Biological Applications of Statistical Analysis

Stuart Surrey
Philadelphia High School for Girls

Index:

Overview:

At the root of any experimental science is the accumulation of recorded data. Analyzing and interpreting that data will ultimately influence the conclusion of the experiment and/or study. Therefore, the ability to determine reliable data from erroneous data becomes an integral component of any investigation. Determining the validity of such data can be achieved utilizing standard statistical analysis techniques. These methods include, but are not limited to, calculating the: mean, mode, median, range, sample variance, standard deviation, chi-square analysis, and the Student's t-test. Consequently, this unit is designed to familiarize high school biology students with certain fundamental concepts of statistical analysis.

Rationale:

Many students in the traditional, comprehensive high schools throughout the city of Philadelphia are below grade level in both reading and math skills. Although this curriculum unit was initially intended to coincide with the first topic of the standard level, International Baccalaureate biology program, it will also be used in other biology classes regardless of the scholastic level of the students. It is essential that all high school students understand the relevance of statistical analysis, not only as it applies to science but, in everyday life as well. As a result, implementation of this unit will occur throughout the academic year depending on the specific course to which the students are enrolled. For example, with regard to the standard level, International Baccalaureate biology class, this unit will be part of the first topic "Statistical Analysis." In the traditional introductory biology course, however, this unit would be part of the following units: Unit 1- Review of the Scientific Method; Unit 2- Introduction to Chemistry; Unit 4- Cellular Respiration and Photosynthesis; Unit 6- Mendelian Genetics; and Unit 10- Ecology; as outlined in the School District of Philadelphia's <u>Planning and Scheduling Timeline for Biology</u> (1).

Science is built upon a foundation of inquiry and experimentation. The data obtained from such investigations has led to and could lead to the further advancement of scientific knowledge and understanding. The reliability of the data, however, is subject to interpretation and is dependent upon a number of factors including: the experimental design, sample or population size, and the accuracy of the instrumentation employed. Validation of the data and the resulting conclusions is based largely upon statistical analysis. Therefore, creation of a curriculum unit which focuses on the implementation of statistical methods and techniques would not only be beneficial to improving the problem solving skills of my students, but it would also introduce them to the importance of statistics.

Most sources will agree that statistics is a mathematical science which focuses on the collection, organization, analysis, and interpretation of experimental data (2, 3, 4, 5). In an article entitled "A Brief History of Statistics in Three and One-Half Chapters: A Review Essay," the following quote was attributed to Stephen Stigler: "If all sciences require measurement – and statistics is the logic of measurement – it follows that the history of statistics can encompass the history of all science." He also pointed out, however, that this philosophy is erroneous and that the role statistics played in science should begin with the development of probability-based statistical methods (6). Articles dealing with the history of statistics tend to agree that statistics had its beginning in the seventeenth century with the development of probability theory and even have attributed its beginning to John Graunt in the 1660's (7, 8). Statistics has become such an integral part of the biological sciences that specific disciplines such as biostatistics and bioinformatics have recently been developed.

Statistical Methods

In general, statistics can be separated into two main branches or categories: descriptive statistics and inferential statistics. The role of descriptive statistics is to numerically or graphically represent the data using such methods as calculating the central tendency, standard deviations, and frequencies. In contrast, inferential statistics is used to make generalizations based on the reported experimental data, such as: hypothesis testing, correlation analysis, and regression analysis (9, 10, 11, 12). The next several paragraphs have been devoted to examining a number of statistical methods that have biological applications.

Descriptive Statistics

Central tendency also referred to as the measure of location, signifies the middle number or central portion of a distribution and includes: the arithmetic mean, harmonic mean, geometric mean, mode, and median. Determining the proper descriptive method depends mainly on the sample distribution and its application. The most common method for determining the average within a given set of data is to calculate the arithmetic mean, the equation for which is listed below (13, 14).

$$\overline{X}_A = \frac{\sum X}{N} \quad \text{(equation 1)}$$

$\overline{X}_A$ is the arithmetic mean, $\sum X$ is the sum of the experimental data and N is the number of trials

Even though it is the most widely used method for determining the average, there are certain situations or conditions for which it should not be used. When dealing with rates for example, the arithmetic mean can give misleading results. Under these conditions the harmonic mean should be used. The following is the equation for the harmonic mean (15).

$$\overline{X}_H = \frac{N}{\sum \frac{1}{X}} \quad \text{(equation 2)}$$

$\overline{X}_H$ is the harmonic mean, N is the number of trials, and $\sum \frac{1}{X}$ is the sum of the reciprocal of the experimental data

