AP Biology Quantitative Skills:

A Guide for Teachers

The College Board
New York, NY

ABOUT THE COLLEGE BOARD

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators and schools.

For further information, visit www.collegeboard.org.

EQUITY AND ACCESS POLICY STATEMENT

The College Board strongly encourages educators to make equitable access a guiding principle for their AP programs by giving all willing and academically prepared students the opportunity to participate in AP. We encourage the elimination of barriers that restrict access to AP for students from ethnic, racial and socioeconomic groups that have been traditionally underserved. Schools should make every effort to ensure their AP classes reflect the diversity of their student population. The College Board also believes that all students should have access to academically challenging course work before they enroll in AP classes, which can prepare them for AP success. It is only through a commitment to equitable preparation and access that true equity and excellence can be achieved.

Pages 16, 17, 44, 45: Graphs from *AP Biology Investigative Labs: An Inquiry-Based Approach.* (New York: The College Board, 2012).

Pages 18, 19, 20, 46, 47, 48, 77, 96, 97: Questions from *AP Biology Practice Exam.* (New York: The College Board, 2012).

1 2 3 4 5

Contents

Acknowledgments

The College Board would like to acknowledge the following individuals for their commitment and dedication toward the completion of this project:

- Elizabeth Carzoli, Castle Park High School, Chula Vista, CA
- John Jungck, University of Delaware, Newark, DE
- Kim Foglia, Division Avenue Senior High School, Levittown, NY
- Sharon Radford, The Paideia School, Atlanta, GA
- Paul Rodriguez, Troy High School, Fullerton, CA
- Jim Smanik, Sycamore High School, Cincinnati, OH
- Gordon Uno, University of Oklahoma, Norman, OK
- Brad Williamson, University of Kansas, Lawrence, KS
- Julianne Zedalis, The Bishop's School, La Jolla, CA

Introduction

Problem Solving and Quantitative Skills in AP Biology

The BIO2010 report of the National Research Council (2003) describes the current trend in biology and biological research as follows:

> *Biological concepts and models are becoming more quantitative, and biological research has become critically dependent on concepts and methods drawn from other scientific disciplines. The connections between the biological sciences and the physical sciences, mathematics, and computer science are rapidly becoming deeper and more extensive.*

Therefore, it is imperative that today's students develop and apply quantitative skills as part of their exploration into biology. A good grasp of quantitative methodology and reasoning is particularly important in the laboratory experience. In fact, many statistical tests and other mathematical tools were developed originally to work out biological problems. *America's Lab Report: Investigations in High School Science* (National Research Council, 2005) sums it up:

> *Laboratory experiences provide opportunities for students to interact directly with the material world (or with data drawn from the material world), using the tools, data collection techniques, models, and theories of science.*

The new AP Biology curriculum framework; AP Biology lab manual, *AP Biology Investigative Labs: An Inquiry-Based Approach*; and AP Exam reflect this emphasis on inquiry and reasoning and the quantitative skills these processes entail. In their laboratory experience, students will now model the behavior of scientists. As they carry out independent investigations, students will observe, explore, and discover knowledge themselves as they engage in scientific problem solving.

The focus of the labs has shifted from students achieving predetermined results to students making independent observations, posing their own questions, and determining how to investigate these questions. Students now must choose which variables they want to investigate (formerly, they often were told what they would investigate). They then conduct lab and field investigations and gather their own data. They must determine the appropriate way to record those data. (Formerly, tables and graphs were provided for students to fill out). They go on to analyze the results of their data gathering, test hypotheses, and use mathematical modeling. (Formerly, the necessary mathematical equations and other quantitative skills would be provided with step-by-step instructions on how to use them). Finally, students communicate the results and conclusions of their laboratory investigations the way scientists all over the world do—through peer review, mini-posters, and presentations, in addition to the more traditional lab notebooks, portfolios, reports, and papers.

In the new laboratory environment, students are directed to resources, which they then must decide how to apply. The focus is on helping students develop the skills they need for using statistics and other techniques to inform their investigations. In so doing, they will develop the reasoning skills essential to good scientific practice.

SCIENCE PRACTICES AND QUANTITATIVE SKILLS

Practice can be defined as a way to coordinate knowledge and skills to accomplish a goal or task. The new AP Biology curriculum framework is based on seven science practices or necessary skills for the biology student of the 21st century. It is impossible to imagine engaging in any one of the science practices without a good grasp of quantitative skills.

For example, the first science practice states, *The student can use representations and models to communicate scientific phenomena and solve scientific problems.* By creating and interpreting graphs, or visual representations, students can illustrate biological concepts and ideas, communicate information, make predictions, and describe systems to promote and document understanding. The second science practice says, *The student can use mathematics appropriately.* It expresses the fact that quantitative reasoning is crucial in solving problems, analyzing experimental data, describing natural phenomena, making predictions, and describing processes symbolically. The fifth science practice, *The student can perform data analysis and evaluation of evidence,* recognizes that quantitative analysis is involved in determining whether the data support a conclusion, drawing conclusions from a data set, assessing the validity of experimental evidence, identifying possible sources of error in an experimental procedure or data set, identifying outliers, and proposing explanations for them. The sixth science practice, *The student can work with scientific explanations and theories*, points to the importance of hypothesis testing and mathematical modeling in scientific inquiry. The other science practices similarly involve the use of the quantitative skills described in this guide.

Perhaps the biggest challenge AP Biology instructors will face in implementing the new curriculum will be getting students to adopt the scientific inquiry method and its language. For example, students who may be used to "proving" that a hypothesis is true or false may struggle as they begin to work with the more scientifically accurate statements *reject the null hypothesis* or *fail to reject the null hypothesis*. In addition, students may or may not already have been introduced to the necessary quantitative skills in mathematics classes. And even students with the educational background to support the skills probably will not be familiar with their application to biological inquiry in the "real world."

HOW TO USE THIS GUIDE

This guide is an introduction to thinking about quantitative skills as they are used to help answer questions about biological phenomena. It explains how biology instructors can help students learn and apply four of the quantitative skills that students will rely on most heavily in their exploration of biology in the AP Biology curriculum framework: graphing, data analysis, hypothesis testing, and mathematical modeling.

Chapter 1: Graphing introduces the graphical representation of data to allow students to discover patterns or relationships in that information and infer causal biological mechanisms.

Chapter 2: Data Analysis helps students discover meaningful patterns in masses of data as they develop their skills in statistics. This quantitative skill helps students understand and deal with the variability and the sometimes ambiguous nature of data that are typical in scientific investigations.

Chapter 3: Hypothesis Testing covers the application of statistical tests to help distinguish between working hypotheses.

Finally, **Chapter 4:** Mathematical Modeling helps students understand complex systems, explore various possibilities in those systems, develop conceptual frameworks, make accurate predictions, and generate explanations. When students learn to model, they improve their ability to solve problems, understand how they know what they know, and transfer that insight to other subject areas.

Each of the four skills covered in this guide means something a little different depending on the field of study. This guide limits the discussion of data analysis primarily to descriptive statistics with a few specific examples of inferential statistics. Inferential statistics are tools that allow one to infer, or conclude, something about the true population—the population beyond the sample. The discussion of hypothesis testing is limited to a small sample of statistical tests and techniques that may be useful to the AP Biology student. Each chapter begins with a list of labs in *AP Biology Investigative Labs: An Inquiry-Based Approach* that use the skills covered in the text. An introductory explanation includes components of the quantitative skill, key terms and their definitions, a discussion of how the skill applies to biology, and information on how the skill can be practiced and used both in the classroom and in the wider world of biological inquiry. Next, examples illustrate the application of the quantitative skill in a lab setting. These examples reflect the type of analysis needed for most of the investigations in the lab manual and follow an imagined run-through to demonstrate the use of the quantitative skill in a lab setting. Preliminary data are provided for students to explore and analyze in order to discover patterns and make observations that can inform their hypotheses and experimental design. Students will then be ready to design and carry out an investigation using the same or similar techniques or procedures in conjunction with a question of their choice. Working through these examples before doing the actual lab allows students to get a feel for the process. Finally, each chapter ends with a list of resources that the instructor may tap into for classroom activities and further examples.

Every laboratory in *AP Biology Investigative Labs: An Inquiry-Based Approach* encourages students to carry out independent investigations based on their own questions that arise after exploring a new technique or method in order to learn more about a biological concept. The resources in this guide will help students to develop good experimental designs to answer those questions, as well as to use quantitative methods to make sense of the results—a critical requirement for drawing valid conclusions. Of course, because students are asking their own questions, some of their investigations will require techniques and skills not covered here. In those cases, the teacher and student will need to build on the approaches presented here to explore further tests and quantitative methods.

This is not meant to be a comprehensive guide to teaching quantitative skills. Instead, this document will provide enough background to allow you to seek out appropriate tools and methods suitable to your own classroom. One place to start seeking out those resources is the NIMBioS website: http://www.nimbios.org/. Furthermore, the information in this guide is not meant to replace the quantitative skills that have been emphasized in previous AP Biology labs, such as preparing buffers by use of logarithms, mathematically modeling the rate of enzymatic actions, and considering the magnification in microscopy. Rather, the information here is intended to help you in the transition to the new AP Biology teaching framework, which includes thinking conceptually about analyzing data, testing hypotheses, and building mathematical models to discriminate among possible causes. As such, the goal of this manual is to guide you in planting the seeds for an intuitive understanding of statistics and mathematical modeling in your students. Be sure to explore the resources at the end of each of the skills chapters to further your own exploration into the new curriculum.

Note that nearly all of the formulas required for the examples in this resource guide are on the AP Biology equations and formulas list, which was created to help teachers and students design their instruction and prepare for the exam (see the appendix). The intent of this guide is to provide a context for understanding how this kind of quantitative thinking informs biology. It is not intended that students should memorize a list of formulas. This list will be available to students when they sit for the exam.

INCORPORATING QUANTITATIVE SKILLS IN THE CLASSROOM

Quantitative skills should be deeply embedded into all laboratory and content instruction as a reflection of how integral these skills are to the study of biological sciences. As such, student assessments that focus on quantitative skills alone are probably inappropriate. Instead, course and lab assessments should include evaluation of how well the appropriate quantitative skills have been applied to solving biological problems. Much of this assessment of quantitative skills should come in the form of feedback from peers and teachers to the student on the appropriate application of quantitative skills.

It is essential that the classroom environment spark an ongoing, nonthreatening dialogue between students and teachers that explores the appropriate application of various skills. This interchange will vary by laboratory, by content, and by the skills that the student and teacher bring to the investigation. Think of a casual lab lunch meeting where members of the lab present their preliminary results and analysis to the entire group. Similarly, give students multiple opportunities to polish their research and analysis based on authentic peer review before making the final presentation of their results, as described at http://www.nabt.org/blog/2010/05/04/mini-posters-authentic-peer-review-in-the-classroom/. Students who are used to more formulaic science labs may question the need for the quantitative skills covered in this guide. It may be helpful to spend some time clarifying what happens when this type of reasoning is not applied or when it is distorted deliberately. Using a catchy or local news report to illustrate how and why quantitative intelligence is important to science and to the students' own lives can provide a number of excellent avenues for discussion. Students will begin to understand why important health and governmental decisions that will affect them and their families should be based on a clear-minded assessment of the science behind the issues.

Guide your students in acquiring the ability to detect when information is credible, incomplete, or suspect. Have the students read reports and ask themselves, *What does this really mean? Does this represent a useful or meaningful result? How could I test this?* Ask students to discuss the "scientific facts" in a news item and the ways in which quantitative analysis may have been used to arrive at the statements reported. Let the students debate their ability to even answer that question. Then remind them to reflect on this discussion as they prepare to undertake their own scientific experiments.

You might want to have the students follow a story of interest to them. Many resources offer reports on, and analyses of, today's scientific issues. A search of resources on the Internet can be instructive in its own right, as students will be faced with the task of assessing which sites come with their own agendas. Emphasize the need for analysis of the sources in research reports. Steering clear of political and cultural hot points may seem challenging, but some sites offer less confrontational defenses of good scientific methods while pointing out flaws in analysis or data in published reports.

In preparation for discussion of the need for quantitative skills, you and your students might want to look at the website of the James Randi Educational Foundation (www.randi.org), which catalogs and dissects numerous examples of poor science, distorted science, and complete nonscience. An English website with a statistical take on bad science is available from Dr. Ben Goldacre, an epidemiologist, at www.badscience.net.

A lighter take on experimental design and analysis would be the *MythBusters* television series. Mixed among the overtly silly are the still very silly experiments that at least use some of the tenets of quantitative skills. Analyzing the discussions among the show's "researchers" about their failures can be a good way to illustrate to students how the wrong question sometimes gets asked in an investigation and how an entirely different question might get answered. Sparing use of these examples will keep your students interested, but caution them to avoid experiments involving catapults or cannons.

CHAPTER 1:
Graphing

LABS USING GRAPHING

1: Artificial Selection
4: Diffusion and Osmosis
5: Photosynthesis
6: Cellular Respiration
9: Biotechnology: Restriction Enzyme Analysis of DNA
12: Fruit Fly Behavior
13: Enzyme Activity

One of the best ways to communicate the results of a scientific investigation is **graphing,** or creating an effective visual representation (a **graph**) of the data that have been counted, measured, and calculated. Investigators often can easily see patterns in a carefully crafted visual display that may not be as readily apparent in a data table of numbers. Visual displays also can clarify how two measured variables affect each other. In a scientific investigation, then, one of the first steps in data analysis and data exploration is to create graphical displays of the data that reflect the questions the investigation purports to answer. Effective graphs convey summary or descriptive statistics as part of the display. This means that a certain amount of data analysis is required to produce appropriate graphical representations of experimental results. Hence, this chapter will also briefly introduce descriptive statistics that can enhance the data graphs. Chapter 2 will cover data analysis more thoroughly. The more graphing experience students gain, the more they will understand that certain plotted shapes or forms are associated with models that make it easier to infer causal mechanisms (i.e., exponential growth and decay functions).

Computer software promotes data exploration by facilitating the rapid transformation of data sets into graphs and enabling quick modifications of these graphs, thus allowing users to bypass other computational tools. All too often, though, students put all their data into a spreadsheet or a similar tool and then let the software create default line graphs or other inappropriate graph presentations. Students need to be able to determine the appropriate parameters of graphs, create the graphs, and then analyze them.

COMPONENTS OF GRAPHING AND PRELIMINARY DATA ANALYSIS

Most AP Biology investigations involve the analysis of data to answer a question. Usually it is not practical to collect data that include the entire population of interest. Instead, it makes more sense to collect a sample of data from the larger population. Most often in an AP Biology investigation, these data are quantitative and may be measurements, counts, behavioral observations, times, and so on. The sample serves as a representation of the larger population, but how well does it represent the true population? Descriptive statistics and graphical displays allow one to estimate and communicate how well sample data represent the true population in order to begin to answer the original research question.

Descriptive statistics and graphical displays are also useful for summarizing larger data sets and for presenting patterns in data. Tables, while useful for collecting and compiling data to answer a question, are often not the best way to make sense and communicate the results of an investigation. For illustration, consider a nonbiological example: exam grades. Suppose that a class of 100 AP Biology students sat for an exam that was worth 100 points. The teacher posts the anonymous results in a table of 100 unsorted numbers. With their own score and the table of results, a student could compare his or her score to others, but such a comparison would likely not be obvious. However, if instead the teacher posted a **mean** (average) score and a **standard deviation** (how variable the scores were) and presented the student scores in a graphical display known as a histogram, students could quickly compare their scores to others in the class. Likewise, the teacher would have a good summary for how the class did as a whole. The same is true for AP Biology investigations. It is all about communication. Descriptive statistics and graphical displays share the same goal of presenting and summarizing data—one with numbers and one with visual displays. Most often the two are combined when presenting the results of an investigation.

The AP Biology laboratory manual is designed to encourage students to ask their own questions by designing and carrying out investigations. This process of inquiry requires data analysis and communication of results. The data collected to answer questions generated by students will generally fall into three categories: (1) normal or **parametric data**, (2) **nonparametric data**, and (3) **frequency or count data**. Normal or parametric data are measurement data that fit a normal curve or distribution. Generally, these data are in decimal form. Examples include plant height, body temperature, and response rate. Nonparametric data do not fit a normal distribution, may include large outliers, or may be count data that can be ordered. A scale such as *big*, *medium*, *small* (qualitative) may be assigned to nonparametric data. Frequency or count data are generated by counting how many of an item fit into a category. For example, the results of a genetic cross fit this type of data as do data that are collected as percentages.

The typical questions asked in an AP Biology lab investigation can likewise be divided into two groups: those questions that compare phenomena, events, or populations (*Is A different from B?*), and those questions that look for associations between variables (*How are A and B correlated?*).

One of the challenges of designing good investigations is determining the type of data that will be needed to answer the principal question. It is far less effective to attempt to apply different analysis tools after the fact. Encourage your students to get into the practice of thinking through the investigation before taking data. Prior to starting an investigation and during the design process, they should take time to think through what kind of question is being asked and what kind of data will be required to answer it. Alternatively, a short exploratory investigation can help inform later decisions. The English statistician R.A. Fisher once said, "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

Table 1 is adapted from Neil Millar's "Simple Statistics Software for Biology Students" (http://www.heckmondwikegrammar.net/index.php?highlight=introduction&p=10310). This table is organized to inform experimental design, analysis, and communication. If the investigator knows the type of question asked and the data required to answer it, then this table will point to descriptive statistics, graphs, and test statistics that can be used. To use this table, students should first consider the kind of data that will be collected. Will it be measurement data? Will it fit a normal curve? Background research can sometimes inform these questions. If it is not "normal" data, then students need to decide whether the data collected are categorical or nonparametric. They then choose a column in the table that matches the data type. The descriptive statistics row contains suggestions for appropriate descriptive statistics to report. Students should consider the type of questions being asked in the investigation and then choose the appropriate row to determine possible graph types and statistical tests to perform. This manual will not cover all of the tests or plots listed in the table, nor is the table itself meant to be a comprehensive collection of all the tools of visual display and statistics. The purpose of the manual and the table is to provide entry-level support for choosing appropriate data analysis and display.

