Statistical Analysis

In preparation for your Science IA this year, you will need to practice and come to love statistical data analysis. As you will be completing a lab, you will be asked to assess your data for both validity and significance. Depending upon your background knowledge, you may need to do some research to complete these calculations. I have included a few websites below to get you started. One of your best tools to become associated with is excel. This will allow you to create graphs and complete calculations with the greatest ease. It is worth your time!

Your assignment will be to learn and practice these skills using the provided data.

The validity tests will help you to determine the consistency of your data. Lab work is all about being repeatable, so you are expected to be able to analyze and discuss your data in the context of reliability.

- Mean/Average:

Add the values of a trial and divide by the number of trials

- Standard deviation: https://www.mathsisfun.com/data/standard-deviation.html

1. Find the mean, or average, of your sample.
2. Subtract the mean from each number in the sample.
3. Square all of the products from the previous step.
4. Add the squared products together.
5. Divide the sum of the squared products by (n-1), where n=how many numbers are in the sample, to find the variance.
6. Calculate the square root of the variance to find the standard deviation.

- Error bars:

www.rwu.edu/sites/default/files/downloads/fcas/mns/adding_error_bars_to_excel_graphs.pdf

Indicate estimated error or uncertainty to give a general sense of how precise a measurement is. Typically, Error bars are used to display either the standard deviation, standard error, confidence intervals or the minimum and maximum values in a ranged dataset. (Using your time to the best advantage, I suggest using an excel spreadsheet and learning the steps to add error bars.)

- r squared value: projects.ncsu.edu/labwrite/res/gt/gt-reg-home.html

Calculation that shows how well the data points fit onto the best fit trend line. (Please look up how to perform this calculation in the excel program or on your calculator.) The closer the r^2 value is to 1, the better.

The significance tests will help you to determine the impact of uncertainties. You will need to evaluate how the uncertainties that exist in all measurements, can alter your perspective of the data trend. Hopefully your data will reveal a trend that will either support or refute your hypothesis, but you will need to analyze and discuss if that trend is actually meaningful. When using these tests, you would create a null hypothesis, statement showing that there is no statistical significance and an alternative hypothesis, statement showing that there is a statistical relationship.

- Z test (large sample/ t test (small sample): www.investopedia.com/terms/z/z-test.asp

When comparing sample populations or data sets (before and after scenario) with normal bell curves, the tests will determine if their means are statistically different. This will provide evidence to support the null or alternative hypothesis.

- chi squared test/ G test: https://libguides.library.kent.edu/SPSS/ChiSquare

Determines if a qualitative variable (present vs absent) has a relationship or not with another variable.

- p values: https://www.wikihow.com/Calculate-P-Value

Evaluates if the data fits the expected model or not. The "expected" values are found through background research to make the comparison. If the newly observed data is different enough from the expected, then variation then a relationship does exist between the manipulated variable and the responding variable.

- f test/anova test: http://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/

http://blog.minitab.com/blog/adventures-in-statistics-2/understanding-analysis-of-variance-anova-and-the-f-test

"Analysis of variance" This compares multiple sample populations with normal bell curves, to determine if curves are statistically different or if the variation is too great to make the determination.