Another useful method for determining the average or central tendency, the equation for which is listed below, is the geometric mean (16),

$$\log \overline{X}_G = \frac{\sum \log X_i}{N} \quad \text{(equation 3)}$$

$\overline{X}_G$ is the geometric mean, $\sum \log X_i$ is the sum of the logarithms for the experimental data, and N is the number of trials

The geometric mean would be especially useful when dealing with population growth. Its main drawback is that it requires positive numbers. The following relationship exists between the aforementioned means when the data being compared is the same (17). The arithmetic mean is greater than or equal to the geometric mean which is greater than or equal to the harmonic mean. The equation for this relationship is given as:

$$\overline{X}_A \geq \overline{X}_G \geq \overline{X}_H \quad \text{(equation 4)}$$

Other methods for determining the central tendency include the mode and the median. The mode refers to the numerical value that occurs most frequently within a set of data. It can be unimodal if only one value occurs more often than the others, bimodal when two values within the distribution occur an equal number of times, or multimodal when more than two numbers within the distribution occur an equal number of times. The last method to be mentioned here, but which does not complete the methods for determining central tendency, is the median. For an odd set of values, the median is simply the middle value when arranged in sequential order.

When there is an even numbered set of values, however, the median is calculated by adding the two middle values and dividing by two.

Upon examining experimental data, not only must one consider the mean but all of the data should also be analyzed. Statisticians refer to this data as the variance or dispersion. Measures of dispersion include the range, spread, or variability within a given set or population (18). The range is simply the numerical difference between the minimum and maximum values. Although calculating the range is rather straight forward, there are several different ways of reporting the range. One method is to simply state the minimum and maximum values. Another method is to report that the range is the difference between the minimum and maximum values. An illustrative example is if the minimum temperature was 22°C and the maximum temperature was 28°C, then the range would be the difference between the two or 6°C. A third method is to calculate the difference between the minimum and maximum values and then add one. To use the same example as above, in this case the range would be 7°C and not 6°C. It is important to keep in mind that the range simply is concerned with the minimum and maximum values and not the number of samples.

The term mean deviation is another method for expressing the variability between the reported data. Simply put, this method describes the degree to which data deviates from the mean. The equation for calculating the mean deviation is given as:

$$MD = \frac{\sum \left| x - \bar{x} \right|}{N} \quad \text{(equation 5)}$$

MD is the mean deviation, $\sum \left| x - \bar{x} \right|$ is the sum of the absolute values of the differences between each sample and the arithmetic mean, and N is the total number of samples.

Statistical methods for examining the variability or dispersion of values relative to the mean include sample variance and standard deviation (19). The sample variance can be calculated using the equation listed below:

$$s^2 = \frac{\sum \left(x - \bar{x} \right)^2}{n - 1} \quad \text{(equation 6)}$$

In the previous equation, s^2 is the sample variance, $\sum \left(x - \bar{x} \right)^2$ is the sum of all of the differences between each sample and the mean which is then squared, and n is the number of samples. Calculating the sample variance involves using the sum of the squares and therefore results in a value which is squared. A more convenient method of reporting the variability is to take the square root of the sample variance which is designated as the standard deviation. Before doing so, however, one must consider whether the experimental data is representative of an entire population or merely a sample of that population. If the data is characteristic of an entire population, the following equation is used to calculate the population standard deviation:

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}} \quad \text{(equation 7)}$$

In this equation, σ is the statistical symbol for the population standard deviation, $\sum (x-\mu)^2$ is the sum of the differences between the individual sample values and the population mean which is then squared, and N population size. In contrast, the sample standard deviation can be calculated with following equation:

$$s = \sqrt{\frac{\sum \left(x - \bar{x}\right)^2}{n-1}} \quad \text{(equation 8)}$$

The major difference between the two equations is the use of N for determining the population standard deviation and n-1 for the sample standard deviation. The rationale for the disparity stems from the difference between the entire population versus that of a sample set. In a population all values within a given distribution are present, however, a sampling of the population usually will not include the extreme values at either end of the distribution. The use of n-1 increases the sample standard deviation and gives a better approximation of the population standard deviation.

Graphical Analysis

The statistical methods mentioned to this point have involved equations dealing with a single variable or univariate sets of data. Experiments involving two variables are referred to as bivariate sets of data whereas multivariate data would involve the analysis of more than two variables. Regardless of whether the experimental design involves univariate, bivariate, or multivariate data, the typical method of presenting the data involves the use of tables. An additional method of displaying data relies on the use of graphs or charts. The overall advantage of utilizing graphs versus tables is that with graphs the data is easier to interpret and/or analyze such as trends (20). Modern statistical graphical techniques include the analysis of two dimensional, three dimensional and even four dimensional graphs. With that in mind, the most common statistical representations include: bar graphs, histograms, pictograms, frequency polygons, ogives, box plots, dot plots, scatter plots, ternary graphs, stem and leaf diagrams, line graphs, contour graphs, icon graphs, matrix graphs, probability graphs, and circle graphs or pie charts (21, 22, 23, 24, 25).