Understanding the five graph types used in this guide will be sufficient for getting started with the investigations in the new lab manual and the AP Biology course. **Bar graphs** are graphs used to visually compare two samples of categorical or count data. Bar graphs are also used to visually compare the calculated means with error bars of

Table 1. Statistical Tests and Graph Styles

<table>
<tr><td colspan="3"></td><td>Parametric Tests (normal data)</td><td>Nonparametric Tests</td><td>Frequency Tests (counts)</td></tr>
<tr><td colspan="3">Descriptive Statistics</td><td>Mean, Standard Deviation, Standard Error, 95% CI</td><td>Median, Quartiles, Interquartile Range</td><td>Percent in Each Category</td></tr>
<tr><td rowspan="5">Comparative Statistics</td><td colspan="2">Graph Type</td><td>Bar Graph</td><td>Box-and-Whisker Plot</td><td>Bar Graph or Pie Chart</td></tr>
<tr><td rowspan="2">Independent Samples</td><td>2 groups</td><td>Unpaired T-test</td><td>Mann-Whitney U-test</td><td rowspan="4">Chi-square Test</td></tr>
<tr><td>≥ 2 groups</td><td>Anova</td><td>Kruskal-Wallis Test</td></tr>
<tr><td rowspan="2">Matched Samples</td><td>2 groups</td><td>Paired T-Test</td><td>Wilcoxon Test</td></tr>
<tr><td>≥ 2 groups</td><td>Matched Anova</td><td>Friedman Test</td></tr>
<tr><td rowspan="3">Association Statistics</td><td colspan="2">Graph Type</td><td>Scatterplot</td><td>Scatterplot</td><td>Mosaic Chart</td></tr>
<tr><td colspan="2">Test for an Association</td><td>Pearson Correlation</td><td>Spearman Rank Correlation</td><td>Chi-square Test for Association</td></tr>
<tr><td colspan="2">Linear Relationship</td><td>Linear Regression</td><td>–</td><td>–</td></tr>
<tr><td colspan="6">Source: Redrawn from "Statistics for AS Biology," available as part of a download at: http://www.heckmondwikegrammar.net/index.php?highlight=introduction&p=10310</td></tr>
</table>

normal data (Figure 1). **Scatterplots** are graphs used to explore associations between two variables visually. **Box-and-whisker plots** allow graphical comparison of two samples of nonparametric data (data that do not fit a normal distribution). **Histograms,** or frequency diagrams, are used to display the distribution of data, providing a representation of the central tendencies and the **spread** of the data. The creation and use of **line graphs** are explained in depth in Appendix B of *AP Biology Investigative Labs: An Inquiry-Based Approach*. The other types of graphs will be explored in this chapter. See Table 1 for suggested graph types that work well with specific types of data.

Bar Graphs

Many questions and investigations in biology call for a comparison of populations. For example, Are the spines on fish in one lake without predators shorter than the spines on fish in another lake with predators? or Are the leaves of ivy grown in the sun different from the leaves of ivy grown in the shade? If the variables are measured variables, then the best graph to represent the data is probably a bar graph of the means of the two samples with standard error indicated (Figure 1). (Means and standard error will be discussed in Chapter 2.)

In Figure 1, the **sample standard error bar** (also known as the *sample error of the sample mean*) is a notation at the top of each shaded bar that shows the sample standard error (SE, in this case, ±1). Most of the time, bar graphs should include standard error rather than standard deviation (discussed in Chapter 2). The standard error bars provide more information about how different the two means may be from each other. Sample standard error bars are not particularly easy to plot on a graph, however. In Excel, for example, the user needs to choose the "custom error bar" option. An Internet search will yield links to video instructions on "how to plot error bars in Excel."

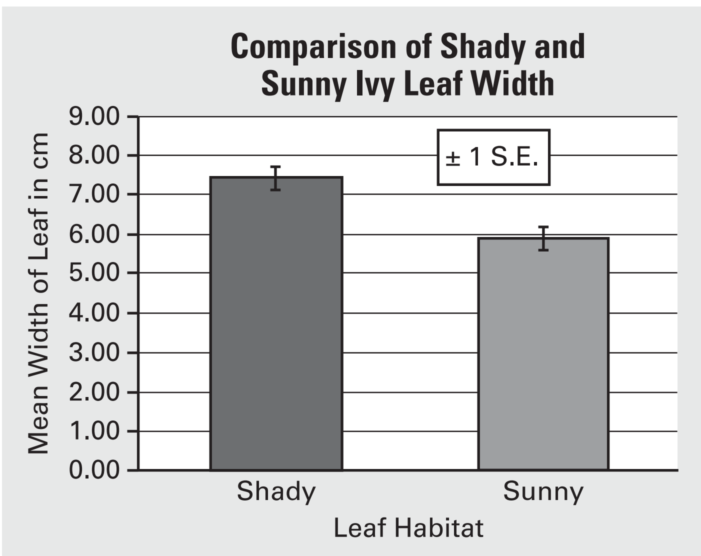


Figure 1. A Bar Graph

Figure 2 shows a variation of a bar graph that only plotted a point for each mean, but there are still standard error bars around each mean. Note that the light intensity (the independent variable, or the variable manipulated) is on the *x*-axis, and the rate of photosynthesis (the dependent variable) is on the y-axis. A standard bar graph could have been used, but the bars would have cluttered up the display. For this reason, only the means were plotted. The points imply a function between the two variables. Had a line been drawn between the points, this would have been a line graph. However, since data were not taken at all light intensities, a line is not appropriate. With more advanced mathematical techniques, one could draw a line of best fit to describe the relationship.

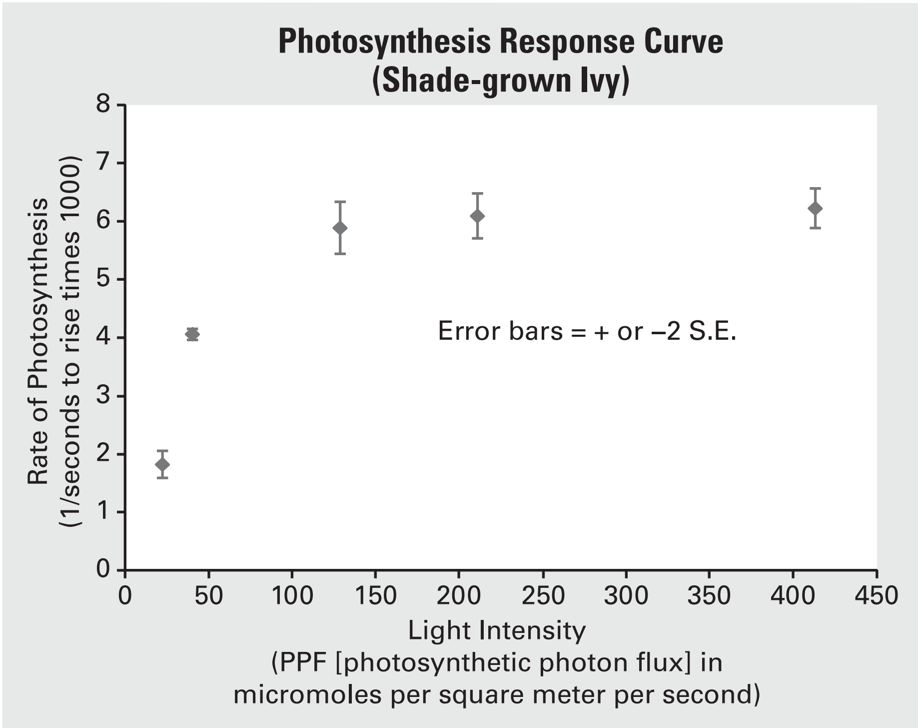


Figure 2. Variation of a Bar Graph with Only the Means Plotted

Scatterplots

When comparing one *measured* variable against another—looking for trends or associations—it is appropriate to plot the individual data points on an x-y plot, creating a scatterplot. If the relationship is thought to be linear, a linear regression line can be calculated and plotted to help filter out the pattern that is not always apparent in a sea of dots (Figure 3). In this example, the value of r (square root of R^2) can be used to help determine if there is a statistical correlation between the x and y variables to infer the possibility of causal mechanisms. Such correlations point to further questions where variables are manipulated to test hypotheses about how the variables are correlated.

Students can also use scatterplots to plot a manipulated independent x-variable against the dependent y-variable. Students should become familiar with the shapes they'll find in such scatterplots and the biological implications of these shapes. For example, a bell-shaped curve is associated with random samples and normal distributions. A concave upward curve is associated with exponentially increasing functions (for example, in the early stages of bacterial growth). A sine wave–like curve is associated with a biological rhythm.

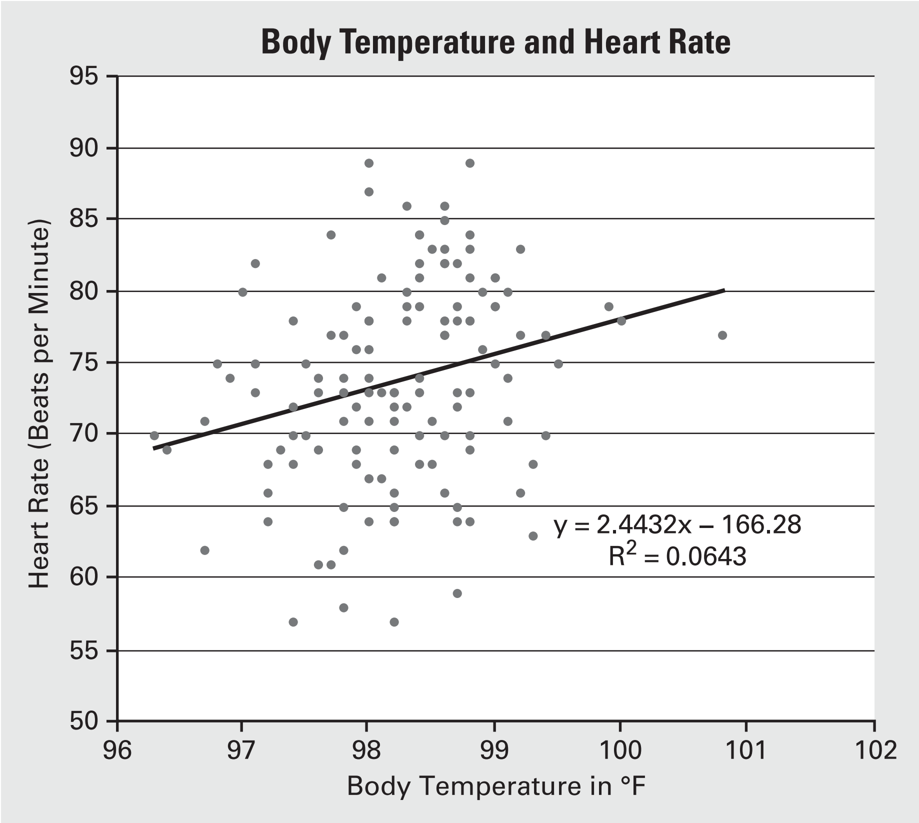


Figure 3. A Scatterplot with a Linear Regression Line

Box-and-Whisker Plots

Figure 4 shows a table with nonparametric data. (See the next section for tips on identifying nonparametric data.) The appropriate descriptive statistics are medians and quartiles, and the appropriate graph is a box-and-whisker plot (whisker plot). In the graph, the ticks at the tops and bottoms of the vertical lines show the highest and lowest values in the dataset, respectively. The top of each box shows the upper quartile, the bottom of each box shows the lower quartile, and the horizontal line represents the median. The graph allows the investigator to determine at a glance, in this case, that the ash leaves appear to decay the fastest and the beech leaves take longer to decay.

Excel and most other spreadsheet programs do not plot box-and-whisker plots automatically. If the data to be graphed are best plotted with box-and-whisker plots, use Google to find a video or instructions for making a box-and-whisker plot in Excel.

Bag	% Decay		
Number	Ash	Sycamore	Beech
1	51	40	34
2	63	33	15
3	44		26
4		52	21
5	48	48	
6	32	35	11
7	70	44	19
8	48	63	32
9	57	40	

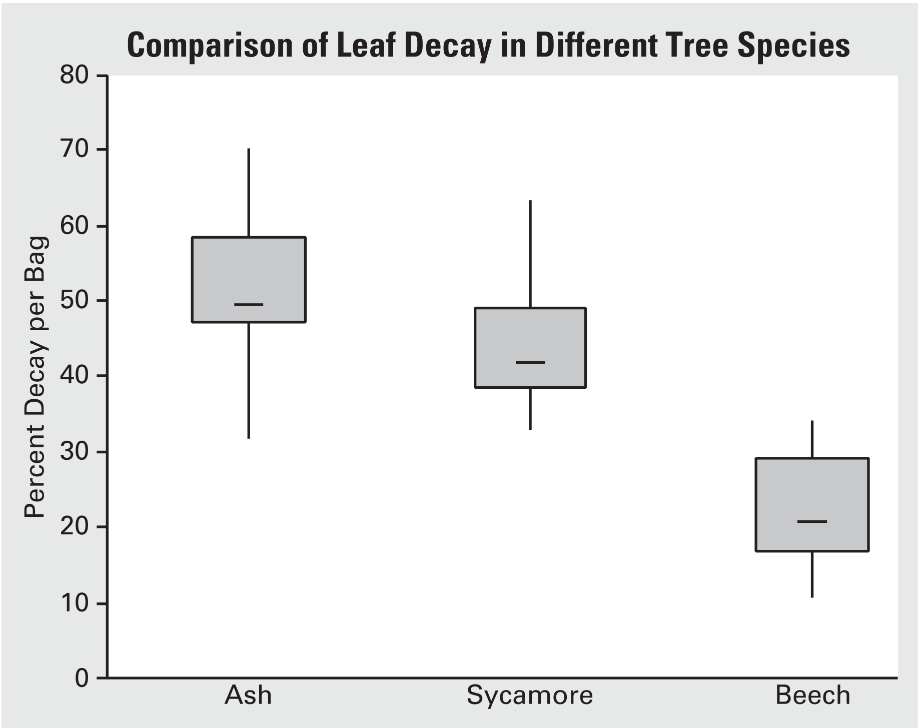


Figure 4. Nonparametric Data and Their Representation in a Box-and-Whisker Plot
[*Source: Redrawn from "Merlin_examplels.xls," available as part of a download at: http://www.heckmondwikegrammar.net/index.php?highlight=introduction&p=10310*]

Histograms

When an investigation involves measurement data, one of the first steps is to construct a histogram to represent the data's distribution to see if it approximates a normal distribution (which will be defined shortly). Creating this kind of graph requires setting up **bins**—uniform range intervals that cover the entire range of the data. Then the number of measurements that fit in each bin (range of units, shown in Figure 5) are counted and graphed on a frequency diagram, or histogram. If enough measurements are made, the data can show an approximate normal distribution, or bell-shaped distribution, on a histogram. These constitute parametric data. The normal distribution is very common in biology and is a basis for the predictive power of statistical analysis. If the data do not approximate a normal distribution (that is, they are nonparametric data), then other descriptive statistics and tests need to be applied to those data. Keep in mind, though, that even though a distribution does not reflect a perfect bell curve, it doesn't mean that the actual population is not normally distributed. The shape of the distribution could be due to sampling bias or errors. Figure 5 shows histograms with parametric and nonparametric data.

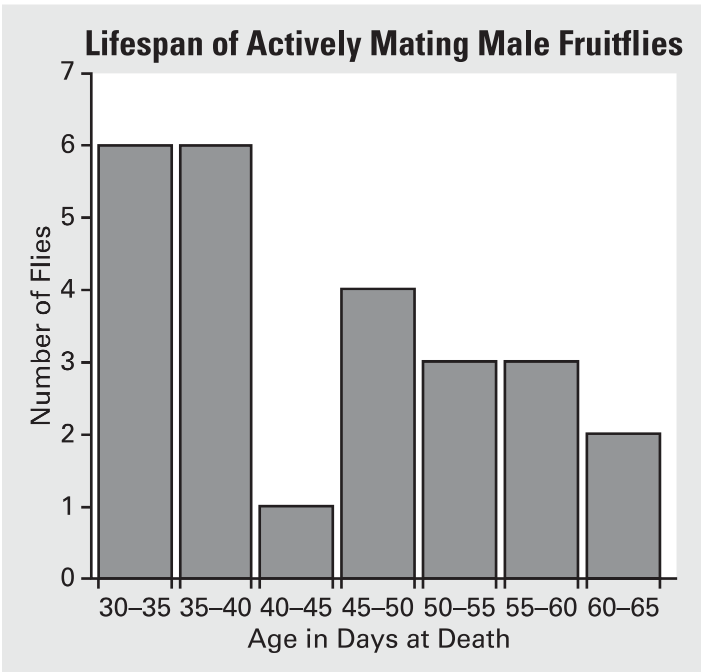


Figure 5. Histogram Showing Nonparametric Data [*Source: Hanley, James A., and Stanley H. Shapiro. "Sexual Activity and the Lifespan of Male Fruitflies: A Dataset That Gets Attention."* Journal of Statistics Education *2, no. 1 (1994).*]

ELEMENTS OF EFFECTIVE GRAPHING

Students will usually use computer software to create their graphs. In so doing, they should keep in mind the following elements of effective graphing:

- A graph must have a title that informs the reader about the experiment and tells the reader exactly what is being measured.
- The reader should be able to easily identify each line or bar on the graph.
- Axes must be clearly labeled with units as follows:
  - The x-axis shows the independent variable. Time is an example of an independent variable. Other possibilities for an independent variable might be light intensity or the concentration of a hormone or nutrient.

- The y-axis denotes the dependent variable—the variable that is being affected by the condition (independent variable) shown on the x-axis.
- Intervals must be uniform. For example, if one square on the x-axis equals five minutes, each interval must be the same and not change to 10 minutes or one minute. If there is a break in the graph, such as a time course over which little happens for an extended period, it should be noted with a break in the axis and a corresponding break in the data line.
- It is not necessary to label each interval. Labels can identify every five or 10 intervals, or whatever is appropriate.
- The labels on the x-axis and y-axis should allow the reader to easily see the information.

- More than one condition of an experiment may be shown on a graph by the use of different lines. For example, the appearance of a product in an enzyme reaction at different temperatures can be compared on the same graph. In this case, each line must be clearly differentiated from the others—by a label, a different style, or colors indicated by a key. These techniques provide an easy way to compare the results of experiments.
- The graph should clarify whether the data start at the origin (0,0) or not. The line should not be extended to the origin if the data do not start there. In addition, the line should not be extended beyond the last data point (extrapolation) unless a dashed line (or some other demarcation) clearly indicates that this is a prediction about what may happen.
- For most of the labs in the lab manual, students should include standard error in their analysis and use standard error bars on their graphical displays when appropriate (see the discussion of bar graphs earlier in this chapter).
- For detailed information about line graphs, see Appendix B in *AP Biology Investigative Labs: An Inquiry-Based Approach*.

USING GRAPHING PROGRAMS

Graphical representations of data can be created using Excel and other spreadsheet programs, but scientific plotting or data display generally is clumsy with that software. To make graphing and analysis more accessible in Excel, students should consider downloading the Merlin Excel Statistical Package, available at http://www.heckmondwikegrammar.net/index.php?highlight=introduction&p=10310.

Other graphing programs and statistical packages usually provide better graphing tools, along with better and more efficient tools for the statistical tests and data analysis. Students who wish to explore data display, data analysis, and hypothesis testing more deeply will benefit from learning to use a dedicated software package. Many of the commercial packages are prohibitively expensive, but a free and very powerful software package that will do the job—R—is available at http://www.r-project.org/. It is the software of choice for many researchers, but it is not very accessible for the novice user. For the highly motivated, though, it is a useful option.

ALIGNMENT TO THE INVESTIGATIONS

Sample Alignment to Investigation 5: Photosynthesis

What factors affect the rate of photosynthesis in living leaves?

In this investigation, students learn how to measure the rate of photosynthesis indirectly by using the floating leaf disk procedure to gauge oxygen production. As students consider variables that might affect the rate of photosynthesis and the floating disk procedure itself, a number of questions emerge that lead to independent student investigations requiring the application of graphing skills.

Alignment to the Curriculum Framework

To have a rich foundation in biology, students need to apply quantitative methods to the study of biology, especially for the laboratory experience. Graphing is one of these methods. Investigation 5: Photosynthesis provides opportunities for students to design experimental plans, collect data, graph data, and apply mathematical routines to draw conclusions from their graphs.

In this investigation, students will monitor the number of leaf disks that float when oxygen is produced during photosynthesis. They have a number of options to design their own procedure for collecting and graphing the data. They may choose to count the number of disks floating at the end of time intervals or time how long each disk takes to float. Counting is one of the most common activities of biological observation. Each procedure generates a different kind of data: counting produces nonparametric data, while measuring the time for each disk to rise generates parametric data. Both kinds of data have their strengths and weaknesses. After students collect data, they plot their results on a graph(s) and draw conclusions about the effect(s) of environmental variables on the rate of photosynthesis.

