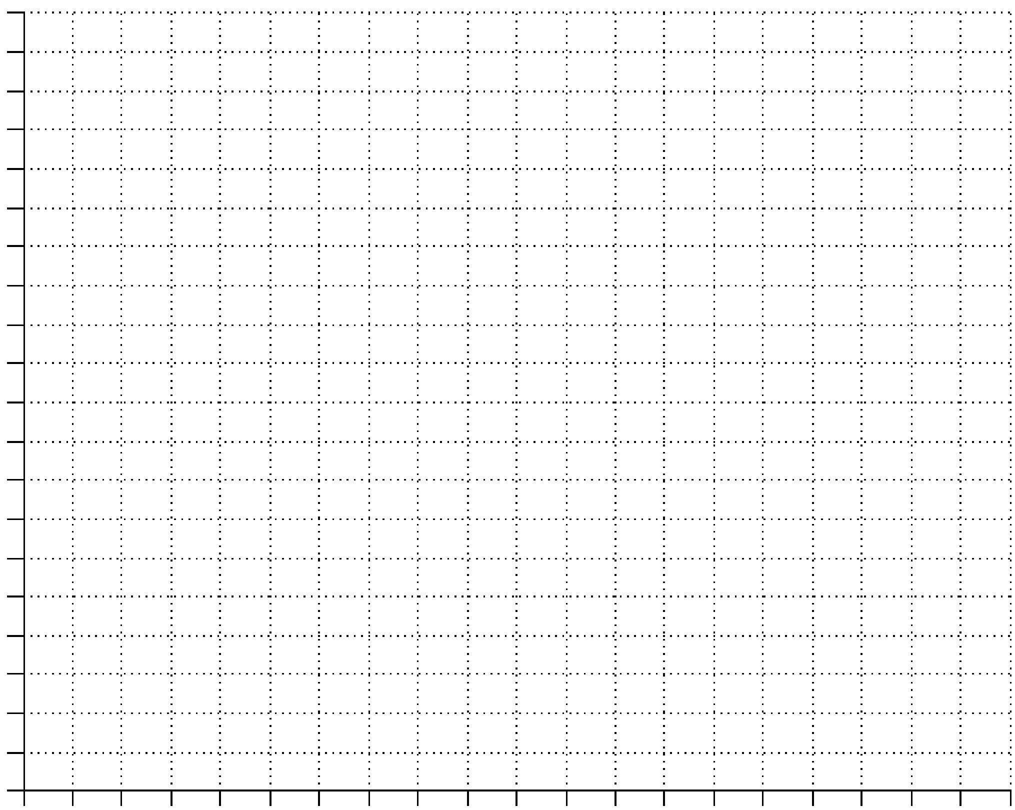
# of dice rolled	# of successes in the rule

# 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. 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.



G. Choose two points from your data set or the graph and draw a line between them. (You may want to use different coloured pencils to tell your lines apart.) Use the point-slope equation of a line to find the equation of the line the connects the points. Rewrite the equation in slope-intercept form.

Recall: $y - y_1 = m(x - x_1)$ $\quad m = \frac{y_2 - y_1}{x_2 - x_1}$

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. 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 minimized the sum of the squared distances between the points on the graph, and the equation of the line at each point. 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 line.

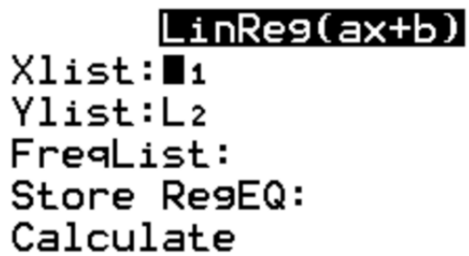
1. Turn on Diagnostics:

   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.

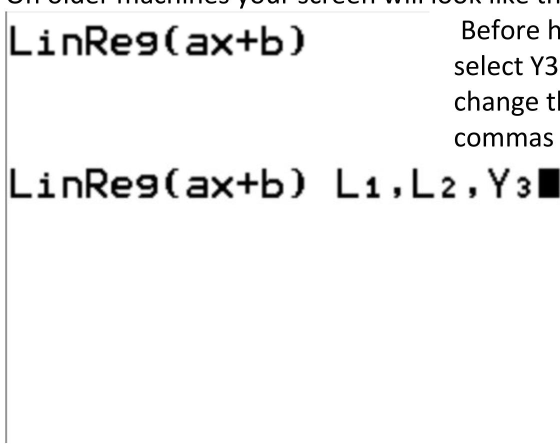
(You should only have to do this once unless someone clears the memory.)

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:



Accept the defaults for the List locations unless you used a different list. Leave the frequency line blank. On the Store RegEQ line type: VARS then select Y-VARS, then FUNCTION, then Y3. Then scroll down to Calculate and press ENTER.