Inferential Statistics

As previously mentioned, the role of descriptive statistics is to numerically or graphically represent the data using such methods as calculating the central tendency, standard deviations, and frequencies. In contrast, however, inferential statistics is used to make generalizations based on the reported experimental data, such as: hypothesis testing, correlation analysis, and regression analysis (26, 27, 28, 29). Two very important statistical tests used in the biological sciences to test for the significance of the experimental data are the chi-square test also referred to as "goodness of fit" and the Student's t-test.

The statistical method for examining the differences between the observed results with those of the expected results is known as the chi-square test. This test, which is also referred to as the "goodness of fit", assesses the variability between the observed results with those of the expected or theoretical results. The equation below, in which O and E are the observed and expected values, is used to calculate the chi square (30, 31, 32).

$$\chi^2 = \sum \frac{(O - E)^2}{E} \qquad \text{(equation 9)}$$

Once the chi square is calculated, the degrees of freedom must be determined. This value is one less than the total number of trials within a sample of size n.

Another statistical test, developed by William Gossett in 1908 while working at the Guiness Brewery in Dublin, Ireland, is used for determining the reliability of data obtained from small populations. His work resulted in what is commonly referred to as the Student's t-test or t-distribution. There are a number of different t-tests which are used for a variety of different conditions. The equation for a one sample t-test is given as:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \qquad \text{(equation 10)}$$

In the above equation, $\bar{x}$ is the sample mean, μ is the population mean or a hypothesized value, s is the sample standard deviation, and n is the number of samples. When comparing the sample means from two different groups, the equation for the t-test becomes:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{(equation 11)}$$

In the above equation, $\bar{x}_1$ and $\bar{x}_2$ are the sample means, s_1^2 and s_2^2 are the sample variances, and n_1 and n_2 are the sample sizes. When dealing with paired samples which involves a direct correlation between two samples the equation for the t-test becomes:

$$t = \frac{\bar{x}_d - \mu_d}{\frac{s_d}{\sqrt{n}}} \qquad \text{(equation 12)}$$

In equation 12, $\bar{x}_d$ and s_d represent the sample mean and the standard deviation of the differences, μ_d is the hypothesized value, and n is the number of sample differences (33, 34, 35).

Null Hypothesis

A widely used statistical method for testing the validity of a hypothesis involves the use of what is called the "null hypothesis" which is generally designated as H_o. In this test, one compares the experimental results with an accepted value. If these values are in agreement then the null hypothesis is accepted. However, if these results are significantly different, then the null hypothesis must be rejected in favor of an alternative hypothesis, designated as H_a (36, 37, 38). The null hypothesis may involve the use either the Student's t-test or chi square test. Both of these tests involve the use of critical value tables in determining the probability. From this, one can determine whether the null hypothesis will be accepted or rejected.

Limitations of Statistics

The phrase "There are three kinds of lies: lies, damned lies, and statistics." was made popular by Mark Twain in his autobiography published in 1906. The exact origin of the phrase, however, remains in question, but skepticism regarding the use of statistics can be traced back as far as 1891 (39). Misconceptions and/or misuse of statistics include: hypothesis testing, percentage method, sampling sizes, random sampling methods, statistical uncertainty, statistical analysis studies, the use of questionnaire studies, objectives versus data collection, the need for positive results, and the use of statistics for persuasion. The misuse of statistics can therefore be attributed to several factors including: the lack of technical expertise, biased opinions, inadequate data collection methods, and the flawed evaluation of experimental data (40, 41).

Objectives:

It is envisioned that this unit will enable high school biology students to understand, calculate and utilize a number of statistical analysis techniques in determining the exactness of experimental data. They will be able to evaluate the variability within a sample/population as well as between the data collecting instruments or sensors. In addition, this unit is aligned to both the School District of Philadelphia's standardized core curriculum for biology (42) as well as a number of Pennsylvania academic standards for science and technology (43). More specifically, targeted standards are included within: 3.1.10 "Unifying Themes," 3.2.10 "Inquiry and Design," and 3.7.10 "Technological Devices" and are delineated in the appendix.