Science Practices Applied

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

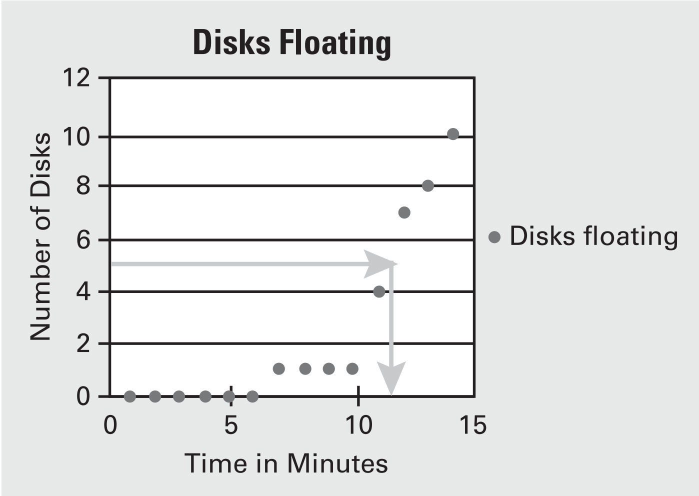
SP 5 The student can perform data analysis and evaluation of evidence.

Learning Objective Addressed

LO 4.14 The student is able to apply mathematical routines to quantities that describe interactions among living systems and their environment, which result in the movement of matter and energy (4.A.6 & SP 2.2).

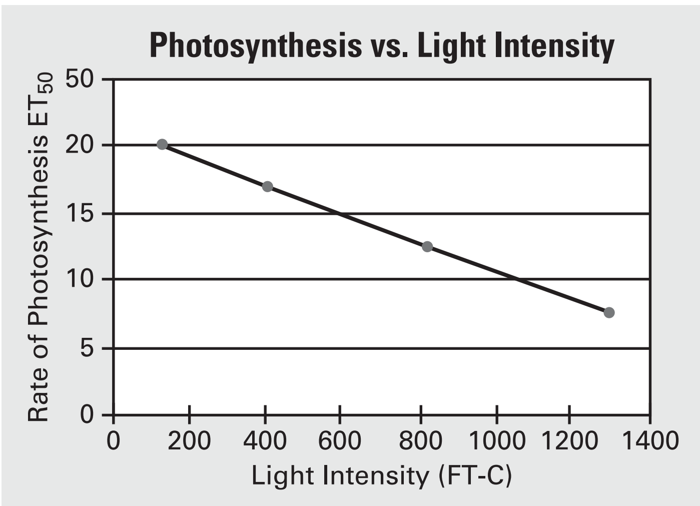
Example of Sample Data and Constructed Graphs

Analysis and presentation of data are difficult for most students, and graphing data allows students to visualize patterns from which they can draw conclusions. The more data students graph, the sooner they begin to understand that certain plot shapes or forms are associated with models that make it easier to infer causal mechanisms. The following is an example of a graph of results that a student might construct for Investigation 5: Photosynthesis:



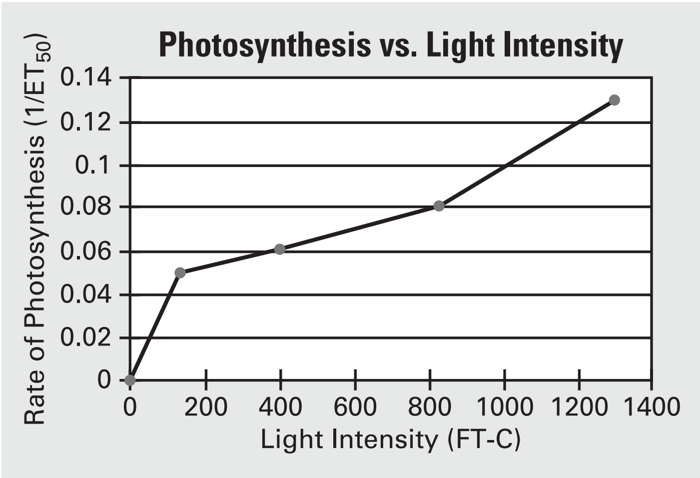
The data in the graph were collected by the following method, although others work as well. The disks floating during exposure to a specific environmental variable were counted at the end of each time interval. The median was chosen over the mean as the summary statistic; the median generally provides a better estimate of the central tendency of the data because, on occasion, a disk fails to rise or takes a long time to do so. Consequently, for this sample, the median time for five disks to rise was somewhere between 11 and 12 minutes.

A term coined by G.L. Steucek and R.J. Hill [1985] for this relationship is ET50, the estimated time for 50 percent of the disks to rise. That is, rate is a change in a variable over time. The time required for 50 percent of the leaf disks to float is represented as Effective Time = ET50. The following is a sample graph of a photosynthesis light response curve utilizing the ET50 concept. In this example, the student was investigating the effect of light intensity on the rate of photosynthesis.



Source: Steucek and Hill, 1985

In this example, note that the shape of this curve is not the expected curve that rises and levels off. This is because the times to float are the inverse of the rate of photosynthesis. Taking the reciprocal of ET50 (1/ET50) allows the graphic presentation to more closely express the physical phenomenon, as shown in the following graph.



Source: Steucek and Hill, 1985

ALIGNMENT TO THE EXAM

Graphing and Exam Questions

One of the best ways to communicate scientific phenomena, solve scientific problems, and reflect the results of a scientific investigation is graphing. Graphing allows students to create visual representations of data that have been counted, measured, and calculated; graphical displays of data also reflect the question(s) that the investigation purports to answer. In addition, certain plotted shapes or forms are associated with models that make it easier for students to infer causal mechanisms. Students need to be able to determine what type of graph (e.g., histogram, line graph) most appropriately reflects their collected data and then create the graph and use it to draw conclusions, make predictions, and pose questions for further investigation. The following sample exam questions from the *AP Biology Practice Exam* represent items that assess students' ability to glean information about biological phenomena based on graphical analysis and the application of quantitative skills. The questions are based on the learning objectives in the curriculum framework.

Sample Multiple-Choice Question Using Graphing

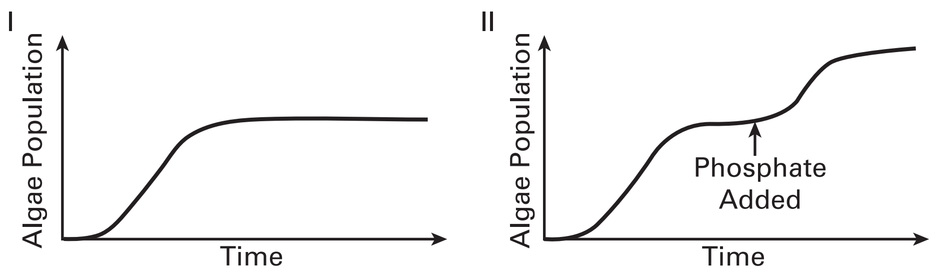
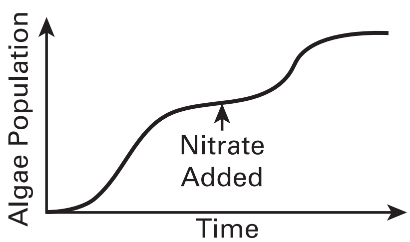


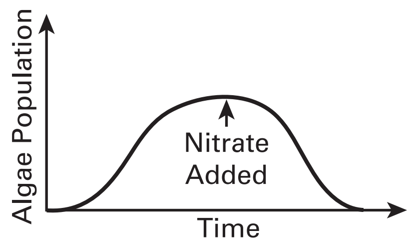
Figure I shows the growth of an algal species in a flask of sterilized pond water. If phosphate is added as indicated, the growth curve changes as shown in Figure II.

Which of the following is the best prediction of the algal growth if nitrate is added instead of phosphate?

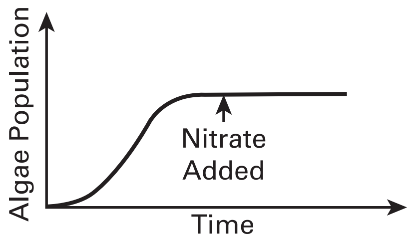
(A)



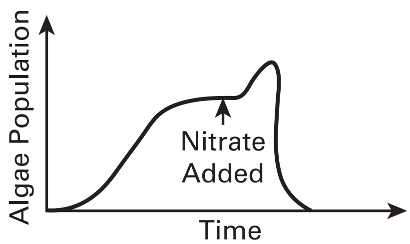
(B)



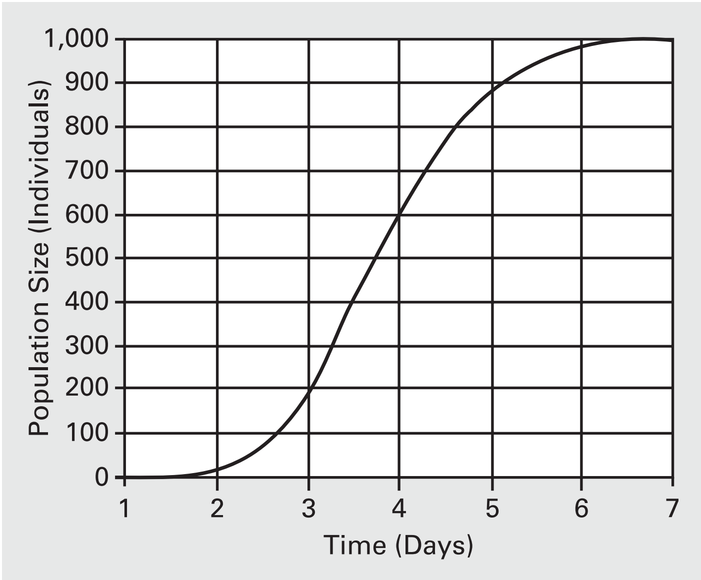
(C)



(D)



Sample Grid-In Question Using Graphing



Use the graph above to calculate the mean rate of population growth (individuals per day) between day 3 and day 5. Give your answer to the nearest whole number.

Sample Free-Response Question Using Graphing

Plants lose water from their aboveground surfaces in the process of transpiration. Most of this water is lost from stomata, microscopic openings in the leaves. Excess water loss can have a negative effect on the growth, development, and reproduction of a plant. Severe water loss can be fatal. Environmental factors have a major impact on the rate of plant transpiration.

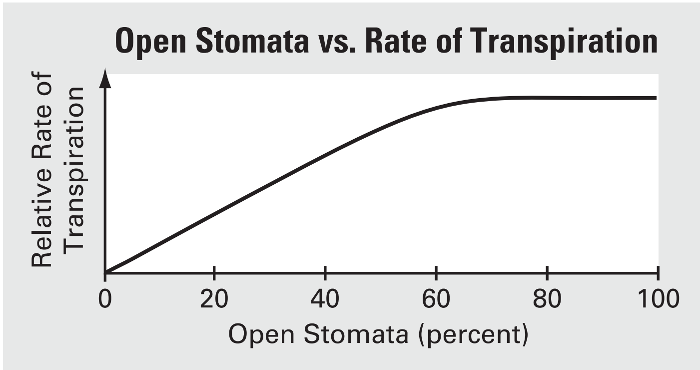
Transpiration Rate Versus Temperature

Temperature (°C)	20	23	27	28
Transpiration Rate ($mmol/m^2.sec$)	1.5	3	5	4.5

(a) Using the data above and the axes provided, **draw** a graph showing the effect of temperature change on the rate of transpiration. **Explain** the shape of the curve from 23 degrees to 28 degrees.

(b) Humidity is another environmental factor that affects transpiration rate. Using the axes provided, **draw** a curve that illustrates what you predict would be the rate of transpiration with increasing humidity and constant temperature. **Justify** the shape of the curve based on your prediction.

(c) The curve below illustrates the rate of transpiration related to the percent of open stomata on the leaf of a particular plant. **Explain** why the curve levels off with increasing percentage of open stomata per area of the leaf.



(d) The data below show the density of stomata on the leaf surfaces of three different species of plants. **Describe** the environments in which each plant most likely evolved. **Justify** your descriptions.

	Stomata Density (# of stomata/mm^2)	
Plant	In Upper Epidermis	In Lower Epidermis
Anacharis	0	0
Water Lily	420	0
Black Walnut	2	465

RESOURCES

The following resources include ideas for classroom activities and lessons, as well as examples for learning and teaching as you further your exploration into the new AP Biology curriculum. Each resource includes the approximate time involved, a description of the resource and where to find it, the quantitative skills the activity or lesson involves, and where those skills might be applied.

MR. GIRARD'S BIOLOGICAL SCIENCES WEBSITE http://phsgirard.org/AcademicBiology.html	APPROXIMATE TIME 2 class periods

DESCRIPTION OF RESOURCE

This website provides a section on graphing with the use of a ready-made PowerPoint presentation. The examples can be used to help students understand the various types of graphs that exist, compare graphs, and make use of graphs for analyzing data. The website has links to downloads that can be used as exercises, such as using populations for human growth to help students create a graph and identify factors that affect population growth based on data provided.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1–2.3) The student can use mathematics appropriately.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 1.26 The student is able to evaluate given data sets that illustrate evolution as an ongoing process. [See SP 5.3]

LO 4.11 The student is able to justify the selection of the kind of data needed to answer scientific questions about the interaction of populations within communities. [See SP 1.4, 4.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

THE BIOLOGY CORNER: ECOLOGY http://www.biologycorner.com/flash/mark_recap.swf	APPROXIMATE TIME 2 class periods

DESCRIPTION OF RESOURCE

You can use this activity to show students how to estimate population size. It teaches students how to estimate population size by sampling and using the mark and recapture method to be able to apply the techniques to solve population problems and compare methods of population estimation. Students use a mathematical formula after they simulate the two techniques and obtain the data on a worksheet. They then calculate population size and compare the data. Students are then given a scenario and asked to estimate population size.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1–2.3) The student can use mathematics appropriately.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 1.26 The student is able to evaluate given data sets that illustrate evolution as an ongoing process. [See SP 5.3]

LO 4.11 The student is able to justify the selection of the kind of data needed to answer scientific questions about the interaction of populations within communities. [See SP 1.4, 4.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

THE BIOLOGY CORNER: DEER PREDATION OR STARVATION http://www.biologycorner.com/worksheets/predator_prey_graphing.html	APPROXIMATE TIME 1 class period

DESCRIPTION OF RESOURCE

This lesson activity contains the results of a lab for predation. You can use the data from this activity to show students how to graph population sizes and analyze data with the questions provided in a worksheet. Students obtain graphing and analysis skills in a brief one-class lesson.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1–2.3) The student can use mathematics appropriately.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 1.26 The student is able to evaluate given data sets that illustrate evolution as an ongoing process. [See SP 5.3]

LO 4.11 The student is able to justify the selection of the kind of data needed to answer scientific questions about the interaction of populations within communities. [See SP 1.4, 4.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

AP BIOLOGY DAILY LESSON PLANS http://www.carolina.com/text/pdf/other/AP_Biology_Daily_Lesson_Plans.pdf	APPROXIMATE TIME 11 class periods

DESCRIPTION OF RESOURCE

This resource contains 11 lesson plans for cell biology, genetics, evolution, anatomy and physiology, botany, and ecology. Students use group interaction and teamwork to complete the activities and questions for each lesson.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1–2.3) The student can use mathematics appropriately.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 4.11 The student is able to justify the selection of the kind of data needed to answer scientific questions about the interaction of populations within communities. [See SP 1.4, 4.1]

HANDBOOK OF BIOLOGICAL STATISTICS http://udel.edu/~mcdonald/statintro.html	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

This online book has a graphing section that provides general guidance in creating graphs. The guide helps students understand the importance of analyzing their data and presenting their results in coherent and effective graphs. You can use this resource at the beginning of and throughout the school year to help students with their graphing skills. Since students can represent their data in many types of graphs, the guide includes discussions on choosing the right type of graph for the data and understanding the ways to represent results. It explains the use of scatter graphs and bar graphs. It discusses types of measurable variables (i.e., independent and dependent variables) and the addition of error bars to the graphs. Throughout the guide are links to further resources, which include explanations for understanding standard errors and nominal variables. The guide also outlines techniques for using Excel and Calc (part of OpenOffice.org) in generating graphs.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1–2.3) The student can use mathematics appropriately.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 1.26 The student is able to evaluate given data sets that illustrate evolution as an ongoing process. [See SP 5.3]

LO 1.31 The student is able to evaluate the accuracy and legitimacy of data to answer scientific questions about the origin of life on Earth. [See SP 4.4]

LO 2.9 The student is able to represent graphically or model quantitatively the exchange of molecules between an organism and its environment, and the subsequent use of these molecules to build new molecules that facilitate dynamic homeostasis, growth and reproduction. [See SP 1.1, 1.4]

LO 2.17 The student is able to evaluate data that show the effect(s) of changes in concentrations of key molecules on negative feedback mechanisms. [See SP 5.3]

LO 2.30 The student can create representations or models to describe nonspecific immune defenses in plants and animals. [See SP 1.1, 1.2]

(continued on next page)

(continued from previous page)

LO 2.32 The student is able to use a graph or diagram to analyze situations or solve problems (quantitatively or qualitatively) that involve timing and coordination of events necessary for normal development in an organism. [See SP 1.4]

LO 4.11 The student is able to justify the selection of the kind of data needed to answer scientific questions about the interaction of populations within communities. [See SP 1.4, 4.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

LABWRITE RESOURCES http://www.ncsu.edu/labwrite/res/res-homepage.htm	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

You can use these student homework tutorial modules throughout the school year to aid students who are struggling with the basic skills of graphing. *LabWrite Resources* contains a section on graphing resources that includes a step-by-step Excel tutorial and a flowchart on how to use and present certain types of data. This information helps the student decide what type of graph would best represent the type of data collected in a particular lab. The resource includes links with descriptions of the different types of bar graphs, histograms, line graphs, and scatterplots followed by examples illustrating when to use each type of graph. Tips on how to prepare and modify a graph for final presentation, such as the inclusion of error bars for summarizing data, are also provided. This resource is particularly helpful because it contains many useful examples to go with the description and explanation of graphing techniques.