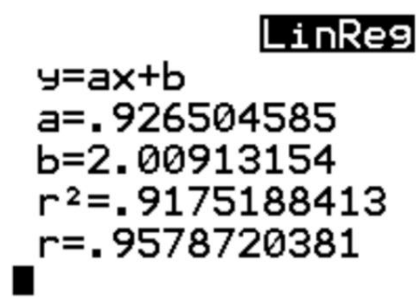
On older machines your screen will look like this:



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 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. Be sure to add it to the graph above (use a different colour than the others and label it clearly).

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?

O. What do you think this value means in the context of the example?

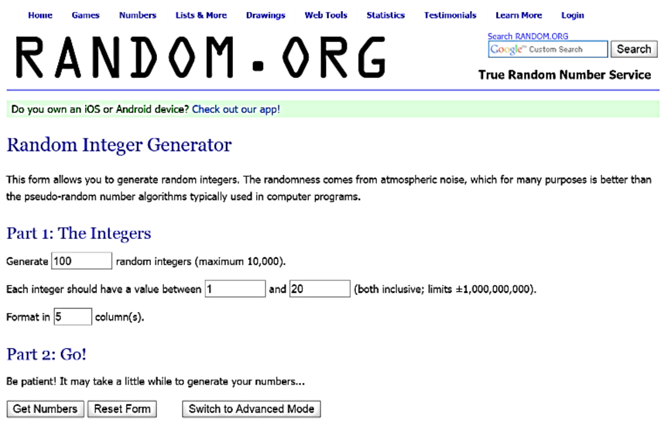
P. What is the y-intercept of your best-fit equation? What does this value mean in the context of the example?

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? Why or why not?

Check with your instructor.

S. We don't have that many dice to roll physically (nor would we really want to), but we can use technology to simulate the roll of many dice using a random number generator.
Access the website random.org. Scroll down to the *Numbers* section and select *Integer Generator*.



You will want several large sets of numbers (for instance: 100, 200, 500, 1000 each at least once, but preferably, 5 times each). Set your integer range to match the values on your dice. An example for an icosahedral dice is shown here. Set the number of random outcomes you want, and then select the Get Numbers button. For larger sets, increase the number of columns of the display for easier counting. Hint: you may want to generate these lists where a printer is available, and count the successes later.

Add the data you gather to the tables below.

# of dice rolled	# of successes in the rule

# of dice rolled	# of successes in the rule

# of dice rolled	# of successes in the rule

T. 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. Extend your scatterplot in your calculator to see all the lines along with the new data points.

U. 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.

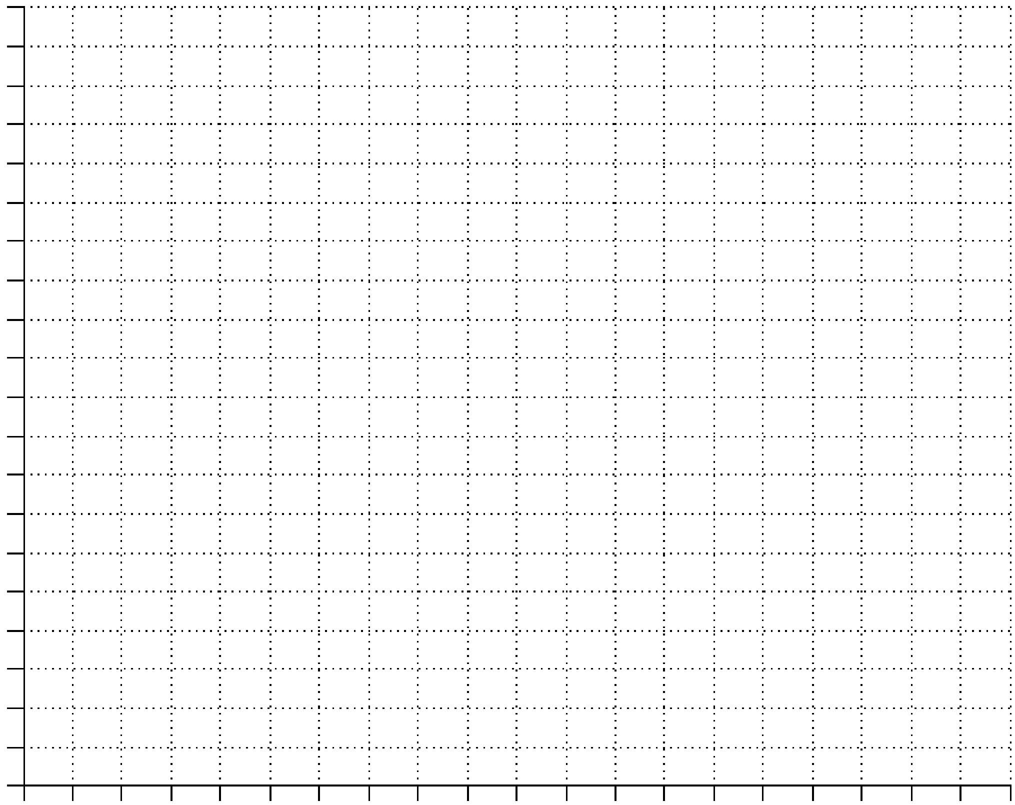
V. We are now going to transform our data set. Instead of the total dice rolled as input (x) and the number of successes as output (y), we are going to create a new variable which will be the ratio of successes to the total on each toss. Before doing this, you will need to remove the point (0,0) from the data set. We are going to store this new variable in L_3. To calculate this quickly in the TI-83/84:

Select [2nd] [2] (for L_2) then [÷] [2nd] [1] (for L_1) then [STO▶] [2nd] [3] (which will store the result in L_3). Your screen will look like this:

$L_2/L_1 \rightarrow L_3$

Then press [ENTER] to execute.