Strategies:

A key element in this unit is having students work collaboratively in conducting experiments and analyzing the data obtained using a variety statistical techniques. To attain this goal, the students will use graphing calculators, a variety of Vernier sensors, and such computer software programs as Logger Pro and Microsoft Excel.

In an ongoing effort to improve student achievement throughout the district, the School District of Philadelphia adopted an initiative for all high schools. The six step plan includes the following teaching strategies. With the first strategy, students are expected to preview content specific vocabulary on a daily basis. They are also expected to be able to preview, analyze, and connect material presented in textbooks. The remaining strategies include: reciprocal teaching, the ability to summarize material, the use of comprehension connectors or graphic organizers, and the ability to take notes. Parts of this unit will also necessitate the use of cooperative learning strategies which has been a successful pedagogical strategy for many years. The benefits of which have been shown to increase scholastic achievement, improve social skills, as well as team self-esteem. In order for cooperative learning to be an effective teaching strategy, deliberate care must be used in evaluating its ideal classroom design. There are six basic factors that one needs to consider in establishing and maintaining an effective cooperative learning environment. These factors include: team organization, cooperative management, the will to cooperate, the skill to cooperate, basic practices, and structuring the cooperative lesson. A synopsis of each will be presented in the presented in the following paragraphs (44).

From past experiences, team organization tends to be most effective when there is academic heterogeneity among the students rather than random selection. Academic heterogeneity allows for the establishment of teams or groups each of which contains students with high, average, and below average scholastic ability. Administering an entrance test the first week of school is extremely useful in this regard. Groups consisting of no more than four students have been ideal for a variety of reasons. Lateness and absences are real concerns for most of the high schools within the School District of Philadelphia. With four students in a group, individual groups can still function even when half of the students in any one group are absent. From the standpoint of classroom management, teacher determined learning groups tend to eliminate or diminish behavioral problems associated with those groups which were determined by the students.

Classroom management is essential to an effective cooperative learning environment. This can be accomplished through: cooperative management, the will to cooperate, and the skill to cooperate. It is imperative that students understand the guidelines for acceptable classroom behavior. For example, teachers must establish consistency in dealing with unacceptable noise level within the classroom. The will to cooperate is developed over time and is based on positive social interactions and pride within the group. The skill to cooperate is based on the ability of the students to assume specific roles within the group, listen to, and work with each other.

The basic practices inherent to cooperative learning include a number of behavioral skills which include: simultaneous interaction, positive interdependence, and individual accountability. Within a cooperative learning environment, the students are encouraged to interact with members within the group. This freedom is usually not permissible within a traditional classroom setting. Positive interdependence comes from the achievement of individual students within the group and from the entire group as a whole. Individual accountability can be addressed with the aid of a variety of assessments. For instance, students can be given individual grades for a project, or they can be made aware of their part of a group grade.

Effective classroom management depends, in large part, upon the structure of the lesson. Not only does it involve the arrangement of the students within the group, but it is also dependent upon the manner in which individual lessons are designed and presented. These structures, designs, or activities are meant to improve such areas as team building, information sharing, thinking skills, communication skills, and content mastery. A brief list of classroom structures and lesson designs include: brainstorming, jigsaw, numbered heads together, rally table, round robin, roundtable, student teams achievement division (STAD), team projects, and think pair share. A detailed review of each activity can be found in <u>Cooperative Learning</u> (45).

By improving their note taking skills, students should be able to utilize, practice, and/or engage in summarizing, comprehension connectors, and structured note taking. For those reasons, I intend to teach my students the highly successful method of note taking that was developed by Walter Pauk, an English professor at Cornell University in the 1950's. The Cornell Method, as it is referred to, involves writing a key word, phrase, or concept on the left hand side of a sheet of paper. In a column, on the right hand side of the sheet of paper, relevant material about the concept is written in short sentences or phrases. Finally, at the bottom of the page, the material listed is then summarized into a short paragraph. This widely used method enables students to improve their skills in summarizing material presented in both lecture and written form (46).

In order to address and improve reading comprehension, my students will participate in reciprocal teaching techniques. This is another cooperative learning activity which is designed to encompass four skills: summarizing, questioning, clarifying, and predicting. Each student within the group will be responsible for reading a specific section within their textbook or assigned reading material, summarizing that material, and reporting out to the rest of his or her group. This pedagogical strategy has been reported to be successful in both small groups as well as in large classroom settings (47).

Classroom Activities:

As previously stated, the primary goal of this curriculum unit is to introduce students to a number of commonly used statistical methods in biology. While doing so, the students will be able to evaluate the variability in data collection and the validity of their conclusions based on the experimental data.