SCIENCES PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1–2.3) The student can use mathematics appropriately.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 2.9 The student is able to represent graphically or model quantitatively the exchange of molecules between an organism and its environment, and the subsequent use of these molecules to build new molecules that facilitate dynamic homeostasis, growth and reproduction. [See SP 1.1, 1.4]

LO 2.32 The student is able to use a graph or diagram to analyze situations or solve problems (quantitatively or qualitatively) that involve timing and coordination of events necessary for normal development in an organism. [See SP 1.4]

LO 4.11 The student is able to justify the selection of the kind of data needed to answer scientific questions about the interaction of populations within communities. [See SP 1.4, 4.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

EASTERN ILLINOIS UNIVERSITY BIOLOGY 1100: GENERAL BIOLOGY RESOURCES http://castle.eiu.edu/bio_data/resources/bio1100/index.php	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

You can use this resource when classic AP Biology labs are done in the classroom or when demonstrating techniques in data collection prior to labs. This resource from Eastern Illinois University for Biology 1100: General Biology provides 2003 and 2007 Microsoft Excel graphing tutorials for most of the classic AP Biology labs. The tutorials walk students through simple graphs as well as graphs involving multiple data sets. Each link provides images with clear explanations for novice Excel users.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 2.9 The student is able to represent graphically or model quantitatively the exchange of molecules between an organism and its environment, and the subsequent use of these molecules to build new molecules that facilitate dynamic homeostasis, growth and reproduction. [See SP 1.1, 1.4]

LO 2.32 The student is able to use a graph or diagram to analyze situations or solve problems (quantitatively or qualitatively) that involve timing and coordination of events necessary for normal development in an organism. [See SP 1.4]

INTERPRETING ECOLOGICAL DATA http://www.biologycorner.com/worksheets/interpreting_data.html	APPROXIMATE TIME 1 class period or homework

DESCRIPTION OF RESOURCE

You can use these short and easy ready-made graphs to help students interpret ecological data for a class activity in groups or as homework. Six different types of charts and data tables have short questions about the data or interpretation of the graphs. This is a great way to help students start looking at data and inquiring about the trends they see or are prompted to see in the data and results.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 2.9 The student is able to represent graphically or model quantitatively the exchange of molecules between an organism and its environment, and the subsequent use of these molecules to build new molecules that facilitate dynamic homeostasis, growth and reproduction. [See SP 1.1, 1.4]

LO 2.32 The student is able to use a graph or diagram to analyze situations or solve problems (quantitatively or qualitatively) that involve timing and coordination of events necessary for normal development in an organism. [See SP 1.4]

HUMAN POPULATION GROWTH http://www.biologycorner.com/worksheets/humanpop_graph.html	APPROXIMATE TIME 2 class periods

DESCRIPTION OF RESOURCE

The objective of this activity is for students to create a graph of human population growth by using the data provided and to use the graph to predict future human growth. Students will then identify factors that affect population growth.

SCIENCE PRACTICES APPLIED

SP 1 (1.1–1.5) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 4 (4.4) The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 (5.1–5.3) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 2.9 The student is able to represent graphically or model quantitatively the exchange of molecules between an organism and its environment, and the subsequent use of these molecules to build new molecules that facilitate dynamic homeostasis, growth and reproduction. [See SP 1.1, 1.4]

LO 2.32 The student is able to use a graph or diagram to analyze situations or solve problems (quantitatively or qualitatively) that involve timing and coordination of events necessary for normal development in an organism. [See SP 1.4]

CHAPTER 2:

Data Analysis

LABS USING DATA ANALYSIS

1: Artificial Selection

2: Mathematical Modeling: Hardy-Weinberg

3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

7: Cell Division: Mitosis and Meiosis

As your students start making observations and collecting data for an investigation, they will probably notice patterns. These patterns may or may not be real, or valid. Quantitative data analysis is one of the first steps toward determining whether an observed pattern has validity. Data analysis also helps distinguish among multiple working hypotheses. Every laboratory in the manual that calls for any data collection—even a mathematical modeling lab such as Mathematical Modeling: Hardy-Weinberg—requires some level of **data analysis**, or quantitative description and exploration of the masses of data to discover meaningful patterns relevant to your investigation. This analysis includes descriptive statistics, tabular and graphical data display (see Chapter 1), and inferential statistics. **Descriptive statistics** serves to summarize the data. It helps show the variation in the data, standard errors, best-fit functions, and confidence that sufficient data have been collected. **Inferential statistics** involves making conclusions beyond the data analyzed—using your experimental sample data to infer parameters in the natural population. These skills will be described in this chapter and throughout the rest of the publication.

Most of the data your students collect will fit into two categories: measurements or counts. **Measurements** are recordings of quantitative data, such as absorbance, size, time, height, and weight. Most measurements are continuous, meaning there is an infinite number of potential measurements over a given range. **Count data** are recordings of qualitative, or categorical, data, such as the number of hairs, number of organisms in one habitat versus another, and the number of a particular phenotype. Measurements and counts require different statistical tools. Count data usually do not lend themselves to the kind of analysis that this chapter will introduce. Chapter 3, Hypothesis Testing, includes more about using count data.

Many of the labs in the new lab manual introduce a technique or procedure and then ask students to design and carry out an investigation using that technique or procedure in conjunction with a question of their choice. A run-through of the technique or procedure is one of the first steps students can take before designing a more rigorous investigation based on their own questions. The examples at the end of this chapter demonstrate how such a run-through can help provide preliminary data for students to explore and analyze in order to discover patterns and make observations that can inform their hypotheses and experimental design without spending a lot of time, effort, and supplies on a wild goose chase.

COMPONENTS OF DATA ANALYSIS

Measurements in biology should reflect the variations in a given population. Obviously, a researcher cannot collect data on an entire population but rather must choose smaller samples to inform the investigation. These samples can then be analyzed using descriptive and inferential statistics.

Descriptive Statistics

Descriptive statistics is used to estimate important parameters of the sample data set. Examples include sample standard deviation, which describes the variability in the data; measurements of central tendencies such as mean, median, and mode; and sample standard error of the sample mean, which helps you determine your confidence in the sample mean. The same parameters (mean, standard deviation, etc.) can also describe the entire or true population that you are studying, but collecting the data to compute these statistics is most often not possible. That's where inferential statistics comes in.

Inferential Statistics

Inferential statistics includes tools and methods (statistical tests) that rely on probability theory and an understanding of distributions to determine precise estimates of the true population parameters from the sample data. This is a key part of data analysis and allows you to support and draw conclusions from your data about the true population.

WHY BOTHER WITH DATA ANALYSIS?

Why not provide the results of an experiment and let the reader do the work? Authentic research experiences involve more than the conception, design, and carrying out of an investigation. They involve arriving at conclusions about the work and communicating those conclusions to the larger community. That communication involves making claims and/or arguments from evidence. The arguments or claims are only as strong as the logic and the evidence. In order to draw sound conclusions on a question, it is important to have an estimate on the reliability of the data for the analysis along with a description of the types of tests and results used to support or reject hypotheses. What are the bases of your decisions?

Novice researchers are often at a loss to explain their sampling procedure. They usually do not know if they have an adequate sample to support their claims. Often they will take a sample of three or fewer since that is what they have done in the past. Next they will average these three data points and make an argument that is not supported with data. On the other hand, with appropriate techniques such as standard error, students can generate measures of confidence that lead to greater precision. Have they collected enough data? Can they make their claims with confidence? Is there a difference or are the results they observe due to chance? As you work through this chapter, note how the descriptive and inferential statistics help to inform the entire process of the investigation.

USING DATA ANALYSIS

When an investigation involves measurement data, one of the first steps is to construct a histogram, or frequency diagram, to represent the data's distribution (see Chapter 1). If enough measurements are made, the data can show an approximate normal distribution, or bell-shaped distribution, on a histogram; if they do, they are parametric data. The normal distribution is very common in biology and is a basis for the predictive power of statistical analysis. If the data do not approximate a normal distribution (that is, they are nonparametric data), then other descriptive statistics and tests need to be applied to those data. Keep in mind, though, that even if a sample distribution may not appear to be a perfect bell curve, the actual population could still be more or less normally distributed. The shape of the sampling distribution could be due to sampling bias or errors. Students may have a good reason to assume that they have measured a normal distribution. If so, they may want to consider a larger **sample size** before continuing. An important part of designing an experiment is gathering enough information to develop an understanding of the nature of the variable being studied.

For a normal distribution, the appropriate descriptive statistics for the data set include the mean (average), sample size, standard deviation, and standard error. Each is important. The mean of the sample is the average (the sum of the numbers in the sample divided by the total number in the sample). The mean summarizes the entire sample and might provide an estimate of the entire population's true mean. The sample size refers to how many members of the population are included in the study. Sample size is important when students try to estimate how confident they can be that the sample set they are trying to analyze represents the entire population.

Both the standard deviation measure and the standard error measure define boundaries of probabilities. The **sample standard deviation** is a tool for measuring the spread (variance) in the sample population, which in turn provides an estimate of the variation in the entire sample set. A large sample standard deviation indicates that the data have a lot of variability. A small sample standard deviation indicates that the data are clustered close to the sample mean.

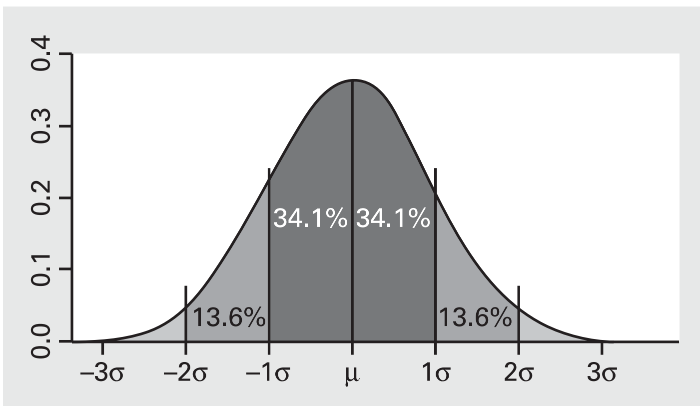


Figure 6. Standard Deviations and a Normal Distribution

A little more than two-thirds of the data points will fall between +1 standard deviation and −1 standard deviation from the sample mean. More than 95% of the data falls between ±2 standard deviations from the sample mean.

Sample standard error (SE) is a statistic that allows students to make an inference about how well the sample mean matches up to the true population mean. If one were to take a large number of samples (at least 30) from a population, the means for each sample would form an approximately normal distribution—a distribution of sample means. Normally, you would not do hundreds of individual investigations on a population. This distribution of sample means, then, is a theoretical construct that helps us define our boundaries of confidence in our sample. This distribution also has parameters, such as a standard deviation. Standard error is the equivalent of the standard deviation of the sampling distribution of the means and is calculated from the following formula:

$$\frac{s}{\sqrt{n}}$$

where s = the sample standard deviation and n = the sample size.

A sample mean of ±1 SE describes the range of values about which an investigator can have approximately 67% confidence that the range includes the true population mean. Even better, a sample with a ±2 SE defines a range of values with approximately a 95% certainty. In other words, if the sampling were repeated 20 times with the same sample size each time, the **confidence limits,** defined by ±2 SE, would include the true population mean approximately 19 times on average. This is the **inference;** it is a statistic that allows investigators to gauge just how good their estimate of the true population mean actually is. With this understanding, the investigator can establish ahead of time a reasonable sample size for this population and the degree of confidence needed.

For most of the labs in the lab manual, students should be sure to include standard error in their analysis and use standard error bars on their graphical displays when appropriate (see Chapter 1).

The path through data analysis for the typical AP Biology lab will mirror the steps just described if the investigation involves normally distributed and continuous sample data (parametric data). However, some measurement data will not be normally distributed. The data distribution may be skewed or have large or small outliers (nonparametric data). In such cases, the descriptive statistic tools are a bit different. Generally, the parameters calculated for nonparametric statistics include medians, modes, and quartiles, and the graphs are often box-and-whisker plots (see Chapter 1). The process is the same, however. Students will use these tools to help create a picture of their data and understand the limitations of what their data can tell them. Chapter 3, Hypothesis Testing, includes an explanation of a decision table (Table 1 in Chapter 1) that should help guide students in choosing tools to use based on the data they collect.

Once students have developed a hypothesis, designed an experiment, collected data, and applied a number of descriptive statistics that summarize the data visually, they can apply the standard error statistic as an inference to describe the confidence they have that the means of the sample represent the true population means. The next step is hypothesis testing, which is discussed in Chapter 3.

EXAMPLES OF DATA ANALYSIS

Example 1: Comparing English Ivy Leaves in Sunny and Shady Environments

English ivy has been planted throughout the United States around buildings as a ground cover. Recently, a student noticed that the ivy leaves growing on the shady side of a building were larger than ivy leaves growing on the sunny side of the same building. The student formulated the question, *Do shady English ivy leaves have a larger surface area than sunny English ivy leaves?* It would not be practical to collect all of the leaves growing. Instead, as in most biological investigations, it is advisable to choose smaller samples to inform an investigation. These samples should be as random and as unbiased as possible. The student collected and measured the maximum width, in centimeters, of 30 leaves from each habitat. To try to reduce bias in her sample, she had a friend collect the leaves without telling the friend what the purpose was. (There is still a problem with this technique. Can you think of a better technique for sampling?) Table 2 shows her results.

Table 2. A Student's Leaf Measurement Data

Shady Leaves (in cm)	Sunny Leaves (in cm)
3.7	3.2
5.2	3.5
5.4	4.1
5.7	4.3
5.8	4.4
5.8	4.6
6.0	5.0
6.1	5.0
6.5	5.2
6.5	5.2
6.6	5.3
6.8	5.4
7.0	5.6
7.3	5.7
7.3	5.7
7.4	5.8
7.7	6.0
7.9	6.0
8.0	6.4
8.1	6.5
8.1	6.7
8.2	6.7
8.3	7.1
8.9	7.1
9.0	7.1
9.4	7.3
9.9	7.5
9.9	7.9
9.9	8.0
10.4	8.2

Staring at these numbers probably won't help the student answer her question about the leaf size in different habitats. This is where descriptive statistics comes into play.

The student should notice that the measurements are made in centimeters and are an example of continuous measurement data. These data are not counts or categories; they vary continuously. Since they are measurement data, one of the first steps would be for the student to construct a histogram and see if it approximates a normal distribution (Figure 7).

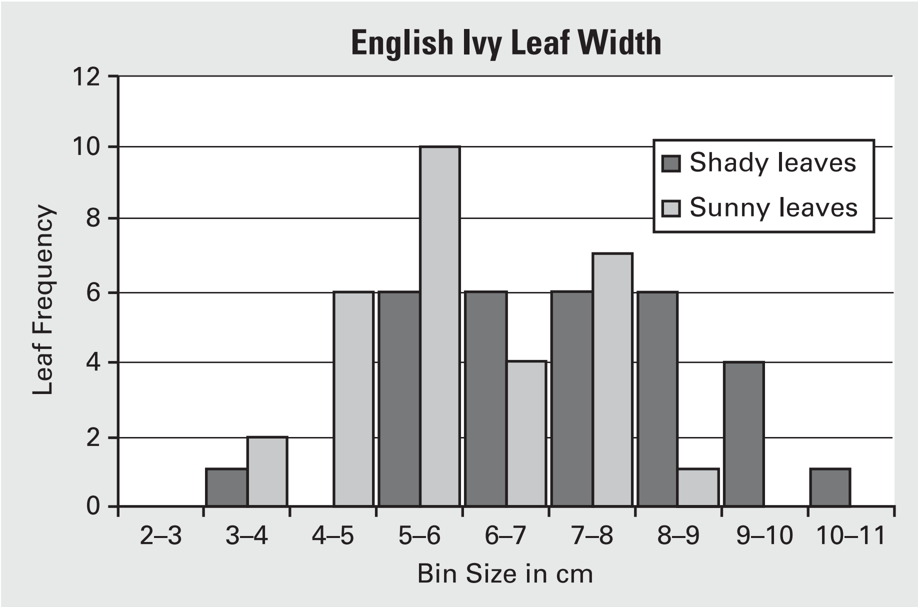


Figure 7. Overlapping Distributions for Both Shady and Sunny Leaves Plotted on the Same Graph

The student now should remember that just because neither distribution reflects a perfect bell curve does not mean that the actual population is not normally distributed. The shape of this distribution may be due to sampling bias or errors. In fact, there is good reason to think that even though the student tried to select leaves randomly, she did not in fact use a sampling protocol that would maximize a random sample. One possibility would be for her to grid off the population of ivy, use a random number table to select sampling areas, and then sample all of the leaves in that one area or point. But that would be for later. Right now, the student has these data to work with. Experience suggests that these distributions will approximate a normal distribution and that they both have about the same spread. The student should bear in mind, however, that this assessment might have to change if a better sampling protocol suggested a skewed, or abnormal, distribution.

Now the student should recall that the standard error (SE) is calculated from the following formula:

$$\frac{s}{\sqrt{n}}$$

where s = the sample standard deviation and n = the sample size.

As discussed earlier in the chapter, the sample mean ±1 SE describes the range of values about which the investigator can have approximately a 67% confidence that the range includes the true population mean. Even better, a sample with a ±2 SE defines a range that has approximately a 95% confidence. In other words, if the sampling were repeated 20 times with the same sample size each time, the confidence limits would include the true population mean approximately 19 times, on average.

The student enters her data in a spreadsheet and then creates formulas to calculate the descriptive statistics, which are shown in Table 3.

Table 3. Descriptive Statistics

	Shady Leaves	Sunny Leaves
Mean	7.43	5.88
Standard Deviation	1.63	1.32
N	30	30
Standard Error	0.30	0.24

She then produces the bar chart shown in Figure 8 to compare the means. Notice that she includes error bars of ±1 SE.

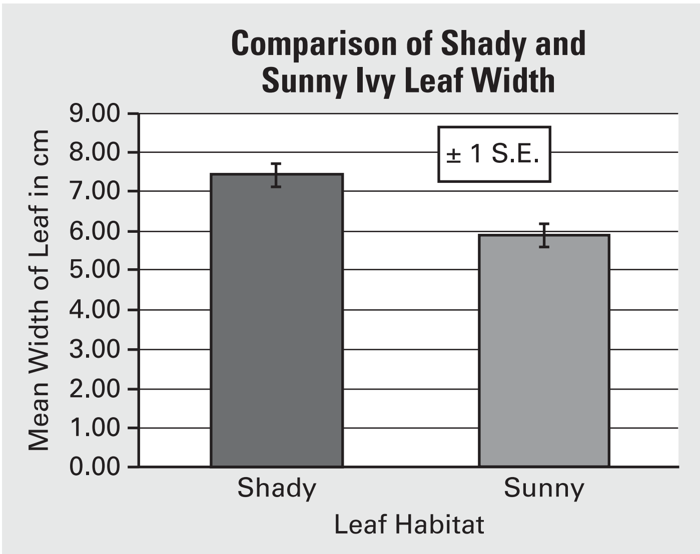


Figure 8. Bar Chart with Standard Error Bars Comparing the Means for Shady and Sunny Ivy Leaf Width

The student now has a graph and a number of descriptive statistics that summarize her data. In addition, she can apply the standard error statistic as an inference to describe the confidence she has that the means of her sample represent the true population means. It is evident from the graph that the error bars for the shady leaf mean do not overlap with the error bars for the sunny leaf mean. In fact, had the student chosen to plot ±2 SE error bars, they, too, would not have overlapped. This non-overlap strongly suggests that a more rigorous statistical test will define a high probability that the two populations are indeed different from each other. This analysis will be saved for Chapter 3: Hypothesis Testing. This student now has a wealth of information about leaf size and habitat to develop a rigorous experimental design. Next, she needs to develop her hypothesis, design her experiment, collect data, and run her tests and analysis.

Example 2: Body Temperature

One of the most familiar numbers in biology is the typical body temperature for healthy humans—37 °C, or 98.6 °F—a figure that has been accepted for more than 100 years. Anyone who has ever tried to measure his or her own internal body temperature, however, is aware of issues that can bring into question the accuracy of the measured body temperature.

Many variables can affect a body temperature measurement. For example, temperature measurements taken from different parts of the body can generate a range of readings for a single person, as can the readings from different instruments, such as a digital thermometer, an infrared thermometer, or a glass-and-mercury thermometer. And this amount of variability doesn't even take into account that an individual's body temperature actually varies on both a daily and a monthly basis due to biological rhythms.