Timing the reaction progress of HCl with 3 different metals through temperature

Time, t/s	Temperature of HCl with excess Ca, T/°C (±0.1°C)			Temperature of HCl with excess Mg, T/°C (±0.1°C)			Time, t/s	Temperature of HCl with excess Zn, T/°C (±0.1°C)			Temperature of HCl with excess Fe, T/°C (±0.1°C)		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3		Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
0	19.1	21.4	20.9	18.7	19.6	19.4	0	19.2	19.1	18.9	18.5	18.3	18.5
5	20.6	24.9	25.1	19.5	25.2	31.2	20	19.9	19.7	19.5	18.9	18.5	18.9
10	30.5	35.1	37.4	42.6	35.0	44.8	40	20.1	19.8	19.6	19	18.5	19.2
15	40.2	44.7	46.4	53.3	44.5	52.2	60	20.3	20.0	19.8	19.2	18.6	19.4
20	48.6	48.6	52.9	56.7	52.5	55.8	80	20.5	19.9	19.9	19.3	18.7	19.5
25	52.6	52.1	57.3	57.9	56.3	57.1	100	20.5	20.1	19.9	19.4	18.7	19.7
30	56.5	55.4	60.5	57.7	57.4	57.9	120	20.5	20.3	20.0	19.5	18.8	19.9
35	59.3	59.3	64.2	57.3	58.3	57.8	140	20.6	20.5	20.1	19.7	18.9	20.2
40	61.7	61.9	66.6	57.0	58.8	58.6	160	20.7	20.8	20.3	19.8	18.9	20.2
45	64.0	64.2	69.2	-	59.2	58.9	180	20.8	20.9	20.4	20.0	19.0	20.3
50	66.4	66.5	72.2	-	58.9	58.6	200	20.7	20.8	20.6	20.2	19.1	20.5
55	68.3	68.5	74.6	-	58.5	58.1	220	20.5	20.6	20.7	20.3	19.1	20.6
60	72.4	70.5	77.0	-	-	-	240	-	-	20.8	20.5	19.2	20.7
65	74.5	72.4	79.2	-	-	-	260	-	-	20.7	20.6	19.1	20.8
70	74.8	73.6	80.5	-	-	-	280	-	-	-	20.7	19.1	20.9
75	75.1	75.5	81.2	-	-	-	300	-	-	-	20.8	19.2	21.1
80	76.1	77.1	81.5	-	-	-	320	-	-	-	20.9	19.3	21.1
85	76.5	77.3	81.3	-	-	-	340	-	-	-	21.1	19.4	21.2
90	76.5	77.2	80.9	-	-	-	360	-	-	-	21.1	19.5	21.4
95	76.6	77.4	79.1	-	-	-	380	-	-	-	21.2	19.5	21.6
100	76.7	77.2	78.2	-	-	-	400	-	-	-	21.3	19.5	21.5
105	76.4	77	77.5	-	-	-	420	-	-	-	21.4	19.5	-
110	76.0	77.8	76.7	-	-	-	440	-	-	-	21.4	19.6	-
115	75.6	77.0	-	-	-	-	460	-	-	-	21.3	19.7	-
120	75.1	76.0	-	-	-	-	480	-	-	-	21.3	19.8	-
125	74.4	75.3	-	-	-	-	500	-	-	-	21.2	19.7	-

Table 2. Temperature of HCl solution with calcium, magnesium, zinc, and iron (*Dash shows that the reaction has ended)

Name _____ Per. _____

Science IA Statistics

1. Create 3 graphs using the data provided. All graphs will contain a title, labeled axes with units and uncertainty, and the intervals will be appropriate to maximize the space of the graph. You will need to provide a legend for the three samples on each graph. Use excel, it will save you so much effort in the long run…

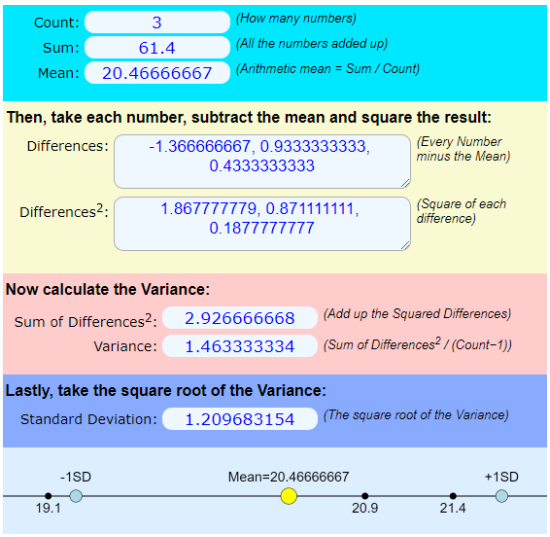
2. Determine the mean for each time, for each metal. This is when the use of the excel program could really come in handy…
   Ex.

Time	Trial 1 ± 0.1°C	Trial 2 ± 0.1°C	Trial 3 ± 0.1°C	Mean
0	19.1	21.4	20.9	20.4 ± 0.1°C

3. Add the mean to each of the three graphs, don't forget to add to your legend…

4. Find the standard deviation for each each time, for each metal.
   Ex.

Time	Trial 1 ± 0.1°C	Trial 2 ± 0.1°C	Trial 3 ± 0.1°C	Mean	Standard Deviation
0	19.1	21.4	20.9	20.4 ± 0.1°C	1.2



If you are not using the excel program, steps are on the left show how the standard deviation was found.

5. Add error bars based upon the standard deviation to the "mean" line on all three of your graphs.

6. These graphs should show a linear line up to a certain point and then taper off as the reaction ends. Calculate the r^2 value for the linear portion of each line – You will calculate 12 r^2 values, each trial plus the mean.

7. These graphs do not have bell curves so our list of significance tests are not applicable. Instead, please discuss qualitatively if there is a realtionship between the different metals reacting and the temperature. Justify your response with evidence from the graphs.