**STAGE 5: data collection & analysis**

 Due Date:

**DATA COLLECTION**

Upon approval of the experimental design, you can begin the data collection process. Students should begin experimentation as soon as possible to account for the length of time that is necessary for the experiment to be conducted multiple times. **It is imperative that you accurately enter ALL of your quantitative data in your FOLDER as you collect it.** Create an organized chart/data table to enter ALL of your results so that you and anyone else viewing your folder can understand the information. Also keep a dated, detailed log of all the other **qualitative observations** that you make during your experiment. Describe any changes you see occurring or other information that is not included in the measurements that you are recording.

This is also where you want to write any reflections that you might have as you complete the experiment. Reflections include what seems to be working well versus what might not be working well. If things are not working in the experiment, discuss what changes can be made (and then make them!). As always, **make sure to DATE all entries that you make during the data collection process** **and don’t forget to take pictures of yourself collecting data!**

**DATA ANALYSIS**

Once data collection is complete, the next step is to summarize and analyze your data using the correct statistical test. Your chemistry teacher will provide help while you complete this phase of your project. Many projects will yield unexpected results that do not support the hypothesis. That is not failure, it is science!

**Summarizing Your Data:**

In order to have the most accurate data for your project, multiple trials should have been performed during the experiment. Summarize your data by calculating the **mean** for each group of trials and determine the standard deviation of those means. The **standard deviation** is the most common method of measuring the variation from the mean that exists for each testing group. A low standard deviation indicates that data points tend to be very close to the mean; a high standard deviation indicates that data points are spread out over a large range of values. When performing these calculations, be mindful of outliers or values that fall well outside of the overall pattern of the data set. Mean and standard deviation values can be calculated using EXCEL software or by visiting http://www.graphpad.com.

**Analyzing Your Data:**

In order to determine whether differences in data collected for each testing group is statistically significant (which would imply that your independent variable DID have an effect), data analysis needs to be completed. Below are descriptions of four types of statistical tests that may be used.

**T TESTS:** http://www.graphpad.com/quickcalcs/

This test is used to compare the data from two different testing groups to determine if they are significantly different. There are two different types:

**PAIRED**: Used when the data values consist of pairs of observations on the same person or object. An example of this would be comparing the heart rate of individuals before and after watching a scary movie.

**UNPAIRED**: Used when the data from one group are not directly linked to the data of the other group. An example of this would be comparing the effect of red light versus blue light on the growth of plants.

**ANOVA:** http://www.vassarstats.net/anova1u.html

The test is used to determine whether there are any significant differences between the means of three or more independent (unrelated) groups. For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels amongst students, dividing students into three independent groups (e.g., low, medium and high-stressed students).

\*Both these tests will give you a **p-value**, which will if your testing group data are statistically significant from each other. The smaller the p-value, the greater the difference between the groups. To say that the values are statistically different, the p-value must be below 0.05.

**CHI-SQUARE TEST:** http://www.graphpad.com/quickcalcs/

This test is used to determine if there are differences between two or more frequency distributions. Both independent and dependent variables need to be **categorical data.** An example would be comparing the foraging height (upper, middle, and lower) of yellow warblers in different tree species (oak, maple, aspen, and hazel).

**COEFFICIENT of CORRELATION**: http://easycalculation.com/statistics/correlation.php

This test is used to determine if the values of the dependent variable are directly related to the independent variable. You would use this test when both the variables are measured on a scale (i.e. time, temperature, length, etc.) An example of this would be to determine the effect of pH value on the growth of plants. After you enter data, you will get an **r-value**, which will indicate how strongly related the two variables are. The closer the value is to 1 or -1, the stronger the relationship is.

**RESOURCE:** [**www.graphpad.com/quickcalcs/index.cfm**](http://www.graphpad.com/quickcalcs/index.cfm)

Once you have performed your statistical analysis, you will need to visually represent your quantitative and qualitative data in an organized manner. Create at least one table that summarizes your data and shows any trends that you have found. Remember that all tables must have an informative title that includes the dependent and independent variables, properly labeled columns and rows (including metric units of measurement when applicable), and data reported to the proper number of significant figures.

**GRAPHING DATA**

Graphs are used to visually represent trends and relationships in your data. Your Chemistry teacher will provide guidance as to what type of graph you should use. Below is a list of commonly used graphs:

* **Time**-**series** **Plot** – can be used if the dependent variable is numerical and the independent variable is time.
* **XY Line Graphs** – shows the relationship between your dependent and independent variables when both are numerical and the dependent variable is a function of the independent variable.
* **Scatter Plots** - might be the proper graph if you are trying to show how two variables MAY be related to one another.

Be sure to include a descriptive title that indicates the independent and dependent variables. For line graphs, be sure to label each axis and indicate the proper unit of measurements. Place the independent variable on the X-axis and the dependent variable on the Y-axis. If you have multiple sets of data, show each series in a different color or symbol and include a legend with clear labels.

**ERROR BARS**

Your graph should include error bars, which are graphical representations of your data’s variability used to indicate the [error](http://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics), or uncertainty in a reported measurement. Error bars often represent one [standard deviation](http://en.wikipedia.org/wiki/Standard_deviation) of uncertainty, one [standard error](http://en.wikipedia.org/wiki/Standard_error), or a certain [confidence interval](http://en.wikipedia.org/wiki/Confidence_interval) (e.g., a 95% interval). These quantities are not the same and so the measure selected should be stated explicitly in the graph or supporting text.

**DON’T FORGET TO MAKE ENTRIES IN YOUR FOLDER!** A lot of the calculations and components in this section can be done on the computer or with a calculator. However, you must still document how you went about obtaining these calculations. You can identify websites or programs that you used, **print out anything that you generate**, and DISCUSS any formulas that were applied and describe how you used everything.

NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ PD: \_\_\_\_\_\_ DATE: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

**DATA COLLECTION AND ANALYSIS RUBRIC**

Once you have collected data from your experiment, you need to be able to gather information and determine trends in the data. Results are NOT the same as raw data (which is what you collected). Results are summarized and analyzed using the correct statistical test. **Please place this rubric behind your tables and graphs in your folder.**

|  |  |  |
| --- | --- | --- |
|  | **POINTS EARNED** | **POINTS POSSIBLE** |
| **Data Tables** |
| Sufficient measurable data has been collected using appropriate SI measurements. (**Minimum of 10 trials)** |  | 2 |
| Each data table has a proper title that identifies the independent and dependent variables. |  | 2 |
| Data is shown in an organized, well-labeled chart/table that is easy to read. |  | 2 |
| There is documented evidence of adequate qualitative data (observations) in the folder.  |  | 2 |
| **Graphs** |
| The proper graph was selected to represent the data (i.e. line vs bar) |  | 2 |
| Each graph has a proper title that identifies the independent and dependent variables. |  | 2 |
| Variables have been placed on the proper axis. |  | 1 |
| Each axis has been properly labeled and a key has been provided (if necessary). |  | 2 |
| Graphs reflect mean values, not raw data. |  | 1 |
| **Statistical Analysis** |
| Means of the raw data have been calculated for each testing group. |  | 1 |
| Standard Deviation for each testing group has been calculated.  |  | 2 |
| Statistical analysis has been completed |  | 1 |
| The proper type of statistical test was used to determine patterns/trends between the testing groups. |  | 2 |
| Error bars are included in the graphs to show the variability of the data. |  | 2 |

**TOTAL POINTS = \_\_\_\_\_\_\_\_ / 24**