A key question that arises for investigation, then, might be, *Is 98.6°F actually the average body temperature for humans?* Tell your students to imagine that they have designed a study to answer this question. They have randomly selected 130 healthy 18- to 40-year-old adults—65 males and 65 females. They then measured their body temperature, orally, at the same time of day to a 10th of a degree precision. Table 4 shows the data for body temperature collected, in degrees Fahrenheit.

Table 4. Body Temperature Data, in Degrees Fahrenheit

96.3	97.4	97.8	98.0	98.4	98.6	99.1	97.4	98.0	98.2	98.5	98.7	99.0
96.7	97.4	97.8	98.1	98.4	98.7	99.2	97.6	98.0	98.2	98.6	98.8	99.1
96.9	97.4	97.8	98.1	98.4	98.7	99.3	97.7	98.0	98.3	98.6	98.8	99.1
97.0	97.5	97.9	98.2	98.5	98.8	99.4	97.7	98.0	98.3	98.6	98.8	99.2
97.1	97.5	97.9	98.2	98.5	98.8	99.5	97.8	98.0	98.3	98.6	98.8	99.2
97.1	97.6	98.0	98.2	98.6	98.8	96.4	97.8	98.1	98.4	98.7	98.8	99.3
97.1	97.6	98.0	98.2	98.6	98.9	96.7	97.8	98.2	98.4	98.7	98.8	99.4
97.2	97.6	98.0	98.3	98.6	99.0	96.8	97.9	98.2	98.4	98.7	98.8	99.9
97.3	97.7	98.0	98.3	98.6	99.0	97.2	97.9	98.2	98.4	98.7	98.9	100.0
97.4	97.8	98.0	98.4	98.6	99.0	97.2	97.9	98.2	98.4	98.7	99.0	100.8

Your students now must determine how well this sample represents the entire population of humans. The data are actually from a sample data set prepared by Allen Shoemaker (Shoemaker, 1996). This particular data set has been modified from the results of a study published in the *Journal of American Medical Association* (Mackowiak, Wasserman, and Levine, 1992). The original data were reported in degrees Celsius, but the modified data set has been converted to degrees Fahrenheit. Real data tend to be messy, so this data set has been cleaned up a bit to make the analysis and tool application more accessible for novice learners.

Have your students upload the data set from http://www.amstat.org/publications/jse/datasets/normtemp.dat.txt into a spreadsheet such as Excel, a statistics software package such as SPSS, or a Web-based statistics package such as those available at the Statistics

Online Computational Resource (SOCR) website: http://www.socr.ucla.edu/SOCR. Once students have loaded the data, have them work through each of the following sections in their own spreadsheets to prepare them for applying these tools to their own data in the future. The instructions and illustrations in the rest of this example are based on the Excel spreadsheet.

Exploring the Sample Data

Before the students do any calculations, ask them to consider how these sample data are distributed. To do so, they will need to create a histogram, or frequency diagram, of all the sample data points (Figure 9). (They can use the Excel help function to explore how to do this task or, better yet, use Google to search for "How to make a histogram in Excel 2010" to find videos and other forms of instruction.) In Figure 9, note that the sample distribution resembles an approximately normal distribution, or bell curve. One of the first steps in deciding on the procedure and techniques that will be applied during data analysis is to determine the nature of the sample data. A normal distribution suggests a number of standard analysis tools.

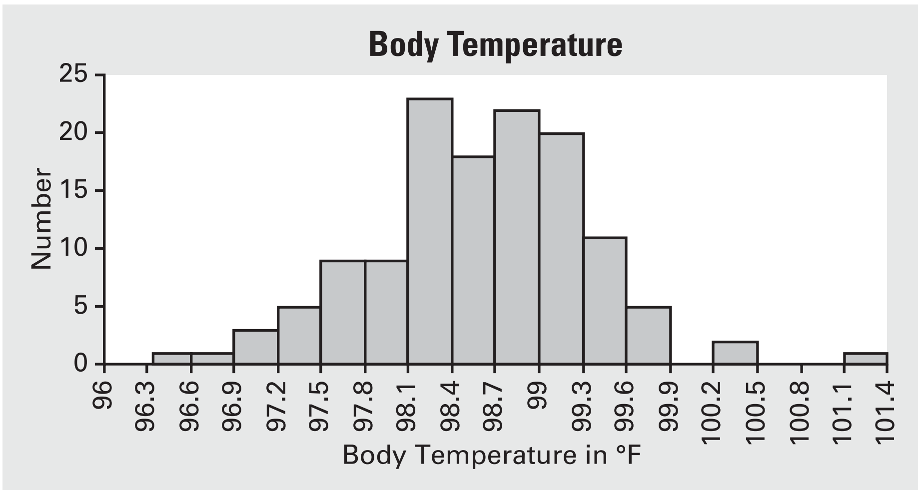


Figure 9. Histogram of Human Body Temperature from Sample Data

Start with Descriptive Statistics

Point out that there are 130 data points. Later, your students will need to decide whether this data set is an adequate sample to answer the following question with confidence: *How do these data compare with the accepted 37°C or 98.6°F values for body temperature?* First, though, the students must calculate a sample mean for these data. It's the sample mean that they will want to compare with the accepted 98.6°F body temperature for a normal healthy adult. They can calculate the sample mean from the data. (Use the = average function if using Excel):

Sample mean = 98.25°F[1]

[1] Note that by convention, descriptive statistics rounds the calculated results to the same number of decimal places as the number of data points plus 1. For example, the temperatures were measured to a 10th of a degree precision. Therefore, the sample mean and other similar statistics are rounded to the nearest 100th of a degree.

Next, the following question should be in their minds: *So, does 98.25°F represent an abnormal mean temperature (representing a true difference between the accepted mean and the one we calculated), or is the difference an acceptable average that just happened by chance, since we tested only 130 individuals and not the entire human race?*

Next, how might variability in the data or the width of the distribution be described? Explain that one measure of the spread of the data is the sample standard deviation. Use the sample standard deviation formula in Excel (=STDEV) to represent the variation in the sample data:

Sample standard deviation = 0.73°F

The nature of the sample's distribution can now be described by combining the sample mean with the sample standard deviation:

98.25 ± 0.73°F

Since the data were approximately normal, around 68% of the temperatures from above should be between 97.51 and 98.99. The standard deviation of the sample will also be used later to estimate the standard error of the sample.

For this study, the students randomly chose 130 people to have their body temperature measured. Discuss what would happen if the students took another random sample of 130 from the true population. Would that sample mean equal the first sample mean? Not likely; it might be closer or farther away from the true mean of the entire population of humans. What if students repeated samplings like this 10 more times, 100 more times, or even 1,000 more times? Again, compared with the actual true mean (unknowable in this case, but rather theoretical), the "sample" of all possible sample means ends up having an approximately normal distribution centered on the true mean. Because the normal distribution is also a probability distribution, the students can determine precise estimates of how confident they are about how close their sample mean is to the true mean. They should recall that the measure of the variation in sample means is known as the *standard error.*

Sample standard error is estimated with this formula:

$$\frac{\text{Sample standard deviation}}{\sqrt{n}\,(\text{sample size})}$$

or where s = sample standard deviation:

$$\frac{s}{\sqrt{n}}$$

for the students' values:

$$\frac{0.73}{\sqrt{130}} = 0.06°\text{F}$$

Like the standard deviation measure, the standard error measure defines boundaries of probabilities. Remember from the earlier discussion that the sample standard error is equivalent to the standard deviation of the sample mean distribution. Therefore, there is around 68% probability that the true population mean lies within the boundaries of the sample mean ±1 sample standard error:

98.25 ± 0.06°F for a 68% confidence

(Students can infer with 68% confidence that the true mean for the population lies between 98.19 and 98.31°F.) Two sample standard errors on either side of the sample mean (98.13 to 98.37°F) define a region that the students can infer includes the true population mean with a little more than 95% confidence. See Figure 10.

	A	B	C	D	E	F	G	H
1	**temp.**	**bins**	**freq.**					
2	96.3	96	0					
3	96.7	96.3	1					
4	96.9	96.6	1					
5	97	96.9	3					
6	97.1	97.2	5					
7	97.1	97.5	9					
8	97.1	97.8	9					
9	97.2	98.1	23					
10	97.3	98.4	18					
11	97.4	98.7	22					
12	97.4	99	20					
13	97.4	99.3	11					
14	97.4	99.6	5					
15	97.5	99.9	0					
16	97.5	100.2	2					
17	97.6	100.5	0					
18	97.6	100.8	0					
19	97.6	101.1	1					
20	97.7	101.4				mean =	98.25	degrees F
21	97.8							
22	97.8					standard deviation =	0.73	degrees F
23	97.8							
24	97.8					standard error =	0.06	degrees F
25	97.9							
26	97.9					95% confidence interval =	0.13	degrees F
27	98							
28	98					95% confidence lower limit =	98.12	degrees F
29	98					95% confidence upper limit =	98.38	degrees F
30	98							
31	98							
32	98							
33	98.1							

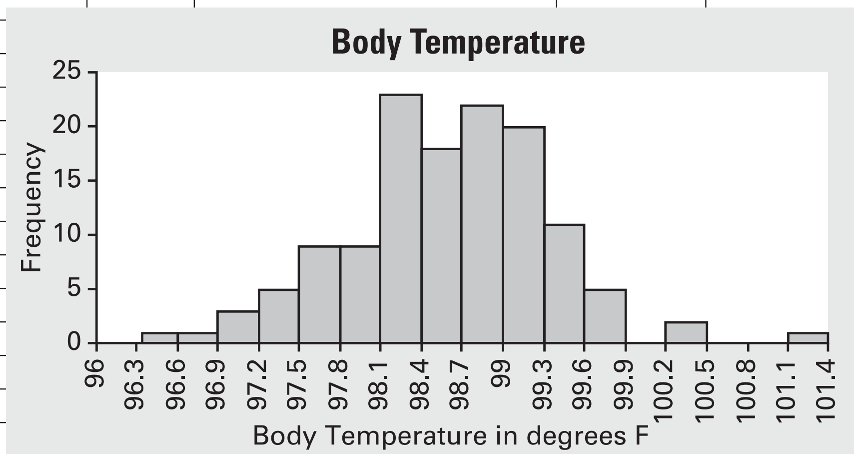


Figure 10. Sample Data Analysis Spreadsheet Showing Human Body Temperature Data

The students' data now have been described, and it is beginning to look like the mean of the body temperature in their sample is different from the assumed value of 98.6°F. Example 4 of Chapter 3: Hypothesis Testing will return to this example.

ALIGNMENT TO THE INVESTIGATIONS

Sample Alignment to Investigation 1: Artificial Selection

Can extreme selection change expression of a quantitative trait in a population in one generation?

Most investigations that help students derive an understanding of natural selection are either computer or structured simulations. An alternative for students to gain insight into natural selection is by studying artificial selection, just as Darwin relied on examples of artificial selection in farm animals to make his case in *On the Origin of Species*. Using Fast Plants® Seed, students identify and quantify several traits that vary in the population and then cross-pollinate selected plants, collect the seeds, and plant them. Students use statistical tools to analyze their data to determine if the observed phenotypes of the sample of the second-generation population differ significantly from the first-generation population.

Alignment to the Curriculum Framework

Quantitative reasoning is an essential part of inquiry in biology, and many mathematical tools, including statistical tests, were developed to work out biological problems. Quantitative analysis is involved in determining whether collected data support a hypothesis or conclusion, drawing conclusions from a data set, assessing the validity of experimental evidence, indentifying possible sources of errors in an experimental design or data set, recognizing outliers, and proposing explanations for them. Data analysis using statistical tools helps show variations in data, standard errors, best-fit functions, and confidence that sufficient data from the experimental sample(s) have been collected. Using statistics helps students inform their investigations by building on other quantitative skills such as graphing, data mining, and problem solving.

Investigation 1: Artificial Selection provides opportunities for students to design experimental plans, collect data, graph data, and use statistical tools to quantitatively analyze their data to draw conclusions and make predictions about artificial selection in a sample population.

Science Practices Applied

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

Learning Objectives Addressed

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time, and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change (1.A.1 & SP 1.5, SP 2.2).

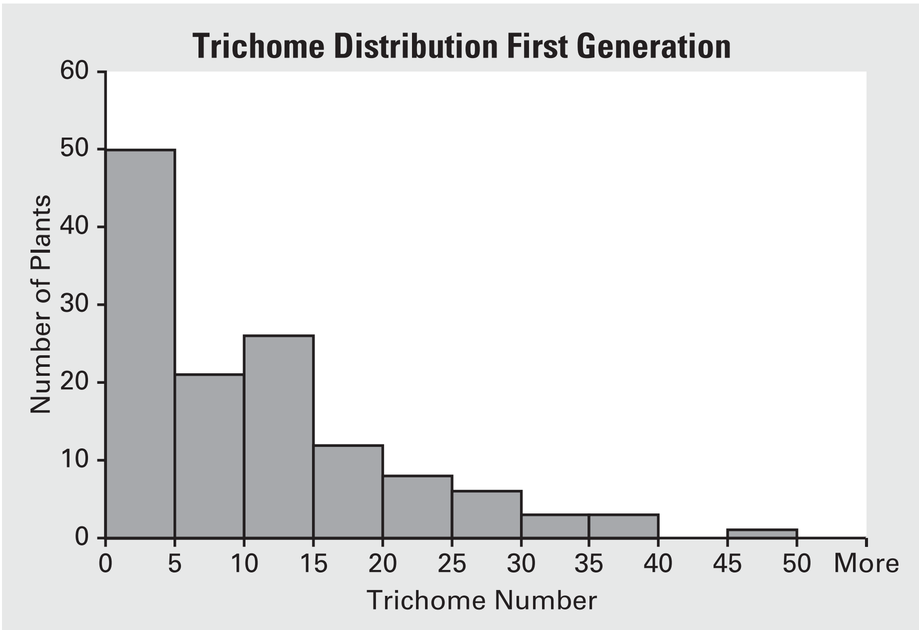
LO 1.2 The student is able to evaluate evidence provided by data to qualitatively and quantitatively investigate the role of natural selection in evolution (1.A.1 & SP 2.2, SP 5.3).

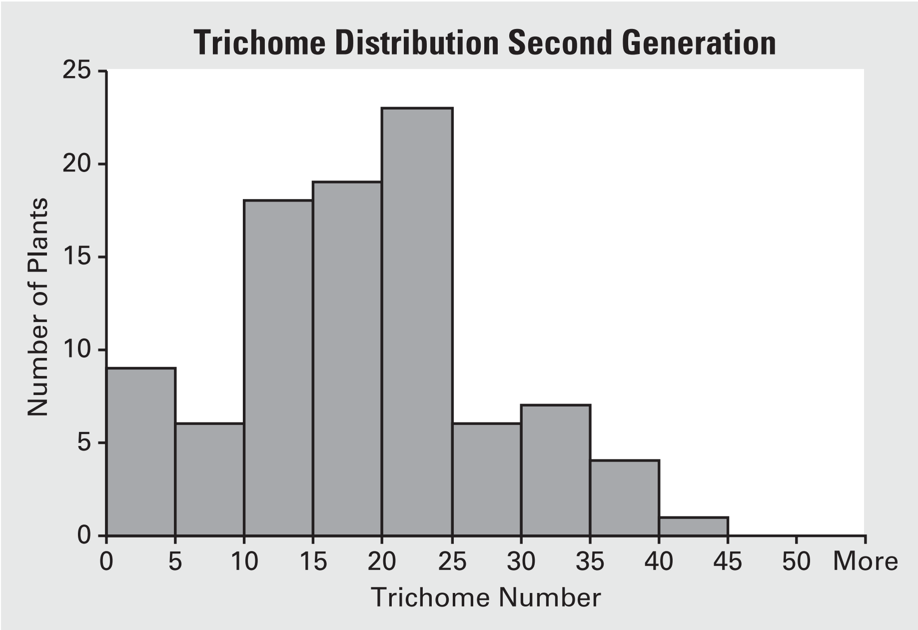
LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future (1.A.1 & SP 2.2).

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time (1.A.2 & SP 5.3).

Example of Sample Data Analysis

When an investigation involves measurement data (such as students will collect to quantify traits that vary in the population of Fast Plants®), a suggestion is to construct a histogram, or frequency graph, to represent the distribution of the trait, i.e., the number of plants exhibiting the trait within a specific interval. If enough measurements are made, data often show a normal or bell-shaped distribution on a histogram. Plant height at first flower would be an example of a normally distributed trait. The following two figures represent histograms that show typical first and second generation trichome distribution for Investigation 1: Artificial Selection:





The appropriate descriptive statistics for a normally distributed data set, such as plant height, include the mean (average), sample size, standard deviation, and standard errors. Plotting the means of two generations along with the ±2 standard errors defines the boundaries of uncertainty of the sample mean (with approximately a 95 percent confidence). However, while the second generation distribution of trichome numbers is approaching a normal distribution, the first generation is highly skewed. For this reason a student doing a trichome selection experiment may choose to employ nonparametric statistics outlined in the decision table in Table 1. Nonparametric methods are generally more conservative than parametric methods. With this analysis, the student can make inferences about the confidence she has that her sample population accurately represents the true population. The student now has a wealth of information about artificial selection to explore possible advantages or disadvantages that selected traits might confer on individuals in different environmental conditions. By applying inferential statistics, the student can draw conclusions and make predictions beyond the data analyzed that are subject to further investigation.

ALIGNMENT TO THE EXAM

Data Analysis and Exam Questions

Quantitative data analysis is one of the first steps toward determining whether an observed pattern has validity and helps distinguish among multiple working hypotheses. Data analysis includes tabular and graphical data display, descriptive statistics, and inferential statistics. Through data analysis, students can draw conclusions, support or refute hypotheses, and make predictions. The following sample exam questions from the *AP Biology Practice Exam* represent items that assess students' ability to critically analyze quantitative data to glean information about biological phenomena. The questions are based on the learning objectives in the curriculum framework.

Sample Multiple-Choice Question Using Data Analysis

An experiment to measure the rate of respiration in crickets and mice at 10°C and 25°C was performed using a respirometer, an apparatus that measures changes in gas volume. Respiration was measured in mL of O_2 consumed per gram of organism over several five-minute trials and the following data were obtained.

Organism	Temperature (°c)	Average respiration (mL O_2/g/min)
Mouse	10	0.0518
Mouse	25	0.0321
Cricket	10	0.0013
Cricket	25	0.0038

According to the data, the mice at 10°C demonstrated greater oxygen consumption per gram of tissue than did the mice at 25°C. This is most likely explained by which of the following statements?

(A) The mice at 10°C had a higher rate of ATP production than the mice at 25°C.

(B) The mice at 10°C had a lower metabolic rate than the mice at 25°C.

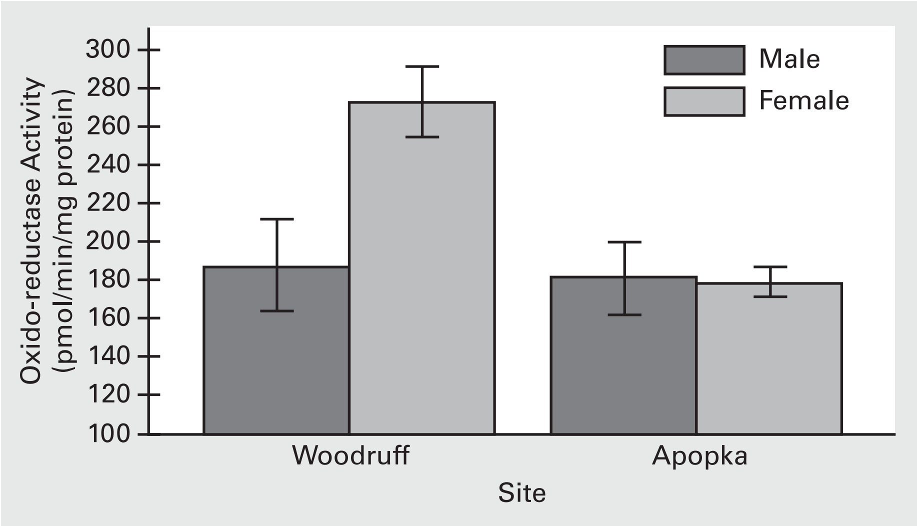
(C) The mice at 25°C weighed less than the mice at 10°C.