W. Clear out all your equations from the [Y=] screen. Enter the proportion value from C into Y1. Then run your LinReg function on L_1 and L_3. Use [ZOOM] [9] ZoomStat to adjust the window to look at the resulting scatterplot. Sketch the graph below.



What do you notice about the line and the points? What is the correlation value now? What is the equation of the line?

X. 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.

Check with your instructor.

Y. Now we are going to look at groups of the data, trials done with the same sample size. For each set of 5 samples taken at the same sample size, find the range of the proportion. Estimate the standard deviation by dividing it by 4. Complete the table below with the data using the sample size (number of total dice tossed), and the estimate of the standard deviation at that sample size.

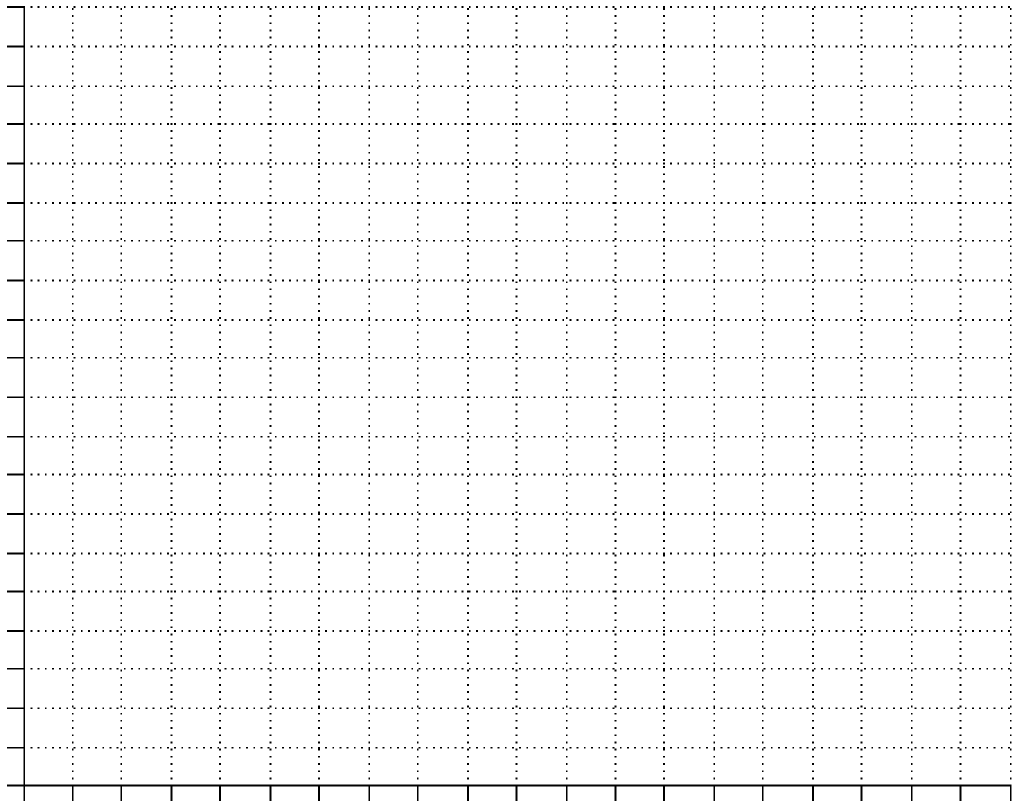
# of dice rolled	# of successes in the rule

Z. What do you notice about the standard deviation estimates as sample size increases?

AA. Clear out the old data from your calculator (or use a different pair of lists) and enter the data in the table. When that this done, go to STAT, then CALC menu and then scroll down to A: PwrReg. This function gives a regression equation in the form $y = ax^b$. Recall that is b is close to one, this is a linear equation with a zero intercept. If b is close to two, then this is a parabolic graph. If b is close to ½, this is a square root function. If b is negative, the power is in the denominator.

BB. What equation did you get? (Note: you may need to eliminate the (0,0) point from your data to make this work.)

CC. Plot the points and the equation you obtained below. How good is the fit?



DD. The Central Limit theorem says that the spread of results of a sample decreases proportional to $\frac{1}{\sqrt{n}}$. Does this agree with the results you obtained above? Why or why not?

Check with your instructor.