Lesson 1 – Calculating the Average Mass of Beans

The first two weeks of the School District of Philadelphia's standardized curriculum for biology is devoted to a review of the scientific method. Among the behavioral objectives for this unit is the ability of the students to perform experiments with appropriate laboratory equipment, use metric units and significant figures in measurements, and critique experimental data. In this experiment, the students will work in collaborative groups of four. Each group will be given fifty beans of four different varieties of *Phaseolus vulgaris*. It has been reported that there are in excess of 2 500 varieties within this genus and species (48). The beans to be examined include: red kidney beans, black beans, pinto beans, and white beans. For each variety, the students will divide the beans into ten equal samples. Using a digital, electronic balance the students will obtain the mass for each sample for the four different beans thereby having a total of forty different measurements. With the use of a graphing calculator, such as Texas Instruments' popular TI-83, one can calculate the following statistical data: the arithmetic mean, the sum of all samples, the sum of all of the samples squared, the sample standard deviation, the population standard deviation, and the number of samples, the minimum sample value, the median value, and the maximum sample value. For reporting the data, the students will generate a data table which will include their raw data and the statistical information obtained from the graphing calculator for each different bean variety. The students will then compare their results with those of the other groups. From the results of this activity, the students will be able to examine the degree of variability between different varieties of beans as well as the variability in measurements between different groups of students.

Lesson 2 – Measuring the Rate of Diffusion of Various Solutions

Unit three of the standardized biology curriculum, Cell Structure and Function, encompasses a five week time frame as outlined in the Planning and Scheduling Timeline for Biology (49). During this time, the students will be able to examine the structural composition of various cellular organelles and their functional relationships. A major portion of this unit is devoted to the structure and function of cellular membranes. This activity will focus on the transport of materials across a semi-permeable membrane. The title of the experiment is Diffusion through Membranes and it was taken from Biology with Computers (50). The basic procedure involves filling three dialysis bags with different sodium chloride solutions. The bags are then placed in distilled water with a Vernier conductivity probe. Students will be able to measure the conductivity of the different sodium chloride solutions (1%, 5%, and 10%) over a one minute period. The conductivity probe is connected to a laptop computer via a Vernier interface. After completion of the experiment the data will be plotted and the statistical data obtained using Vernier Logger Pro software. Students will be able to graph the conductivity versus time, calculate the slope, and determine the rate of diffusion for each solution. They will also be able to calculate relative differences in the rate of diffusion between the solutions.

Lesson 3 – Photosynthesis and Cellular Respiration

Typically, Unit 4: Cellular Respiration and Photosynthesis, covers about four weeks from the end of November up to the end of December. This laboratory activity is designed to supplement that unit. In this experiment, which again was taken from Biology with Computers, the students will be able to compare the rate of photosynthesis or respiration in spinach leaves depending on the presence or absence of light (51). Using Vernier oxygen gas sensors, the students will be able to measure the release of oxygen when photosynthesis is taking place in the presence of light according to the following equation:

$$6\ H_2O + 6CO_2 + \text{light energy} \rightarrow C_6H_{12}O_6 + 6\ O_2 \qquad \text{(equation 11)}$$

Conversely, they will be able to measure the decrease in oxygen due to its consumption during respiration in the dark as given by the following equation:

$$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ H_2O + 6\ CO_2 + \text{energy} \qquad \text{(equation 12)}$$

Students will be able to compare the variability of their results with other groups within the class and suggest possible reasons for the variability. They will also be able to calculate the rates of either photosynthesis or respiration by using

Lesson 4 – Using Chi-Square for Data Analysis

Genetics is one area in which statistical analysis has had a major role in the biological sciences. Consequently, the unit on Mendelian genetics which generally occurs towards the end of January and lasts until the middle of February would be the ideal opportunity to introduce students to statistical significance and the concept of the null hypothesis. This laboratory activity entitled Using the Chi-Square Test for Statistical Analysis of Experimental Data was taken from the Advanced Placement Biology Lab Manual (52). This lab is designed so that students will be able to calculate the chi-square for two experimental conditions. Using a table of critical values and determining the degrees of freedom within each condition, the students will be able to accept or reject the null hypothesis. If the null hypothesis is rejected, the students will be required to offer an alternative hypothesis.

Annotated Bibliography:

Cited References

1. The School District of Philadelphia, Planning and Scheduling Timeline for Science: Grade 10 Biology, 2006.
The School District of Philadelphia outlines the sequence of units to be covered, the time allocated for each unit, and the resources to be used in teaching each unit.

2. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 7-14.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

3. Brase, Charles Henry, and Brase, Corrinne Pellillo, Understandable Statistics 6th edition, Boston: Houghton Mifflin Company, 1999, 4-18.
The authors provide many useful guided exercises in this introductory statistics textbook.

4. Peck, Roxy, Olsen, Chris, and Devore, Jay, Introduction to Statistics and Data Analysis 2nd edition, Belmont, CA: Brooks/Cole, 2005. 1-5.
Of the textbooks used in the creation of this curriculum unit, Introduction to Statistics and Data Analysis 2nd edition seemed to be the most advanced.

5. "Statistics." Wikipedia [Web]. April, 2011 <http://en.wikipedia.org/wiki/Statistics>
This was a fairly good introduction into the history and scope of statistics.

6. Fienberg, Stephen E., "A Brief History of Statistics in Three and One-Half Chapters: A Review Essay." April, 2011, <http://www.jstor.org/stable/2246307>
In this review essay, Stephen Feinberg reviews seven books covering three and one-half centuries of statistics and probability.

7. Fienberg, Stephen E., “A Brief History of Statistics in Three and One-Half Chapters: A Review Essay.” April, 2011, <http://www.jstor.org/stable/2246307>
In this review essay, Stephen Feinberg reviews seven books covering three and one-centuries of statistics and probability.

8. “Statistics.” Wikipedia [Web]. April, 2011 <http://en.wikipedia.org/wiki/Statistics>
This was a fairly good introduction into the history and scope of statistics.

9. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 16-18.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

10. Brase, Charles Henry, and Brase, Corrinne Pellillo, Understandable Statistics 6th edition, Boston: Houghton Mifflin Company, 1999, 15-16.
The authors provide many useful guided exercises in this introductory statistics textbook.

11. Peck, Roxy, Olsen, Chris, and Devore, Jay, Introduction to Statistics and Data Analysis 2nd edition, Belmont, CA: Brooks/Cole, 2005, 8.
Of the textbooks used in the creation of this curriculum unit, Introduction to Statistics and Data Analysis 2nd edition seemed to be the most advanced.

12. “Statistics.” Wikipedia [Web]. April, 2011 <http://en.wikipedia.org/wiki/Statistics>
This was a fairly good introduction into the history and scope of statistics.

13. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 67-87.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

14. Brase, Charles Henry, and Brase, Corrinne Pellillo, Understandable Statistics 6th edition, Boston: Houghton Mifflin Company, 1999, 89-94.
The authors provide many useful guided exercises in this introductory statistics textbook.

15. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 79-80.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

16. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 80-81.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

17. “Mean.” Wikipedia [Web]. April, 2011 <http://en.wikipedia.org/wiki/Mean>
This was an advanced coverage of statistical mean.

18. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 88-104.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

19. Brase, Charles Henry, and Brase, Corrinne Pellillo, Understandable Statistics 6th edition, Boston: Houghton Mifflin Company, 1999, 102-119.
The authors provide many useful guided exercises in this introductory statistics textbook.

20. Pagano, Marcello, and Gauvreau, Kimberlee, Principles of Biostatistics, Belmont, CA: Duxbury Press, 1993, 15.
Although quite outdated for a biostatistics textbook, it still had some useful fundamental statistical concepts.

21. Pagano, Marcello, and Gauvreau, Kimberlee, Principles of Biostatistics, Belmont, CA: Duxbury Press, 1993, 15-24.
Although quite outdated for a biostatistics textbook, it still had some useful fundamental statistical concepts.

22. Brase, Charles Henry, and Brase, Corrinne Pellillo, Understandable Statistics 6th edition, Boston: Houghton Mifflin Company, 1999, 33-74.
The authors provide many useful guided exercises in this introductory statistics textbook.

23. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 24-66.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

24. Devore, Jay, and Peck, Roxy, Statistics: The Exploration and Analysis of Data 4th edition, Pacific Grove, CA: Brooks/Cole, 2001, 41-87.
I found this textbook to be better written and more basic than the other Peck, et al textbook used in this curriculum unit.

25. "Selected Topics in Graphical Analytic Techniques." June, 2011 <http://statsoft.com/textbook/graphical-analytic-techniques/>
This was an extremely useful and detailed account of graphical analysis techniques.

26. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 17-18.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

27. Brase, Charles Henry, and Brase, Corrinne Pellillo, Understandable Statistics 6th edition, Boston: Houghton Mifflin Company, 1999, 16.
The authors provide many useful guided exercises in this introductory statistics textbook.