(D) The mice at 25°C were more active than the mice at 10°C.

Sample Multiple-Choice Question Using Data Analysis

Testosterone oxido-reductase is a liver enzyme that regulates testosterone levels in alligators.

One study compared testosterone oxido-reductase activity between male and female alligators from Lake Woodruff, a relatively pristine environment, and from Lake Apopka, an area that has suffered severe contamination. The following graph depicts the findings of that study.



The data in the graph best support which of the following claims?

(A) Environmental contamination elevates total testosterone oxido-reductase activity in females.

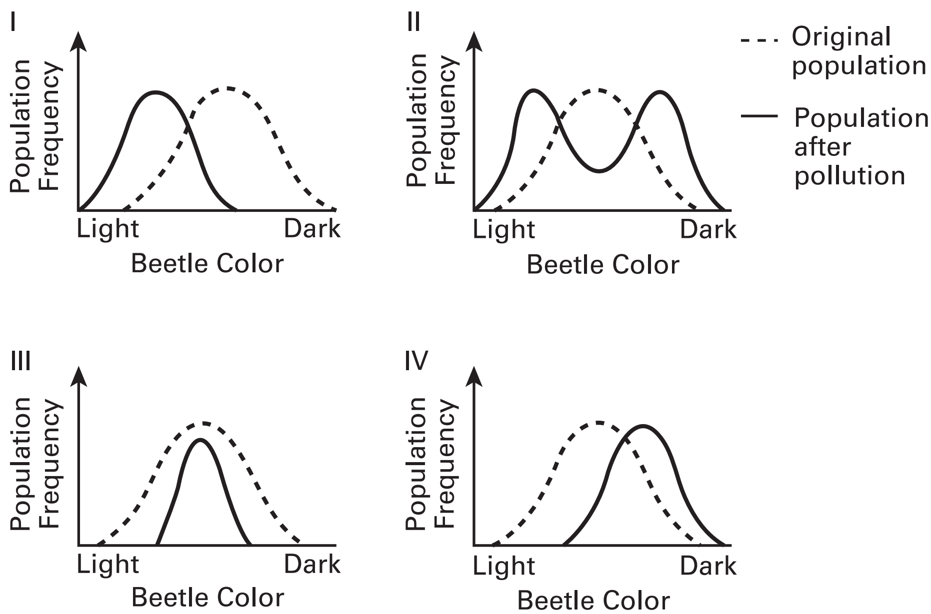
(B) Environmental contamination reduces total testosterone oxido-reductase activity in females.

(C) Environmental contamination elevates total testosterone oxido-reductase activity in males.

(D) Environmental contamination reduces total testosterone oxido-reductase activity in males.

Sample Multiple-Choice Question Using Data Analysis

In a hypothetical population of beetles, there is a wide variety of color, matching the range of coloration of the tree trunks on which the beetles hide from predators. The graphs below illustrate four possible changes to the beetle population as a result of a change in the environment due to pollution that darkened the tree trunks.



Which of the following includes the most likely change in the coloration of the beetle population after pollution and a correct rationale for the change?

(A) The coloration range shifted toward more light-colored beetles, as in diagram I. The pollution helped the predators find the darkened tree trunks.

(B) The coloration in the population split into two extremes, as in diagram II. Both the lighter-colored and the darker-colored beetles were able to hide on the darker tree trunks.

(C) The coloration range became narrower, as in diagram III. The predators selected beetles at the color extremes.

(D) The coloration in the population shifted toward more darker-colored beetles, as in diagram IV. The lighter-colored beetles were found more easily by the predators than were the darker-colored beetles.

Sample Free-Response Question Using Data Analysis

Fruit flies (*Drosophila melanogaster*) with a wild-type phenotype have gray bodies and red eyes. Certain mutations can cause changes to these traits. Mutant flies may have a black body and/or cinnabar eyes. To study the genetics of these traits, a researcher crossed a true-breeding wild-type male fly (with gray body and red eyes) with a true-breeding female fly with a black body and cinnabar eyes. All of the F1 progeny displayed a wild-type phenotype.

Female flies from the F_1 generation were crossed with true-breeding male flies with black bodies and cinnabar eyes. The table below represents the predicted outcome and the data obtained from the cross. **Explain** the difference between the expected data and the actual numbers observed.

F_2 Generation Phenotypes			
Body Color	Eye Color	Number Predicted	Number Observed
Gray	Red	244	455
Black	Cinnabar	244	432
Gray	Cinnabar	244	42
Black	Red	244	47

RESOURCES

The following resources include ideas for classroom activities and lessons, as well as examples for learning and teaching as you further your exploration into the new AP Biology curriculum. Each resource includes the approximate time involved, a description of the resource and where to find it, the quantitative skills the activity or lesson involves, and where those skills might be applied.

HANDBOOK OF BIOLOGICAL STATISTICS http://udel.edu/~mcdonald/statintro.html	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

This online book provides a clear explanation of many statistical tests. The explanation of the Chi-square test for goodness of fit will provide a solid background, including how the test works, how to graph results, and how to formulate the null hypothesis. This book is a valuable resource when you want to apply other statistical tests to inquiry labs.

SCIENCE PRACTICES APPLIED

SP 2 The student can use mathematics appropriately.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVE ADDRESSED

LO 3.14 The student is able to apply mathematical routines to determine Mendelian patterns of inheritance provided by data sets. [See SP 2.2]

Any laboratory investigation that involves genetic crosses and predictions, or any laboratory investigation in which statistics can be applied.

EXERCISES IN DATA ANALYSIS **http://www.math.dartmouth.edu/~matc/DataAnalysis/DataLabs.html**	**APPROXIMATE TIME** **1–2 class periods for each exercise**

DESCRIPTION OF RESOURCE

Exercises in Data Analysis includes four lessons in data analysis that can easily be done in a high school classroom. The topics include quantitative analysis of the distribution of colors of M&Ms from several sources, standard deviation analysis of peas in pods, use of the T-test to analyze radish seedling growth, and use of statistical analysis on measurements of various body parts (e.g., arm length, height, and foot length). The authors provide charts for student data plus a set of questions and suggestions for a report.

SCIENCE PRACTICES APPLIED

SP 2 The student can use mathematics appropriately.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluate evidence.

THE BIOLOGY PROJECT: BIOMATH http://www.biology.arizona.edu/biomath/tutorials/Quadratic/QuadraticFunctionApplications/QuadraticFunctionsApplication.html	APPROXIMATE TIME 1–2 class periods

DESCRIPTION OF RESOURCE

The Biomath section of The Biology Project, a Web resource, gives students practice in using quadratic functions in a logistical population model and in population genetics. Clear explanations and practice problems help students understand these sometimes confusing topics. The population genetics extends the Hardy-Weinberg theorem to include fitness, thus relating these calculations to selection. This resource can be used in class or as a homework assignment.

SCIENCE PRACTICE APPLIED

SP 2 The student can use mathematics appropriately.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.2 The student is able to evaluate evidence provided by data to qualitatively and quantitatively investigate the role of natural selection in evolution. [See SP 2.2, 5.3]

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

MEDICAL TESTING http://www.math.dartmouth.edu/~mqed/UNR/MedicalTesting/MedicalTesting_z65.pdf	APPROXIMATE TIME 1+ class periods

DESCRIPTION OF RESOURCE

This medical testing module is part of a series produced at the University of Nevada at Reno for the National Numeracy Network under a grant from the Woodrow Wilson Fellowship Foundation through the National Council on Education and the Disciplines. It explores the probabilities and percentages involved in medical testing, including genetic testing. The lessons in this module can be used as an extension into the issues surrounding medical testing. An instructor's sheet is included to help you guide discussion and explain calculations.

SCIENCE PRACTICES APPLIED

SP 2 The student can use mathematics appropriately.

SP 3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

SP 5 The student can perform data analysis and evaluation of evidence.

SCIENCE PRACTICES ADDRESSED

LO 3.13 The student is able to pose questions about ethical, social or medical issues surrounding human genetic disorders. [See SP 3.1]

LO 3.14 The student is able to apply mathematical routines to determine Mendelian patterns of inheritance provided by data sets. [See SP 2.2]

BIOQUEST CURRICULUM CONSORTIUM: EVOLUTION THROUGH NATURAL SELECTION http://bioquest.org/esteem/esteem_details.php?product_id=7080#	APPROXIMATE TIME 1+ class periods (depending on which lessons are used)

DESCRIPTION OF RESOURCE

The student works with the Hardy-Weinberg equation in a computer model to simulate natural selection. As described by the author, "This workbook simulates the population genetics of a single gene with two alleles, allowing the user to set the initial allele frequencies and enter parameters for a variety of different selection models; the program then tracks the population through up to 10,000 generations." This activity can be done in class or as a homework assignment.

SCIENCE PRACTICES APPLIED

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 5 The student can perform data analysis and evaluation of evidence.

SP 6 The student can work with scientific explanations and theories.

LEARNING OBJECTIVES ADDRESSED

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.7 The student is able to justify data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and the effects of selection in the evolution of specific populations. [See SP 2.1]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 1.25 The student is able to describe a model that represents evolution within a population. [See SP 1.2]

LO 2.28 The student is able to use representations or models to analyze quantitatively and qualitatively the effects of disruptions to dynamic homeostasis in biological systems. [See SP 1.4]

LO 3.24 The student is able to predict how a change in genotype, when expressed as a phenotype, provides a variation that can be subject to natural selection. [See SP 6.4, 7.2]

POPULATION DYNAMICS FROM MATHBENCH BIOLOGY MODULES http://mathbench.umd.edu/homepage/popn_dynamics.htm	APPROXIMATE TIME 1+ class periods (depending on which lessons are used)
DESCRIPTION OF RESOURCE	
The activities in the Population Dynamics section of the MathBench Biology Modules provide clarification of, and extensions to, concepts learned in class. Lessons such as “The Mystery of the Missing Housefly,” “Bacterial Dynamics: *E. coli* ate my homework,” and “Mutation and Equilibrium” lead students into the math behind exponential growth, carrying capacity, and equilibrium. These activities can be used in class, as a homework assignment, or as extra credit.	
SCIENCE PRACTICES APPLIED	
SP 2 The student can use mathematics appropriately. SP 5 The student can perform data analysis and evaluation of evidence. SP 6 The student can work with scientific explanations and theories.	
LEARNING OBJECTIVES ADDRESSED	
LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2] LO 4.25 The student is able to use evidence to justify a claim that a variety of phenotypic responses to a single environmental factor can result from different genotypes within the population. [See SP 6.1] LO 4.26 The student is able to use theories and models to make scientific claims and/or predictions about the effects of variation within populations on survival and fitness. [See SP 6.4] LO 4.27 The student is able to make scientific claims and predictions about how species diversity within an ecosystem influences ecosystem stability. [See SP 6.4]	

CELLULAR PROCESSES FROM MATHBENCH BIOLOGY MODULES http://mathbench.umd.edu/homepage/cell_processes.htm	APPROXIMATE TIME 1+ class periods (depending on which lessons are used)

DESCRIPTION OF RESOURCE

Activities in the Cellular Processes section of the MathBench Biology Modules make topics such as diffusion and osmosis fun. Students use skills such as graphing and generating equations to analyze diffusion and osmosis. After working with the math involved in these concepts, students confront biological questions such as, *Why don't sharks get thirsty?* and *How long will it take a macrophage to search and destroy a virus in your lungs?* These activities suit in-class work as well as homework or extra credit assignments.

SCIENCE PRACTICES APPLIED

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

SP 7 The student is able to connect and relate knowledge across various scales, concepts, and representations in and across domains.

LEARNING OBJECTIVES ADDRESSED

LO 2.9 The student is able to represent graphically or model quantitatively the exchange of molecules between an organism and its environment, and the subsequent use of these molecules to build new molecules that facilitate dynamic homeostasis, growth, and reproduction. [See SP 1.1, 1.4]

LO 2.10 The student is able to use representations and models to pose scientific questions about the properties of cell membranes and selective permeability based on molecular structure. [See SP 1.4, 3.1]

LO 2.11 The student is able to construct models that connect the movement of molecules across membranes with membrane structure and function. [See SP 1.1, 7.1, 7.2]

LO 2.12 The student is able to use representations and models to analyze situations or solve problems qualitatively and quantitatively to investigate whether dynamic homeostasis is maintained by the active movement of molecules across membranes. [See SP 1.4]

CHAPTER 3:
Hypothesis Testing

LABS USING HYPOTHESIS TESTING

1: Artificial Selection
2: Mathematical Modeling: Hardy-Weinberg
3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST
4: Diffusion and Osmosis
5: Photosynthesis
6: Cellular Respiration
7: Cell Division: Mitosis and Meiosis
8: Biotechnology: Bacterial Transformation
9: Biotechnology: Restriction Enzyme Analysis of DNA
10: Energy Dynamics
11: Transpiration
12: Fruit Fly Behavior
13: Enzyme Activity

Once students have graphed their data and completed a preliminary data analysis, they may notice a pattern in the data that they would like to investigate further. For example, in Example 1 in Chapter 2, there appears to be a definite size difference between shady and sunny leaves in English ivy. Even though the distributions of leaf size overlap between the two populations, it appears that the shady leaves tend to run larger on average. How might students test this observation? This is where hypothesis testing comes into play during a biological investigation.

INTRODUCTION TO HYPOTHESIS TESTING

A **hypothesis** is a statement explaining that a causal relationship exists between an underlying factor (variable) and an observable phenomenon.Often, after making an observation, you might propose some sort of tentative explanation for the phenomenon; this could be called your *working hypothesis*. Because absolute proof is not possible, statistical **hypothesis testing** focuses on trying to reject a null hypothesis. A **null hypothesis** is a statement explaining that the underlying factor or variable is independent of the observed phenomenon—there is no causal relationship. For example, in the leaf study introduced in Chapter 2, an appropriate null hypothesis might be that the distributions of the leaf widths in sunny and shady habitats are the same—in other words, that there is no difference between the two populations. The **alternative to the null hypothesis** might be that there is a size difference between the two populations. Usually (but not always), an investigator is trying to find an alternative to the null hypothesis—evidence that supports the alternative hypothesis by rejecting the null (based on statistical tests). In the leaf case, one biologically interesting alternative hypothesis might be that leaves in low light grow larger, which allows them to capture more photons of light for photosynthesis. If the null hypothesis (that there is no difference between shady and sunny leaves) can be rejected, then that is support for this alternative hypothesis.

It is important to realize that hypothesis testing does not allow proof, or even acceptance, of the alternative to the null hypothesis. Typically, the decision comes down to whether there is enough evidence to reject the null hypothesis. If evidence to reject the null hypothesis is sufficient, what can be said is that the investigator rejects the null hypothesis—*not that the investigation has proven the alternative hypothesis*. This is a crucial concept for students to understand.

It is also important to remember that a hypothesis and a prediction are different from each other. A hypothesis is a testable statement explaining some relationship between cause and effect, while a prediction is a statement of what you think will happen given certain circumstances. For example, a hypothesis might state that because chromosomes assort independently of one another during meiosis, the gametes carry the random combination of these assorted chromosomes. A prediction, in contrast, might state that when an F1 double heterozygote (a dihybrid) with simple dominance at each independently assorting locus is backcrossed with its doubly homozygous recessive parent, then four genotypically and phenotypically different types of progeny will appear in approximately equal amounts (1:1:1:1 ratio).

If a null hypothesis is clearly rejected (rejected at a significant statistical level, which will be explained shortly), it is usually a good practice in hypothesis testing to develop a second prediction or working hypothesis (especially if the result is counterintuitive) and to perform a statistical test for that option. Similarly, it greatly helps investigators to zero in on the causal relationship they are studying if they can reject an alternative hypothesis that might reasonably account for an observable outcome. An investigator who begins with multiple working (alternative) hypotheses is likely to remain more open scientifically—to not be tied to a single explanation. If two hypotheses can account for observations equally well, scientists normally prefer the simpler hypothesis.

In data analysis, investigators determine the size and confidence they have in various population parameters that were measured, counted, or calculated during the course of the investigation. Hypothesis testing asks the question, *Is there something to these measurements?* or *Is the effect real?*

Hypothesis testing uses various parameters that were introduced in Chapter 2: Data Analysis, such as sample standard error, sample mean, sample variance, and sample median, to calculate comparative **test statistics** (computational procedures that help characterize the data). These test statistics are used to determine a probability distribution of samples—the **null distribution.** The test statistic calculated based on the investigator's data can then be compared to the null distribution. This comparison can be used to determine a precise probability of whether or not the results or even more extreme results occur by chance alone, if the null hypothesis is indeed true. Many different test statistics have been developed for particular types of data and investigative questions. By convention, most biological studies establish a **critical value** of the probability of whether the results or even more extreme results occur by chance alone, if the null hypothesis is indeed true (**probability value,** or **p-value,** of less than 5%). The critical value is a predetermined boundary condition that the investigator establishes before the study.

Probabilities less than a certain critical value will be taken as evidence that the null hypothesis may be false, and probabilities greater than the same critical value will not be taken as evidence that the null hypothesis may be false. Since absolute proof is not possible, how certain do we need to be before coming to a conclusion? Remember, in hypothesis testing, the investigator is testing to see if he or she can reject the null hypothesis. When making this decision, there are four possibilities, as shown in Table 5.

Table 5. The Four Possible Outcomes of Hypothesis Testing

	Possible Outcomes	
Investigator Action	Null hypothesis is true	Null hypothesis is false
Rejects the null	Type I error (false positive) (Example: A blood test indicates a pregnancy when the woman is not pregnant.)	Correct
Fails to reject the null	Correct	Type II error (false negative) (Example: A blood test fails to indicate a pregnancy when the woman actually is pregnant.)

When an investigator's observations in an experiment deviate from the predictions, how much variation should be tolerated before rejecting the null hypothesis? In biological investigations, a 5% critical value is often used as a decision point for rejecting the null hypothesis. There's no reason, though, why an investigator could not set a more stringent critical value of, say, 1% or 0.1%. In life-and-death issues often associated with medical studies, for example, the critical values are often more stringent.

Revisiting Example 1 in Chapter 2, the null hypothesis stated that there is no difference in leaf size between sunny and shady leaf populations. If, after calculating an appropriate test statistic, the student generated a critical or probability value (p-value) of less than 5%, then the student should reject the null hypothesis and state that there is evidence to support that there is a difference between the leaf width in the two populations. With this decision, the student would have only a 5% probability that she had made a Type I error—at least for this round of investigation (see Table 5). In other words, the null hypothesis is false, and if the student repeated the investigation 20 times, she would expect similar results 19 out of 20 times.

