

Analyzing Data: Comparing different data sets

Once a person has collected data from an experiment, she/he must analyze the data. We can use some basic ideas to see if one “treatment” of the independent variable has created a “significantly different” change in the dependant variable compared to another independent variable. The issue here is what is “*significantly different*” mean? In this module we will approach this topic on a qualitative level.

In order to examine this issue, we will weight three different groups of pennies in order to determine whether each group has similar or “significantly different” weights.

Find a balance and the three containers of pennies assigned to your group. (*Please do not mix the pennies from the various containers! Thanks.*)

- Examine each container of pennies and make a prediction. How do you think the weights of the pennies in these 3 contains will compare with one another?

- You and your group should weigh each penny in the containers assigned to your group as precisely as possible. Record your data in a **data table** and attach it to the end of this module.

You probably found that many of the pennies in a specific container were about the same weight, but there was some variation; some of the pennies weigh a little more and others weighted a little less. For a specific container, if you add all of the weights you measured and divide by the number of pennies you will determine the **mean** (or average) weight of the pennies in that container.

- What is the mean weight for each of the 3 containers that you measured?

- In the space below, list all of the weights for one of the containers (pick any one), but this time put them in order from the smallest weight to the largest weight. Draw a line separating the values in half (in other words, find the point at which half of the values are larger and half of the values are smaller). This value is called the **median**. For example, in King County for 2012, the median home price was over \$375,000. Is the median always the same as the mean? Why or why not?

- Imagine Bill Gates builds a huge \$50,000,000 house in Auburn. Which would change more, the mean or median home value in Auburn? Explain your thinking.

- Another measure of central tendency is called the **mode**. The mode is the most frequent value. Look at the penny weights for one of your containers. Is there one value that happened more frequently than others? If so, which one?
- Imagine that one of the containers has an additional penny. If we wanted to predict the weight of the penny before we weighed it, would we use the mean value, the median value or the mode value? Why?

Graphing your data

In the part above you created a data table. That is one way to display data, but if there are a large number of data points, it is difficult to interpret data from a data table, even if you calculate the mean, median and mode.

Another way to illustrate data is to graph the data. We will use this method frequently throughout the year. In this case, we want to graph the number of pennies as a function of the weight of the pennies. You put the number of pennies on the pennies on the y-axis and the weight of the pennies on the x-axis.

This type of graph is called a **histogram**. A histogram is a graphical display of data that uses bars of different heights to show the frequency of different values. Values that occur more frequently will have bars that are higher, while values that occur less frequently will be lower.

If you have not worked with histograms before you may want to review their uses at:

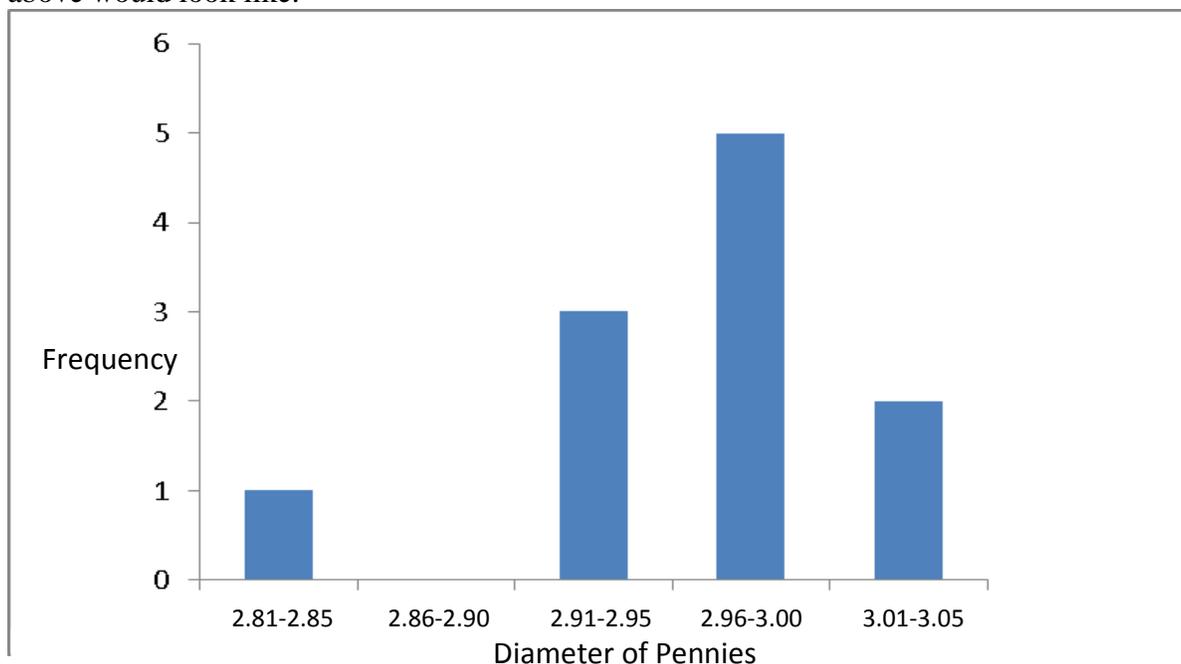
<http://www.mathsisfun.com/data/histograms.html>
<http://quarknet.fnal.gov/toolkits/ati/histograms.html>

If you have questions, let an instructor know). The first step is to organize the data so that it is easier to plot in the graph. For example, a data table of measurements of the diameter of pennies might look something like:

Diameter of the pennies (cm)	Number of pennies
2.81 to 2.85	1
2.86 to 2.90	0
2.91 to 2.95	3
2.95 to 3.00	5
3.01 to 3.05	2
And so on..	

- Attach your sorted data tables for penny weights to the end of this module.

Next you need to create a histogram of the data. A histogram for the example given in the data table above would look like:



- Use paper supplied to draw a histogram graph of the data you collected. Attached your graph to the end of this module. Hint: To conserve paper and to make our graphs the most helpful, your x-axis (the mass of the pennies) should go from about 2.3 grams to at least 3.2 grams. Note: some of your weight may overlap so use a different color pencil for each container of pennies.. Be sure to label the colors on the graph.
- Are all the weights you found for the pennies equally common? If not, which weights were most common?
- Describe the approximate shape of the distribution of weights for each group of pennies.
- Which 2 containers contained pennies that were the most similar in weight to one another? Do you feel that the weights of the pennies on each container is significantly different? Explain your reasoning

- Which 2 containers contained pennies that were the most different in weight to one another? Do you feel that the weights of the pennies on each container is significantly different? Explain your reasoning

- How did the pattern you found compare with your prediction?

- Suggest a reason that the distribution of weights for these pennies may have differed:

You may have found that the data for the pennies fell into two different groups and each of those groups had very little variation in weight. In this case it is relatively easy to see that there are two distinct groups of pennies. In some cases the mean values for the two groups are the same *however the distribution of values is different indicating that the two samples are from two different populations.* Later this quarter you will learn how to calculate the **standard deviation** of the values to estimate the variation in the data using Microsoft Excel. Although we are not going to calculate the standard deviation of any data in this module, but you might appreciate that if there is a lot of variation in the values, the standard deviation is a higher value. In the example with the pennies, the distribution of the masses of the pennies was not very wide. Most of the pennies weighed about the same, so the standard deviation is low.

- In summary, what factors should you consider when you analyze data to determine if two groups are similar or different?