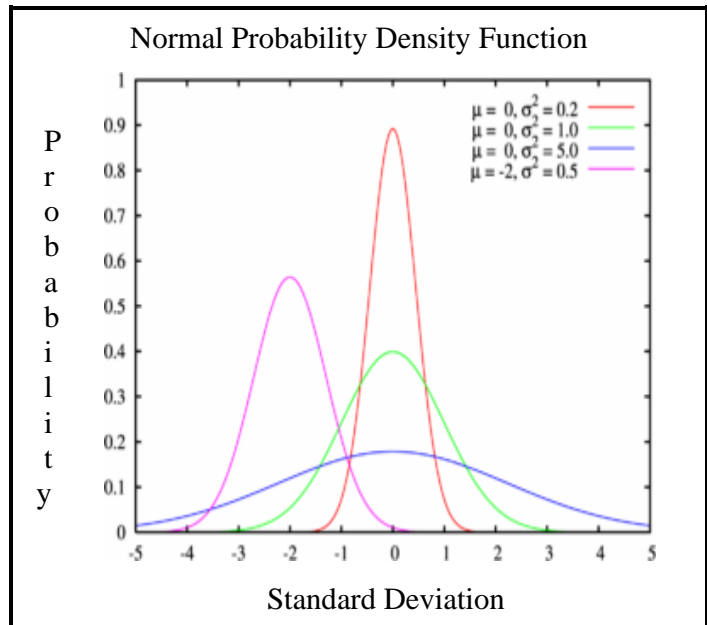


Name _____ Period _____ Date _____

The Normal Distribution

Read the following about the importance of the normal distribution. (Modified from Wikipedia: http://en.wikipedia.org/wiki/Normal_distribution)

The **normal distribution**, also called the **Gaussian distribution**, is an important family of continuous probability distributions, applicable in many fields. Each member of the family may be defined by two parameters, *location* and *scale*: the mean ("average") and standard deviation ("variability"), respectively. Carl Friedrich Gauss became linked with this set of distributions when he analyzed astronomical data using them and defined the equation for the probability density. It is a **bell curve** because the graph of its probability density resembles a bell.



The importance of the normal distribution as a model of quantitative phenomena in the natural and behavioral sciences is due to the central limit theorem. Many psychological measurements and physical phenomena (like photon counts and noise) can be approximated by the normal distribution. While the mechanisms underlying these phenomena are often unknown, the use of the normal model can be theoretically justified by assuming that **many small, independent effects** are additively contributing to each observation.

The normal distribution also arises in many areas of statistics. For example, the sampling distribution of the sample mean is approximately normal, even if the distribution of the population from which the sample is taken is not normal. In addition, the normal distribution maximizes information entropy among all distributions with known mean and variance, which makes it the natural choice of underlying distribution for data summarized in terms of sample mean and variance. The normal distribution is the most widely used family of distributions in statistics and many statistical tests are based on the assumption of normality. In probability theory, normal distributions arise as the limiting distributions of several continuous and discrete families of distributions.

As noted above, most scientific data can be represented by a normal distribution function. The following data was obtained by measuring the mass of a metal strip 24 times with the same balance.

*Mass for a Metal Strip (Grams)			
8.150	8.145	8.148	8.145
8.155	8.155	8.156	8.152
8.142	8.143	8.144	8.148
8.150	8.153	8.152	8.150
8.140	8.149	8.146	8.146
8.148	8.145	8.148	8.150

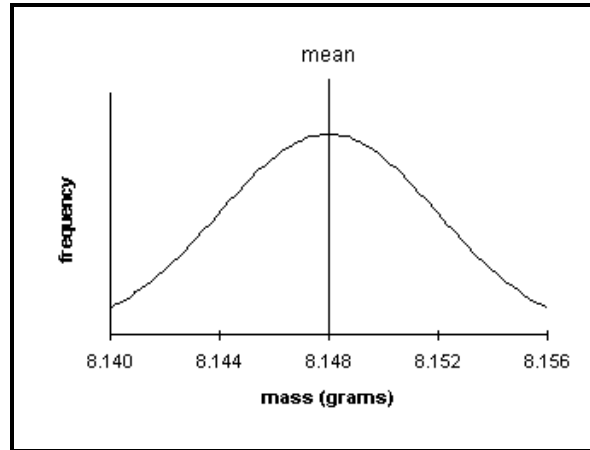
Using Excel, answer the following questions:

1. Type in the above values in column A. Using the Excel formulas, AVERAGE AND STDEV:
 The mean value (AVERAGE) is _____ g.
 The standard deviation (STDEV) is _____g.
2. Run a histogram. In column B, input numbers from 8.140 to 8.156 in steps of 0.001. For example, 8.140, 8.141, 8.142 ... 8.156. Go to "Tools," "Data Analysis," and select "Histogram." For the "Input Range," drag over column A. For the "Bin Range," drag over column B. Select "Output Range" at "\$G\$1," Hit "Run."

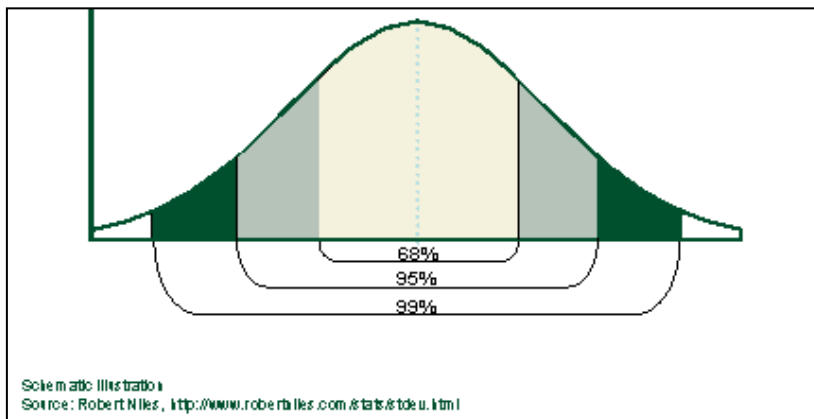
From the frequency distribution, which value occurred most often?

3. Is your answer to question # 2 consistent with the histogram? Explain.

If we were to conduct a very large number of measurements on the metal strip, we would have obtained a bell shaped histogram. A bell shape curve represents the normal distribution.



Facts about a normal distribution:



- The mean (average) is the most probable measurement.
- About 2 out of 3 observations (68.3%) will fall within one standard deviation of the mean. This means that the probability of having a measurement with a value of $\text{mean} \pm \text{standard deviation}$ is 68.3%.
- About 19 out of 20 observations (95.5%) will fall within two standard deviations of the mean. This means the probability of having a measurement with a value of $\text{mean} \pm 2 \text{ standard deviations}$ is 95.5%.

- This means that you are 95.5% confident that a measurement will fall within the limits of 2 standard deviations,

Using the average and standard deviation you found in #1.

4. 68.3% of all the measurements will fall within the limits of _____ g and _____ g.
5. 95.5% of all the measurements will fall within the limits of _____ g and _____ g.
6. This means is that you are 95.5 % confident that any measurement will be between the values of _____ and _____.
7. Any measurement that falls outside of the range of 1.840 to 1.856 g is:
 - a. reliable
 - b. doubtful
 - c. can't tell

Explain your choice.

*This data and information for this worksheet was obtained from "Experimental Errors and Data Analysis" by J.C. de Paulam, <http://www.haverford.edu/chem/302/data.pdf>