USING HYPOTHESIS TESTING

The lab manual presents many opportunities to explore a technique or a question. It should be apparent at this point that graphing data and thinking through data analysis and hypothesis testing before conducting an investigation should greatly improve the experimental design. This preliminary work is often necessary to get a handle on the various population parameters that will be studied in order to determine the kinds of hypotheses and tests that should be involved.

The following steps suggest one path students might follow as they include hypothesis testing in their experimental design. These steps are not set in stone but are offered as guidelines to consider.

1. Choose a statistical test based on the question and types of data that will be collected (see Table 6).

Table 6. Decision Table for Choosing Tools to Use Based on Data Collected and Questions Asked

<table>
<tr><td colspan="3"></td><td>Parametric Tests (normal data)</td><td>Nonparametric Tests</td><td>Frequency Tests (counts)</td></tr>
<tr><td colspan="3">Descriptive Statistics</td><td>Mean, Standard Deviation, Standard Error, 95% CI</td><td>Median, Quartiles, Interquartile Range</td><td>Percent in Each Category</td></tr>
<tr><td rowspan="5">Comparative Statistics</td><td colspan="2">Graph Type</td><td>Bar Graph</td><td>Box-and-Whisker Plot</td><td>Bar Graph or Pie Chart</td></tr>
<tr><td rowspan="2">Independent Samples</td><td>2 groups</td><td>Unpaired T-test</td><td>Mann-Whitney U-test</td><td rowspan="4">Chi-square Test</td></tr>
<tr><td>≥ 2 groups</td><td>Anova</td><td>Kruskal-Wallis Test</td></tr>
<tr><td rowspan="2">Matched Samples</td><td>2 groups</td><td>Paired T-test</td><td>Wilcoxon Test</td></tr>
<tr><td>≥ 2 groups</td><td>Matched Anova</td><td>Friedman Test</td></tr>
<tr><td rowspan="3">Association Statistics</td><td colspan="2">Graph Type</td><td>Scatterplot</td><td>Scatterplot</td><td>Mosaic Chart</td></tr>
<tr><td colspan="2">Test for an Association</td><td>Pearson Correlation</td><td>Spearman Rank Correlation</td><td>Chi-square Test for Association</td></tr>
<tr><td colspan="2">Linear Relationship</td><td>Linear Regression</td><td>–</td><td>–</td></tr>
<tr><td colspan="6">Source: Redrawn from “Statistics for AS Biology,” available as part of a download at: http://www.heckmondwikegrammar.net/index.php?highlight=introduction&p=10310</td></tr>
</table>

2. State the null and alternative hypotheses as precisely as possible.
3. Design and carry out the investigation. (Use exploratory data to help determine adequate sample size, measured parameters, and sampling technique.)
4. Conduct a data analysis, and present graphical and tabular summaries of the data. (Do not report all of the data in the final report—only summary statistics.)

5. Carry out the statistical tests. Include the sample size, test statistic chosen, and p-values in the data reports.
6. Make a conclusion, always stating the amount of evidence in terms of the alternative hypothesis.

The examples that follow illustrate these steps of hypothesis testing. The examples in this manual will cover only a few commonly used test statistics to provide an idea of how students might proceed with their own investigations. Realize that students may need to research other techniques depending on the question, data, and design.

EXAMPLES OF HYPOTHESIS TESTING

These first two examples utilize the *t*-statistic as a test-statistic while making comparisons or looking for associations. The next examples show questions and data using the well known Chi-square statistic. The final examples will look for trends and correlations. Each of the examples works through the aforementioned steps—but not necessarily in the same order. Refer to Table 6 and the guidelines presented earlier as you work through these examples.

Example 1: English Ivy

Let's return to the first leaf study example of Chapter 2. In this case the data have already been collected, and descriptive statistics and graphical and tabular summaries of the data have been presented. Figure 11 summarizes the student's findings comparing the widths of ivy leaves growing in deep shade and in bright sun.

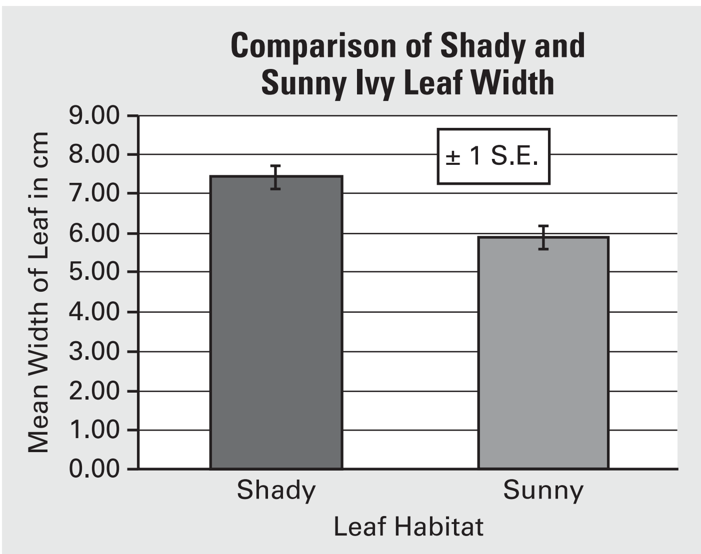


Figure 11. The Widths of Ivy Leaves Growing in Deep Shade and in Bright Sun

The error bars define the range of ±1 standard error. These data suggest that the two populations are significantly different from each other. Based on an understanding of the light requirements for photosynthesis, it makes some sense that shady leaves tend to be larger than sunny leaves to capture more light. These data support that conjecture.

But can this be quantified? This is where statistical hypothesis testing comes into play, as follows:

1. For this example the student established a null hypothesis (H_0) and one alternative hypothesis (H_1):
   a. H_0 = The true population mean width of ivy leaves grown in the shade is the same as the true population mean width of ivy leaves grown in the sun. (This is one of many possible ways of stating that the two populations are the same.) Another way to say this is, *There is no difference between the two true population means.*
   b. H_1 = The true population mean width of ivy leaves grown in the shade is not the same as the true population mean width of ivy leaves grown in the sun. (This is one of many possible ways of stating that the two populations are not the same.) Another way to say this is, *There is a significant difference between the two population means.*
2. The student established the critical value for significance in the investigation before going further. For this investigation she chose to go with convention and indicated that she would accept a p-value of 5% for the critical value.
3. Here is where the decision table can be used to help determine the most appropriate test statistic to apply. For this study, leaf width is a continuous measurement, the data appear to be approximately normally distributed, and the samples are equal and adequate in size. Therefore, we'll select the column for parametric data. The student's question called for testing for a difference between two groups of leaf samples. These samples are independent of each other (neither affects the other). Based on Table 6, the unpaired T-test is likely the best choice of test statistic for this investigation. The T-test helps determine how different two sample populations are from one another by comparing means and standard errors. Calculate the test statistic (using Excel, a calculator, or another statistics software). Based on the test statistic's sampling distribution, determine the p-value. (This is done in the spreadsheet calculations, but students can also use a table as described in a later example.) In Excel, both steps are accomplished with one formula: =T.TEST(array1, array2, tails, type). For this investigation, the parameters in the Excel formula are as follows: "array1 and array2" are the leaf width data, the "tails" equals 2, and the "type" is 2, unpaired. The difficult part here is the tails. This refers to the t-distribution itself and whether you should be looking at only one side of the distribution or both. You might use a search engine to research a complete description.
4. For the two leaf populations, the p-value calculates to 0.016%.
5. A p-value of 0.016% is less than the 5% critical value established earlier. The student in the example decided to reject the null hypothesis that there is no difference between the means of the two populations, in favor of the alternate hypothesis that there is a difference. Note that the student does not accept the alternative hypothesis, but this analysis provides support for the biological hypothesis that low-light leaves are larger than bright-light leaves in English ivy. These results are only one supportive piece of evidence for this hypothesis and point to additional studies that might be made. For example, additional studies might focus on chlorophyll amounts, leaf area, stomata densities, or light response curves. Each such study would be a piece of a

puzzle to answer the larger question about the similarities, or differences, between the sunny and shady leaves.

Example 2: Body Temperature

Let's return to the body temperature and heart rate data in Example 2 at the end of Chapter 2. Recall that it was beginning to look like the mean of the body temperature of the sample is different from the accepted value of 98.6°F. Is the difference between the measured sample mean body temperature (98.25°F) and the accepted value for body temperature (98.6°F) meaningful? In other words, are these really two different measurements, or are they just different due to chance?

According to the calculation of the sample standard error, we can be 95% confident that the true population mean lies between 98.13°F and 98.37°F. Because 98.6°F lies outside those boundaries, is there sufficient evidence to claim that 98.6°F is wrong and that 98.25°F is a more accurate representation of the true population mean? Again, we need to turn to probability distributions to help them make this claim. The T-test can be used to determine how different two sample populations are from one another by comparing means and standard errors. This is a special case of comparing the mean of one distribution (the sample body temperature readings) to a long accepted mean for normal human body temperature. This particular case is not on the decision table in Figure 6.

Figure 12 shows the spreadsheet updated to include this analysis.

To compare the measured sample mean of 98.25°F along with sample standard error to the accepted population mean of 98.6°F, we need to calculate a one-sample T-test statistic. The formula for this is:

$$\frac{(\text{sample mean} - \text{population mean})}{\text{the sample standard error}}$$

or

$$\frac{(98.25\text{-}98.6)}{0.06} = \text{a t-value of } -5.45$$

What does that t-value mean? In this case, the t-value expresses the difference between the two means in terms of the sample standard error. Now consult a T-test table (Table 7 or one available on the Web), or have the spreadsheet program calculate the p-value associated with such a t-value as in Figure 12. The investigator also needs to know the **degrees of freedom** (d) to estimate the p-value. The degrees of freedom refer to the number of ways the values involved in the calculation can vary. They are calculated as *one less* than the number of possible results in the experiment. For this example, $d = N - 1$, or, $130 - 1 = 129$.

	A	B	C	D	E	F	G	H
1	**temp.**	**bins**	**freq.**					
2	96.3	96	0					
3	96.7	96.3	1					
4	96.9	96.6	1					
5	97	96.9	3					
6	97.1	97.2	5					
7	97.1	97.5	9					
8	97.1	97.8	9					
9	97.2	98.1	23					
10	97.3	98.4	18					
11	97.4	98.7	22					
12	97.4	99	20					
13	97.4	99.3	11					
14	97.4	99.6	5					
15	97.5	99.9	0					
16	97.5	100.2	2					
17	97.6	100.5	0					
18	97.6	100.8	0					
19	97.6	101.1	1					
20	97.7	101.4				mean =	98.25	degrees F
21	97.8							
22	97.8					standard deviation =	0.73	degrees F
23	97.8							
24	97.8					standard error =	0.06	degrees F
25	97.9							
26	97.9					95% confidence interval =	0.13	degrees F
27	98							
28	98					95% confidence lower limit =	98.12	degrees F
29	98					95% confidence upper limit =	98.38	degrees F
30	98							
31	98					*t statistic* =	–5.45	
32	98					degrees of freedom = N – 2 or 128		
33	98.1							
34	98.1					*p* 1-tail =	1.22E-07	
35	98.2					*p* 2-tail =	2.44E-07	
36	98.2							
37	98.2							
38	98.2							
39	98.3							
40	98.3							

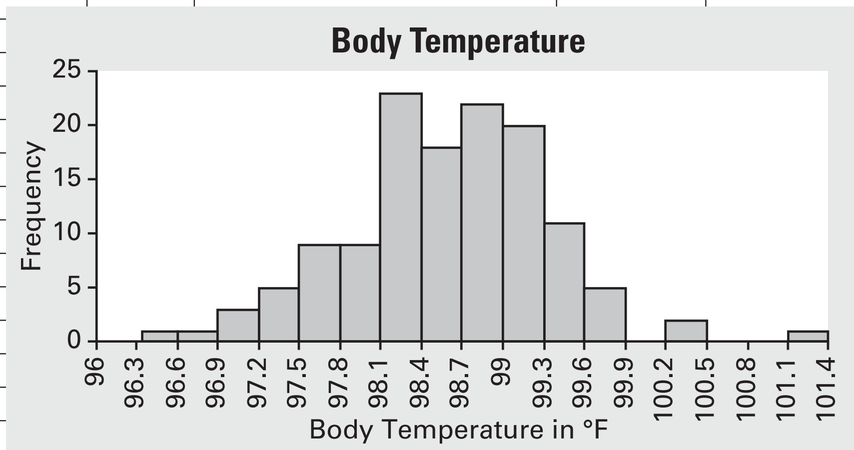


Figure 12. Spreadsheet Updated to Include a Calculation for a One-Sample T-test and a Determination of the p-value for That Statistic

Table 7 shows a condensed T-table. Note that the degrees of freedom go across the top of the column and max out at 100. Also note that the t-values are in two columns—one going from 0 to1.5 and the other going from 1.6 to 3. Where do you read the p-value for $t = -5.45$ and $d = 129$? The minus sign only indicates to which direction the measured sample mean lies relative to the expected value of 98.6°F, so it's best to ignore it. You can't read the actual p-value using this table, though, because the t-value is larger than 3 and the degrees of freedom (129) is larger than 100; the best we can do with the table is to say that the p-value is less than 0.003, which is the p-value for 100 degrees of freedom and a t-value of 3.0. For this reason you may wish to rely on the spreadsheet to calculate the p-value of the t-distribution.

Table 7. A Condensed T-table

t	Degrees of Freedom (*d*) →							*t*	Degrees of Freedom (*d*) →						
	4	6	10	15	20	30	100		4	6	10	15	20	30	100
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000								
0.1	0.925	0.924	0.922	0.922	0.921	0.921	0.921	1.6	0.185	0.161	0.141	0.130	0.125	0.120	0.113
0.2	0.851	0.848	0.845	0.844	0.844	0.843	0.842	1.7	0.164	0.140	0.120	0.110	0.105	0.099	0.092
0.3	0.779	0.774	0.770	0.768	0.767	0.766	0.765	1.8	0.146	0.122	0.102	0.092	0.087	0.082	0.075
0.4	0.710	0.703	0.698	0.695	0.693	0.692	0.690	1.9	0.130	0.106	0.087	0.077	0.072	0.067	0.060
0.5	0.643	0.635	0.628	0.624	0.623	0.621	0.618	2	0.116	0.092	0.073	0.064	0.059	0.055	0.048
0.6	0.581	0.570	0.562	0.557	0.555	0.553	0.550	2.1	0.104	0.080	0.062	0.053	0.049	0.044	0.038
0.7	0.523	0.510	0.500	0.495	0.492	0.489	0.486	2.2	0.093	0.070	0.052	0.044	0.040	0.036	0.030
0.8	0.469	0.454	0.442	0.436	0.433	0.430	0.426	2.3	0.083	0.061	0.044	0.036	0.032	0.029	0.024
0.9	0.419	0.403	0.389	0.382	0.379	0.375	0.370	2.4	0.074	0.053	0.037	0.030	0.026	0.023	0.018
1	0.374	0.356	0.341	0.333	0.329	0.325	0.320	2.5	0.067	0.047	0.031	0.025	0.021	0.018	0.014
1.1	0.333	0.313	0.297	0.289	0.284	0.280	0.274	2.6	0.060	0.041	0.026	0.020	0.017	0.014	0.011
1.2	0.296	0.275	0.258	0.249	0.244	0.240	0.233	2.7	0.054	0.036	0.022	0.016	0.014	0.011	0.008
1.3	0.263	0.241	0.223	0.213	0.208	0.204	0.197	2.8	0.049	0.031	0.019	0.013	0.011	0.009	0.006
1.4	0.234	0.211	0.192	0.182	0.177	0.172	0.165	2.9	0.044	0.027	0.016	0.011	0.009	0.007	0.005
1.5	0.208	0.184	0.165	0.154	0.149	0.144	0.137	3	0.040	0.024	0.013	0.009	0.007	0.005	0.003

To use a computerized spreadsheet to more precisely calculate the p-value, use the TDIST(t-value, degrees of freedom, tails) function. For the t-value of –5.45, the p-value is 0.00000024. Such a small p-value would indicate that the measured sample mean body temperature of 98.25°F is likely different from the assumed value of 98.6°F. In fact, there is only about a 2 in 10,000,000 chance that the differences seen here are due to chance, if in fact the true body temperature were 98.6°F. If the sample size is large enough, and if the observations are independent of each other, then this is strong evidence to make the claim that 98.6°F is not the correct value for the "normal" body temperature—at least for the 130 individuals the students tested. In fact, this was the claim made by Mackowiak, Wasserman, and Levine in a 1992 *Journal of the American Medical Association* article. There is more to the story, but there is also more to explore in this data set.

Comparing Two Samples and Testing for Differences

To this point, we have only explored the body temperature data. But Mackowiak, Wasserman, and Levine also kept track of the gender of each volunteer. *Is there a difference in body temperature between males and females?* To answer this question, we will first need to separate the data on males from the data on females and treat them as individual sample populations—two sample populations but with the same treatment. In biological studies, it is common to compare two sample distributions to determine if there is a difference. Students will likely use this same approach with other investigations.

Let's start:

1. Construct two histograms—one for males and one for females—to get a visual image of their temperature distributions, the overlap of the data, and the shape of the histograms.
2. Calculate descriptive statistics.
3. Display the descriptive statistics in bar graphs.
4. Test for differences between males and females.

It is much simpler to compare two sample distributions if they have the same parameters but vary only in the sample means. For instance, in this data set there are the same number of data points for males and females. What about the other parameters? Are they both approximately normal distributions? Graph the two distributions on the same axes and find out (see Figure 13).

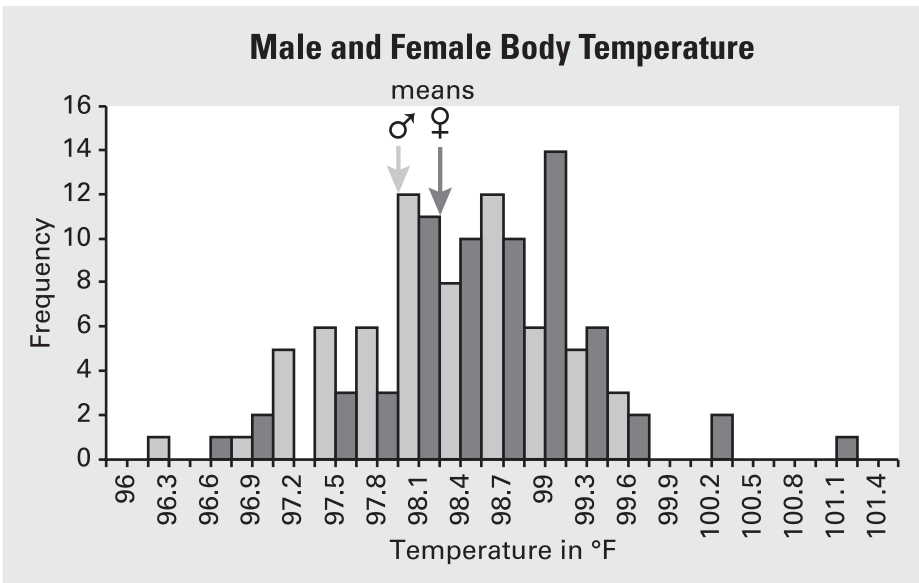


Figure 13. A Comparison of Male and Female Body Temperature from the Sample Data

Note that the two sample distributions overlap but have slightly different sample means. At this point it is too early to say that these two sample distributions are different. The difference is so small that it may be due to chance. While the means are different, the overlap (if we ignore the outliers) is almost 100%. It appears that both distributions resemble the "bell" shape of the normal distribution. With this information, we can proceed with the analysis.