28. Peck, Roxy, Olsen, Chris, and Devore, Jay, Introduction to Statistics and Data Analysis 2nd edition, Belmont, CA: Brooks/Cole, 2005, 8.
Of the textbooks used in the creation of this curriculum unit, Introduction to Statistics and Data Analysis 2nd edition seemed to be the most advanced.

29. "Statistics." Wikipedia [Web]. April, 2011 <http://en.wikipedia.org/wiki/Statistics>
This was a fairly good introduction into the history and scope of statistics.

30. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 357-375.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

31. "Chi Square." June, 2011 <http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage28.htm>
This article not only explained chi square, but a useful self assessment worksheet was also provided.

32. "The Chi Square Statistic." June, 2011<http://math.hws.edu/javamath/ryan/ChiSquare.html>
This was also a good article on chi square and its applications.

33. Devore, Jay, and Peck, Roxy, Statistics: The Exploration and Analysis of Data 4th edition, Pacific Grove, CA: Brooks/Cole, 2001, 376-428.
I found this textbook to be better written and more basic than the other Peck, et al textbook used in this curriculum unit.

34. "Student's t-Test." Wikipedia [Web] June, 2011 <http://en.eikipedia.org/wiki/Student's_t-test>
Wikipedia did a nice job in presenting the history and uses of the Student's t-test.

35. "Student's t-Test of Difference of Means." June, 2011 <http://faculty.chass.ncsu.edu/garson/PA765/ttest.htm>
This was a useful albeit more difficult article to understand.

36. "Null Hypothesis." Wikipedia [Web] May, 2011 <http://en.wikipedia.org/wiki/Null_hypothesis>
This was a rather basic article on the null hypothesis which made it a little easier to follow.

37. Johnson, Dale M., Probability and Statistics, Cincinnati: South-Western Publishing Co., 1989, 234-235.
Although this is a rather old textbook, it is a good introductory level statistics textbook.

38. Devore, Jay, and Peck, Roxy, Statistics: The Exploration and Analysis of Data 4th edition, Pacific Grove, CA: Brooks/Cole, 2001, 352-354.
I found this textbook to be better written and more basic than the other Peck, et al textbook used in this curriculum unit.

39. "Lies, Damned Lies, and Statistics." Wikipedia [Web] June, 2011 <http://en.wikipedia.org/wiki/Lies_damned_lies,_and_statistics>
This was an interesting article into the history surrounding the phrase "Lies, damned lies, and statistics.

40. “Statistics.” Wikipedia [Web]. April, 2011 <http://en.wikipedia.org/wiki/Statistics>
This was a fairly good introduction into the history and scope of statistics.

41. Bhalerao, J. V. “Use Of Statistics And Misconceptions About Its Uses Amongst Users Of Statistics.” June, 2011 <www.researchersworld.com>
There were some grammatical errors in this article, but all things considered I found it very informative regarding the misconceptions surrounding the use of statistics.

42. The School District of Philadelphia, Planning and Scheduling Timeline for Science: Grade 10 Biology, 2006.
The School District of Philadelphia outlines the sequence of units to be covered, the time allocated for each unit, and the resources to be used in teaching each unit.

43. Pennsylvania Department of Education Pennsylvania Teacher’s Desk Reference: A Critical Thinking Guide, Jacksonville, FL: Educational Tools, Inc, 2006.
This reference guide details Pennsylvania’s academic standards for science and Technology for grades 9-12. It also addresses the assessment anchors and teaching strategies for each academic standard.

44. Kagan, Spencer. Cooperative Learning. San Juan Capistrano, CA: Kagan Cooperative Learning, 1992, 4:1-4:9.
Spencer Kagan’s book on cooperative learning is an excellent resource for developing and sustaining a cooperative learning environment in the classroom

45. Kagan 7:1-8:2, 10:1-10:17
Spencer Kagan’s book on cooperative learning is an excellent resource for developing and sustaining a cooperative learning environment in the classroom

46. “Cornell Notes.” *Wikipedia* [Web]. Retrieved April, 2007 from <http://en.wikipedia.org/wiki/Cornell_Notes>,
This article briefly explains the Cornell Method of note taking as developed by Walter Pauk.

47. North Carolina Regional Educational Laboratory. “Reciprocal Teaching.” Retrieved April, 2007 from <http://www.ncrel.org/sdrs/areas/issues/students/atrisk/at6lk38.htm>
The benefits of reciprocal teaching are discussed as well as the steps involved

48. “Phaseolus Vulgaris.” May, 2011<http://www.floridata.com/ref/p/phas_vul.cfm>
This brief article covers the description, location, cultivation, usage, and features of a number of common beans.

