Homework 4

Reading Assignment: The handout on Statistics by February 20
Huff, Chapters 2, 3, and 4; by Thursday, Feb. 22

Written Assignment: Rough Draft: midnight Thursday, Feb. 22
Final Draft: midnight Tuesday, Feb. 27

This assignment is to be performed in groups, but you must each write your own final report. To perform this assignment, at least one member of the group must download the Excel spreadsheet called dice.xls. It can be obtained from the web page at address

http://users.wfu.edu/ecarlson/skeptic/dice.xls

Type in this address, click on **Open it**. If you want to save the program to your own computer, you can do so now.

This program is an automatic dice-rolling program. The dice can roll any number from 1 to 6, just like ordinary dice, but some of the dice are loaded and some are not. Set 1 is not loaded ("fair"), but some of the others are loaded.

You can roll simultaneously anywhere from 1 to 6 dice, all of the same type. The computer will calculate the totals of the dice rolled, which, depending on how many dice you roll, may be any number between 1 and 36. The computer will perform such a roll any number of times from 1 to 3000. You are supposed to study the statistics of the rolls. The program also graphs how many times each total occurs (the blue curve/data), and does the best job it can of fitting the data with a bell curve (the orange curve). You can see a table of the actual number of times each roll occurs just to the left of the graph.

This homework assignment is unlike other assignments, in that you are responding directly to questions that I am putting forward. It is therefore recommended that you number your answers and simply respond to the questions without repeating the question.

1. This question has nothing to do with the dice program. Suppose that nine measurements of a random variable produce the numbers 1, 5, 2, 4, 6, 3, 1, 1, 6. What is the best estimate of the mode, median, and mean of these numbers?

2. This question has to do with the combining of means and standard deviations. Set the dice set to set 2, the number of dice rolled to 1, and the repeat to 3000.
   (a) These dice are obviously not fair. Which numbers come up more often than others? The program automatically calculates the mean and standard deviation of the 3000 repetitions. Record them as well.
   (b) Suppose you take two of these dice and add them together. What would be the expected mean and standard deviation for two dice? Change the "Number" of dice to 2, and check your result.
   (c) Repeat part (b) for three dice, four dice, and five dice. In other words, based on your results for Number = 1, predict the mean and standard deviation for Number = 3, 4, 5, and then try it out and make sure you got (approximately) the predicted results.

3. This is a test of the central limit theorem, which says that if you add many things (dice) together, you should always get a normal distribution. Keep the dice set at set 2 and the repetitions at 3000, but change "Number" successively from 1 to 5. For each value of "Number" make a qualitative comparison between the blue data and the orange bell curve. Explain how this demonstrates the central limit theorem.

Testing whether dice are loaded

For the remainder of this homework assignment, you will be testing whether a set of dice are loaded or not. The steps involved are always the same: (a) form a hypothesis, (b) perform a test of the hypothesis and record the data, (c) analyze the results and either accept or reject the hypothesis.

The formulas you use for the next three problems are always the same. If something has a probability p, and you perform N measurements on it, the number of times that event should occur is given by

$$Np \pm \sqrt{Np(1-p)}$$

That is,

$$\mu = Np \qquad \text{and} \qquad \sigma = \sqrt{Np(1-p)}$$

You will use these values to get a z-value for something to occur by chance, namely

$$z = \frac{|x - \mu|}{\sigma}$$

Based on the z-value, you will conclude either that your hypothesis is not proven, proven, or somewhere in between. For example, if you hypothesize that the number 4 is rolled more often than it should be, and your data bears this out, look at the z-value. If it is less than 2, then your hypothesis is not proven. If it is in the range 2 to 5, then you have evidence that it is true, but you can't prove it. If it is greater than 5, then it is definitely proven.

4. Set the dice set to set 3, the number of dice rolled to 1, and the repeat to 300. The hypothesis we will be testing in this case is that the number 4 comes up more often than it should. The competing hypothesis is that the dice are "fair," that is, that the number 4 comes up as often as it should.
    (a) If the dice were fair, how many times should each number come up? How many times did the number "4" come up? At this point you form the hypothesis "4 comes up more often than other numbers."
    (b) Set the number of repetitions to 600. Assuming the die is fair, how many times should the number 4 come up (μ)? What is the standard deviation in the number of times this number should come up (σ)? What is the actual number of times it came up (x)? (Warning: This has nothing to do with the automatically computed mean and standard deviation).
    (c) What is the z-value for this experiment? Based on this z-value, how confident can you be that dice set 3 is "loaded"?

5. Now we will repeat the previous part with dice set 4. Set the dice set to set 4, the number of dice rolled to 1, and the repeat to 1500. You will have to make up your own hypothesis about this set of dice.
    (a) How many times should you get each of the numbers 1 through 6? Which of the six numbers is farthest from what we expect? Based on 1500 rolls, form a hypothesis about the dice. Your hypothesis will be of the form "The number q comes up more (or less) often than other numbers." You *must* come up with a hypothesis of this form.
    (b) Set the number of repetitions to 3000. Assuming the die is fair, how many times should your chosen number q come up (μ)? What is the standard deviation in the number of times this number should come up (σ)? What is the actual number of times it came up (x)?
    (c) What is the z-value for this experiment? Based on this z-value, how confident can you be that dice set 4 is loaded?

6. Set the dice set to set 5, the repetitions to 1500, and then repeat steps 5(a) through 5(c). Once again, you *must* form a hypothesis that the dice are loaded in some way.