Data Analysis *Handout*

In order to analyze data in large data sets it is often useful to calculate a number that summarizes the data. The most well-known and useful example is the average or mean value.

MEASURES OF CENTRAL TENDENCY

Measures of Central Tendency are statistics you calculate if you want to represent a group of data by a single value. This value may also be referred to as the expected or the most-likely value. The most common measure of central tendency is the mean, or average value. To calculate the mean, add up all of your values divided by the number of values:

$$\text{Mean} = (X_1 + X_2 + X_3 + \ldots X_n) / n$$

Where

X_i = the value of your chemical parameter
n = the total number of values

The Mean is a useful measure of central tendency, but is very sensitive to unusually large or small values (called outliers). For example, the following is a typical data set for fecal chloroform, a type of bacteria, at a sampling station:

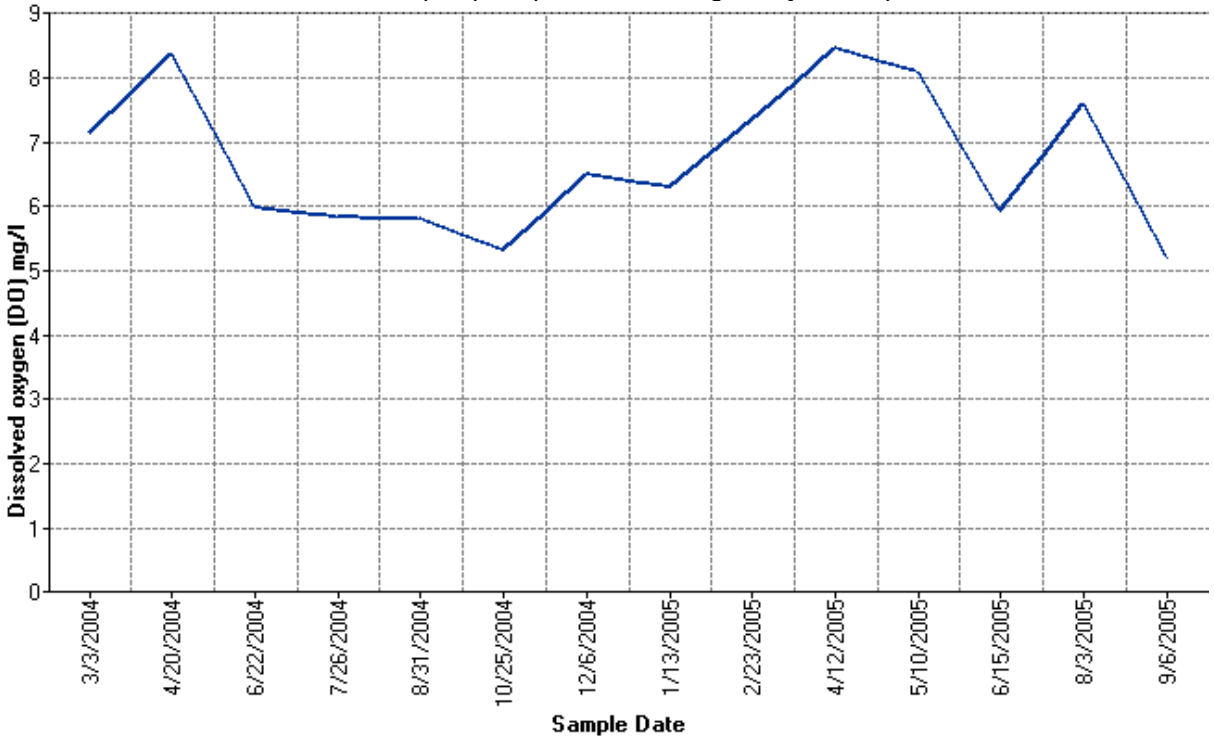
Date	Fecal Coliform (#/100ml)
1/2/1998	50
1/9/1998	40
1/16/1998	90
1/23/1998	500
1/30/1998	80

The Mean of these values is (50 + 40 + 90 + 500 + 80) / 5 or 152 #/100ml. This would be a clear violation of the Class A water quality standard of 100 #/100ml. However, if you remove the largest value (500) the Mean becomes (50 + 40 + 90 + 80)/4, or 65 #/100ml. This Mean value would pass Class A standards.
Given all of the possibilities for measurement error, sampling handling error, and natural variability, you don't want your conclusions to be so heavily influenced by a single odd sample; the value of 500 #/100ml might be the result of sample contamination or just an unusually polluted parcel of water.

Data Analysis *Handout*

Here is an example of Dissolved Oxygen values for Lake Tarpon in Pinellas County graphed over a 2-year period.

2-Year Time Series Plot of Dissolved Oxygen for Lake Tarpon.
(Graph represents average daily values)



It's Your Turn

MATERIALS: Internet access, Excel or equivalent spreadsheet program

PROCEDURE:

1. Go to www.pinellas.wateratlas.org > Research > Data Download Tool.

2. Follow these steps:

   a. Select Data Type (Surface Water Quality)
   b. Select Filters (Water Body Name, Date Range, Parameter)
   c. Type in Lake Tarpon > Select the past 2 years using the drop down > Select Parameters either Nitrogen (TKN) or Dissolved Oxygen (DO) and select Fecal Coliform
   d. Select a station or stations that have recent data
   e. Graph data
   f. Under Graph display options select Data Points > Refresh graph.

Name: Date:

Data Analysis *Handout*

3. After creating the graph, cut and paste it into a word document.

4. Label/Title the graph.

5. Create a table denoting the date and water quality parameter.

6. Calculate the mean value for the water quality data. Show all work. Label units. Circle final answer.

7. Are there any outliers? If yes, recalculate the mean without the outlier value(s).

8. How do outliers affect data interpretation?

9. Get together with another classmate who has the other water quality data (i.e. if you did nitrogen, find a student who did dissolved oxygen). Discuss the quality of your lake based on both calculated means and graphs. Write a summary description of what you have determined.