**Part 5: Errors and uncertainties**

Background: Another goal of this homework is to have you engage in an activity that promotes your understanding of how errors in measurements can lead to uncertainties in the results. This occurs in many instances in real life. For instance, you buy a car, and the car dealer tells you that the car gets 30 miles per gallon. The car has a 10 gallon tank, so you decide to fill it up and drive it to your friend’s house, who lives 280 miles away. However, you run out of gas at mile 270. This is an example of how numbers are often reported in real life, where the number has some uncertainty associated with it. Depending on whether you run the air conditioner, draft behind a truck, or speed, your gas mileage will vary.

When looking at reports of science findings in the news, single numbers are often reported, even though that one number does not represent the TRUE VALUE (much like the 30 miles per gallon). When a scientist reads the reported value, he/she usually assumes that there is some amount of error that is not being reported, or looks to see if the uncertainty is being reported. You see this often when polls are taken. You may read that a politician’s approval rating may be 36 percent plus or minus 4 percent. The 4 percent shows the range of possibilities of the approval rating. That is, the poll shows that the approval rating is somewhere between 32 and 40 percent. This uncertainty reflects errors that occurred that are inherent in the data.

There are two parts to this uncertainty: Accuracy and precision. Accuracy represents the deviation from the true value; precision reflects the clustering of data. The smaller the error (say, we go from 4 to 3 percent), the more precise the result. We say data is more precise when it becomes more clustered. This should not be confused with accuracy. If the politician’s actual, true approval rating was 30 percent, our 36 percent plus or minus 4 percent would not be accurate. Even if we become more precise, and bring our uncertainty down to plus or minus 3 percent, we are not accurate, since 30 is not between 33 and 39. Accuracy reflects how far a measured value deviates from the true, absolute value. So, if the true approval rating is 30 percent, 32 (plus or minus 4 percent) is accurate; 36 (plus or minus 3 percent) is not accurate. In other words, the precision is not necessarily correlated with the accuracy.

The problem: Scientists characterize uncertainty of experimental (or natural observation) results using mathematics. All good experimental data will have error bars, which are determined mathematically from the scatter of results observed in the experiments. In fact, you can tell empirical results, because they have commonly have error bars, from theoretical results that typically don’t have error bars.

Error bars typical indicate one (or two standard) deviations of uncertainty. A standard deviation is just a measure of the variability of a population of numbers. The first step is to calculate the standard deviation.

Consider a population (of 8 values) consisting of the following values:

2, 4, 4, 4, 5, 5, 7, 9

There are eight data points in total, with a [mean](http://en.wikipedia.org/wiki/Mean) (or average) value of 5:

 2+4+4+4+5+5+7+9

Mean = ----------------------------- = 5

 8

To calculate the population standard deviation, first compute the difference of each data point from the mean, and [square](http://en.wikipedia.org/wiki/Square_%28algebra%29) the result:

 (2-5)2 = (-3)2 = 9 (5-5)2 = 02 = 0

(4-5)2 = (-1)2 = 1 (5-5)2 = 02 = 0

(4-5)2 = (-1)2 = 1 (7-5)2 = 22 = 4

(4-5)2 = (-1)2 = 1 (9-5)2 = 42 = 16

Next average these values and take the [square root](http://en.wikipedia.org/wiki/Square_root) to give the standard deviation:

 2+4+4+4+5+5+7+9

Square root ( -----------------------------) = Square root (4) = 2

 8

Therefore, the above has a population standard deviation of 2.

Below, calculate the standard deviation for each of your data points.

You then put this data on the graph, by added a bar that goes out 2 points on both sides of your data point.