49. The School District of Philadelphia, Planning and Scheduling Timeline for Science: Grade 10 Biology, 2006.
The School District of Philadelphia outlines the sequence of units to be covered, the time allocated for each unit, and the resources to be used in teaching each unit.

50. Holman, Scott, and Masterman, David, Biology with Computers 2nd edition, Beaverton, OR: Vernier Software and Technology, 2000, 4:1-4:2T.
This laboratory exercise involves examining the diffusion of substances across membranes.

51. Holman, Scott, and Masterman, David, Biology with Computers 2nd edition, Beaverton, OR: Vernier Software and Technology, 2000, 31A:1-31B:2T.
In this laboratory exercise, the students will examine the rate of photosynthesis or respiration.

52. Advanced Placement Lab Manual, College Entrance Examination Board, 2001, 85-89.
The activity is this laboratory manual is designed to examine the use of chi square.

Student Resources:

Johnson, George, and Raven, Peter, Biology, Austin: Holt, Rinehart and Winston, 2004.
This is the School District of Philadelphia's recommended textbook intended for use with the standardized curriculum for biology.

Damon, Alan, McGonegal, Randy, Tosto, Patricia, and Ward, William, Standard Level Biology Developed Specifically for the IB Diploma, Upper Saddle River, NJ: Pearson, 2007.
This is the textbook which is directly aligned with the standard level International Baccalaureate biology program.

Campbell, Neil A., and Reece, Jane B, AP Edition Biology 7th edition, San Francisco: Pearson, 2005.
The Campbell and Reece textbook is accepted by many as the textbook for the advanced placement biology course.

Holman, Scott, and Masterman, David, Biology with Computers 2nd edition, Beaverton, OR: Vernier Software and Technology, 2000.
This lab manual will be used by the students for a few of the activities in this unit.

Advanced Placement Lab Manual, College Entrance Examination Board, 2001.
The advanced placement lab manual will be used for the chi square lab.

Appendix/Standards:

The Pennsylvania academic standards for science and technology that will be addressed during the course of this curriculum unit were taken directly from the Pennsylvania Teacher's Desk Reference and Critical Thinking Guide and include:

3.1.10 Unifying Themes: There are only a few fundamental concepts and processes that form the framework upon which science and technology are organized – motion and forces, energy, structure of matter, change over time and machines. These themes create the context which the content of the disciplines can be taught and are emphasized in each standard.

B. Describe concepts of models as a way to predict and understand science and technology.
   - Apply mathematical models to science and technology.

D. Apply scale as a way of relating concepts and ideas to one another by some measure.
   - Apply dimensional analysis and scale as a ratio.
   - Convert one scale to another.

E. Describe patterns of change in nature, physical and man made systems.
   - Describe the effects of error in measurements.

3.2.10 Inquiry and Design: The nature of science and technology is characterized by applying process knowledge that enables students to become independent learners. These process skills are developed across grade levels and differ in the degree of sophistication, quantitative nature and application to the content.

B. Apply process knowledge and organize scientific and technological phenomena in varied ways.
   - Describe materials using precise quantitative and qualitative skills based on observations.
   - Develop appropriate scientific experiments: raising questions, formulating hypotheses, testing, controlled experiments, recognizing variables, manipulating variables, interpreting data, and producing solutions.
   - Use process skills to make inferences and predictions using collected information and to communicate, using space/time relationships, defining operationally.

C. Apply the elements of scientific inquiry to solve problems.
   - Organize experimental information using a variety of analytical methods.
   - Judge the significance of experimental information in answering the question.
   - Suggest additional steps that might be done experimentally.

D. Identify and apply the design process to solve problems.
   - Propose and analyze a solution.
   - Evaluate the solution, test, redesign and improve as necessary.
   - Communicate the process and evaluate and present the impacts of the solution.

3.7.10 Technological Devices: Students use tools to observe measure, move and make things. New technological tools and techniques make it possible to enact far-reaching changes in our world. Technology enhances the students' abilities to identify problems and determine solutions. Computers play an integral role in every day life by extending our abilities to collect, analyze and communicate information and ideas.

B. Apply appropriate instruments and apparatus to examine a variety of objects and processes.
   - Compare and contrast different scientific measurement systems; select the best measurement system for a specific situation.
   - Explain the need to estimate measurements within error of various instruments.
   - Apply accurate measurement knowledge to solve everyday problems.

C. Apply basic computer operations and concepts.
   - Analyze and solve basic operating systems problems.

D. Utilize computer software to solve specific problems.

- Apply advanced graphic manipulation and desktop publishing techniques.
- Apply basic multimedia applications.
- Apply advanced word processing, database and spreadsheet skills.