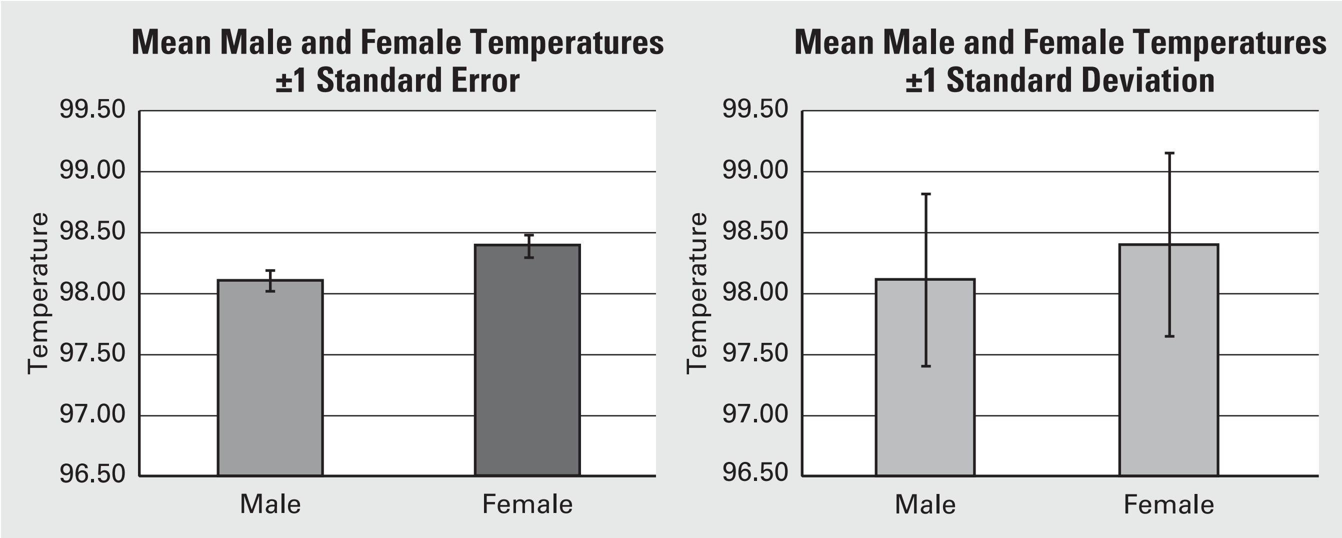


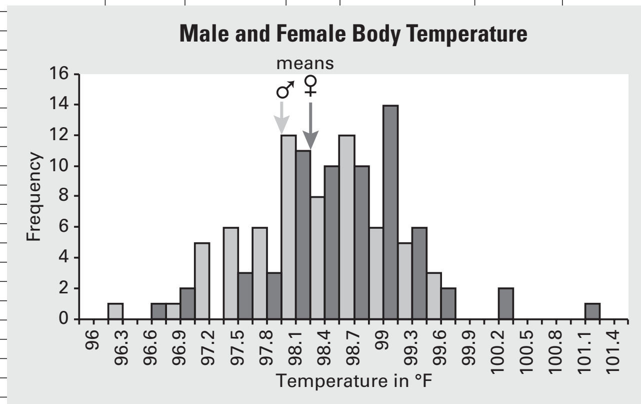
Figure 14. A Comparison of Standard Error and Standard Deviation

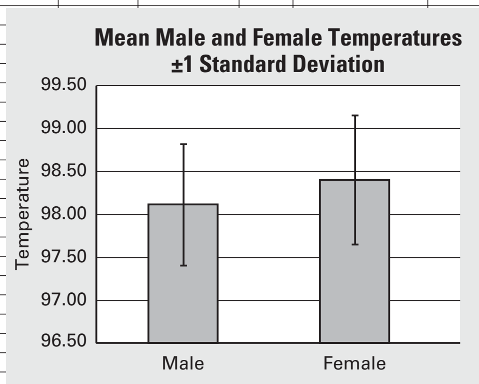
Remember that the estimated standard error gives us an idea about how confident we are that we have estimated the actual true mean of either population. The sample mean plus or minus one sample standard error defines a boundary where we can be confident that approximately 68% of the time the true mean lies within this boundary. (If the error bars were ±2 sample standard errors, then our confidence would rise to just over 95%.) In Figure 13, note that the error bars for the two distributions (male and female) do not overlap. That's an important clue that indicates that the true parameter may indeed be different. This type of estimate is pretty straightforward and easy to calculate. For a rough first estimate of the significance of results in an investigation like this, consider constructing bar graphs with ±2 estimated standard error bars.

Now consider a more rigorous test (Figure 15). Testing for differences with the two-sample T-test gives a p-value of 0.024—a small probability. What does this mean? The 0.024 p-value is a measure of the probability that the difference observed, or an even bigger difference, could come from means that were actually the same—there's only a 24 in 1,000 chance that such a sampling could occur with identical means. With such a small p-value, we can reject the hypothesis that the mean body temperature in males and females is the same. Thus, they can conclude that the mean body temperature of males in their sample is different from the mean body temperature of females in their sample.

The next section explores hypothesis testing with the Chi-square test statistic.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	**male**		**female**												
2	96.3		96.4												
3	96.7		96.7		bins	male freq	female freq								
4	96.9		96.8		96	0	0								
5	97		97.2		96.3	1	0								
6	97.1		97.2		96.6	0	1								
7	97.1		97.4		96.9	1	2								
8	97.1		97.6		97.2	5	0								
9	97.2		97.7		97.5	6	3								
10	97.3		97.7		97.8	6	3								
11	97.4		97.8		98.1	12	11								
12	97.4		97.8		98.4	8	10								
13	97.4		97.8		98.7	12	10								
14	97.4		97.9		99	6	14								
15	97.5		97.9		99.3	5	6								
16	97.5		97.9		99.6	3	2								
17	97.6		98		99.9	0	0								
18	97.6		98		100.2	0	2								
19	97.6		98		100.5	0	0								
20	97.7		98		100.8	0	0								
21	97.8		98		101.1	0	1								
22	97.8		98.1		101.4										
23	97.8		98.2												
24	97.8		98.2		**male**	**female**									
25	97.9		98.2		mean	mean									
26	97.9		98.2		98.10	98.39									
27	98		98.2												
28	98		98.2												
29	98		98.3												
30	98		98.3		stdev	stdev									
31	98		98.3		0.70	0.74									
32	98		98.4												
33	98.1		98.4												
34	98.1		98.4		st error	st error									
35	98.2		98.4		0.09	0.09									
36	98.2		98.4												
37	98.2		98.5		T test										
38	98.2		98.6		p value =	0.023932									
39	98.3		98.6												
40	98.3		98.6		plus or minus 1 SD limit										
41	98.4		98.6		98.80	99.14									
42	98.4		98.7		97.41	97.65									
43	98.4		98.7												
44	98.4		98.7		plus or minus 1 SE limit										
45	98.5		98.7		98.19	98.49									
46	98.5		98.7		98.02	98.49									
47	98.6		98.7												
48	98.6		98.8												
49	98.6		98.8												
50	98.6		98.8												
51	98.6		98.8												
52	98.6		98.8												
53	98.7		98.8												
54	98.7		98.8												
55	98.8		98.9												
56	98.8		99												
57	98.8		99												
58	98.9		99.1												
59	99		99.1												
60	99		99.2												
61	99		99.2												
62	99.1		99.3												
63	99.2		99.4												
64	99.3		99.9												
65	99.4		100												
66	99.5		100.8												
67															





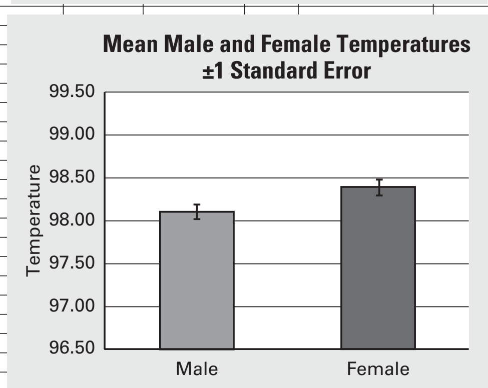


Figure 15. Sample Spreadsheet of Two-Sample T-Test for Human Body Temperature

APPLICATIONS OF CHI-SQUARE TEST RESULTS

The **Chi-square test** is a statistical method that makes a comparison between the data collected in an experiment versus the data an investigator expected to find. The Chi-square test is a way to evaluate the variability that is always present in the real world to get an idea if the difference between real and expected results is due to random chance or if some other factor is involved. For instance, if you toss a coin 10 times, although your expected results are 5 heads and 5 tails, you will often get a different result, without anything being unusual with your coins or the way you tossed them.

The Chi-square test is commonly used in introductory biology classes to test how well the results of genetic crosses fit predicted outcomes based on Mendel's laws of inheritance or to see how well measured gene frequencies in a population match up to Hardy-Weinberg predictions. When the Chi-square test is applied in these kinds of analyses, the goal is to determine whether or not the variation in the results from the expected values is due to chance. In these analyses, students are trying to confirm a theoretical expectation about their data, and they hope to quantify the contribution due to chance events. Here they hope to fail to reject the null hypothesis, i.e., that there is no evidence of a significant difference between the expected and observed results.

In other investigations, however, students may ask a question that requires a different application of the Chi-square test. For example, in a pill bug environmental choice experiment, students may wish to know if pill bugs actually choose one environment over another, or whether they just randomly move about. With this type of investigation, students are trying to discover and verify that an actual pattern exists as opposed to the random variation that often characterizes natural systems. Here they hope to reject the null hypothesis, indicating that their observed results are significantly different from the ones they expected.

The Chi-square test is also often used in medical research studies. When a scientist is testing a new drug, the experiment may be designed so that a control group receives a placebo and an experimental group receives a new drug. Analysis of the data focuses on measured differences between the two groups. The expected values would be that the same numbers of people get better in both the control and experimental groups, which would mean that the drug has no significant effect. If the Chi-square test yields a p-value greater than 0.05, then the scientist would fail to reject the null hypothesis, which would mean that there is not enough evidence that the drug has a significant effect and that any difference between the expected and the observed data is most likely due to random chance alone. If, however, the Chi-square test yields a p-value ≤ 0.05, then the scientist would reject the null hypothesis, which would mean that there is evidence that the drug has a significant effect. The differences between the expected and the observed data are probably not due to random chance alone and can be assumed to have come from the drug treatment. (This is a good opportunity for a class discussion of the complexities and statistical effect of placebo studies.)

Chi-square is calculated based on the following formula:

$$X^2 = \Sigma \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Let's see how it works.

Example 3: Mendel's Monohybrid Crosses

Mendel's laws of inheritance predict that the F_2 offspring in monohybrid crosses will express a 3:1 ratio of dominant to recessive phenotypes (assuming autosomal simple dominance). In one of Mendel's experiments, he crossed F_1 plants that were heterozygous for the seed shape trait. This cross produced 5,474 round seeds and 1,850 wrinkled seeds in the F_2 generation. This result varies slightly from the expected ratio of 5,493 round and 1,831 wrinkled peas. Is the difference between the observed and expected values significant? Is this amount of difference something that is likely to be the result of chance events, or is something else going on? The analysis is shown in Table 8.

Completing the Chi-square Calculation Table

Table 8. A Chi-square Calculation Table for One of Mendel's Peas Crossings Round versus Wrinkled, F_2 Generation

Tested Variables	Observed	Expected	Observed – Expected	(Observed – Expected)^2	(Obs - Exp)^2/ Expected
Round	5,474	5,493	–19	361	0.07
Wrinkled	1,850	1,831	19	361	0.20
				X^2 Total	0.27
				Degrees of Freedom =	1

Interpreting the Chi-square Value

With the Chi-square calculation table completed, look up the Chi-square value of 0.27 in Table 9. To know which column and row to use, you must determine the degrees of freedom to be used and the acceptable probability that the Chi-square obtained is caused by chance alone.

Table 9. Chi-square Distribution Table

Degrees of Freedom	Probability (p) Value									
	ACCEPT NULL HYPOTHESIS								REJECT	
	0.99	0.95	0.80	0.70	0.50	0.30	0.20	0.10	**0.05**	0.01
1	0.001	0.004	0.06	0.15	0.46	1.07	1.64	2.71	**3.84**	6.64
2	0.02	0.10	0.45	0.71	1.30	2.41	3.22	4.60	**5.99**	9.21
3	0.12	0.35	1.00	1.42	2.37	3.67	4.64	6.25	**7.82**	11.34
4	0.30	0.71	1.65	2.20	3.36	4.88	5.99	7.78	**9.49**	13.28
5	0.55	1.14	2.34	3.00	4.35	6.06	7.29	9.24	**11.07**	15.09
6	0.87	1.64	3.07	3.38	5.35	7.23	8.56	10.65	**12.59**	16.81
7	1.24	2.17	3.84	4.67	6.35	8.38	9.80	12.02	**14.07**	18.48

Degrees of Freedom

Which row in the Chi-square distribution table should be used? The rows in the Chi-square distribution table refer to degrees of freedom, which are calculated as *one less* than the number of possible results in the experiment. For example, a pill bug in a choice chamber experiment may have two choices (wet vs. dry, cold vs. warm, etc.); thus, the degrees of freedom would be 1 (2 – 1 = 1). For this specific cross there are only two possible results—round or wrinkled. Therefore, the degrees of freedom are 1. In Table 9 that is row 1.

Probability = p

Which column in the Chi-square distribution table should be used? Move along row 1 until you can find the Chi-square value of 0.27. It is somewhere between the 0.70 column and the 0.50 column. Remember, the probability of whether the results of an investigation differ from the null results by chance alone is called the *p-value*. A p-value of 0.05 means that there is a 5% chance that the difference between the observed and the expected data is a random difference and a 95% chance that the difference is real and repeatable—in other words, a *significant difference*. Therefore, if an investigator's p-value is greater than 0.05, he or she would fail to reject the null hypothesis—that the difference between the observed results and the expected results is due to random chance and is not significant. For Mendel's data, we fail to reject the null hypothesis: There is not enough evidence to indicate a difference between his observed results and his expected results, which suggests his observed results are expected.

Example 4: Snail Movement

Timothy Stewart (Stewart, 2007) and his classes conducted a mark and recapture investigation with snails in a stream. This investigation can lead to similar ones in your classes. One of the questions the students asked was, *Do snails tend to move upstream or downstream after their initial capture?* If the snails had no preference, then the expectation is that half of all the snails would move upstream, and half would move downstream. To answer this question, your students would want to get an estimate of how confident they are that the differences they observe are due to snail preference and not to chance. If snails do have a preference, then the students would have to reject the null hypothesis. Stewart and his classes conducted this experiment in two streams—one with a sandy streambed and one with a rocky streambed. In the sandy stream, 43 of 50 snails were recovered upstream from their release point. In the rocky stream, 22 out of 38 snails were recovered upstream from their release point. Remember, the students are investigating whether snails tend to move upstream or downstream. The students recorded their data in Tables 10a and 10b.

Table 10a. Sandy Streambed

Tested Variables	Observed	Expected	Observed – Expected	(Observed – Expected)^2	(Obs - Exp)^2/ Expected
Upstream	43	25	18	324	12.96
Downstream	7	25	–18	324	12.96
				X^2 Total	25.92
				Degrees of Freedom =	1

Table 10b. Rocky Streambed

Tested Variables	Observed	Expected	Observed – Expected	(Observed – Expected)^2	(Obs - Exp)^2/ Expected
Upstream	22	19	3	9	0.47
Downstream	16	19	–3	9	0.47
				X^2 Total	0.94
				Degrees of Freedom =	1

For the sandy streambed, the null hypothesis (that there is no difference between upstream and downstream movement) is rejected with a high degree of confidence (25.92 is well beyond the numbers on the first row of Table 9). There is an obvious and significant tendency of the snails to move upstream. What about the rocky bottom stream? In the first row of Table 9, 0.94 lies between 0.50 and 0.30. The students should fail to reject the null hypothesis for the rocky stream; there is not enough evidence of an apparent preference for upstream or downstream movement of snails.

Should the students just stop there? There is such a clear pattern in the sandy stream, so why shouldn't they expect the same in the rocky stream? What hypotheses could they propose that might explain these differences? How would they investigate these hypotheses? Is the stream velocity significantly different between the two streams? What could affect the outcome that was obtained? The statistical analyses we use to answer such questions are very robust, and we expect that their use will lead to other questions; but they still need to be interpreted within the context of the experimental design. Chi-square is a useful tool that your classes can employ in some of their own investigations.

Looking for Trends or Associations

The body temperature and heart rate data can be extended to investigate how two measured variables affect each other. For example, *Is heart rate in humans related to body temperature? Does one increase or decrease as the other increases or decreases?* In the same data we've been using, there is another column of numbers—heart rate. First, construct a scatterplot of the data, plotting one variable against the other. Figure 16 shows the scatterplot of heart rate and body temperature. Ask students which variable should go on the *x*-axis and which on the *y*-axis.

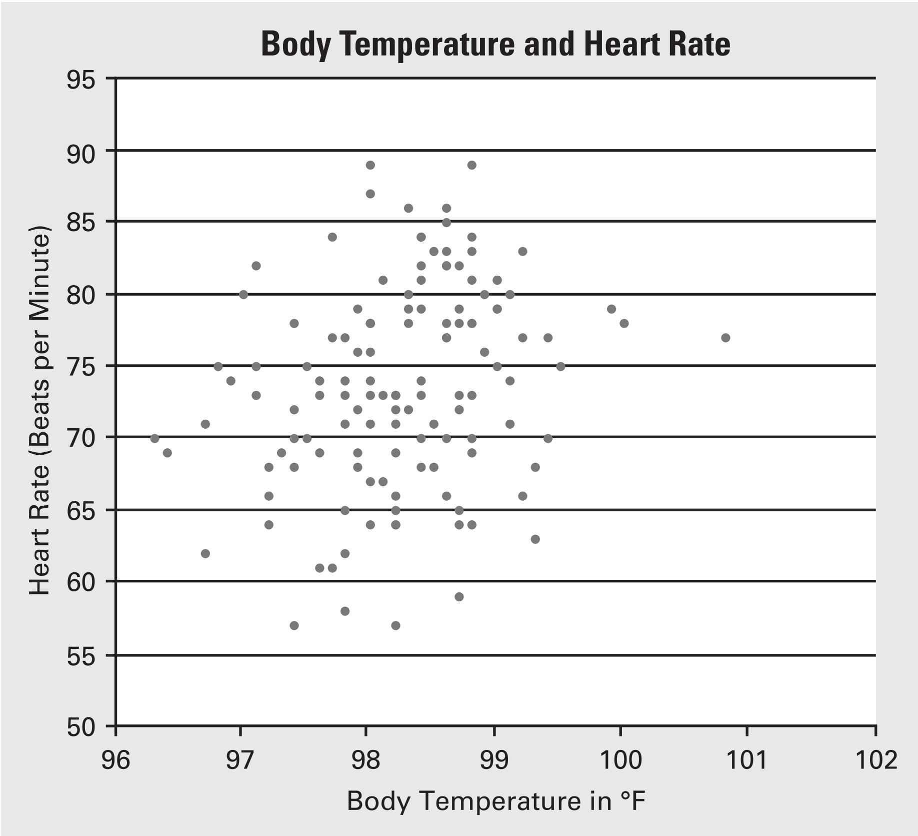


Figure 16. Scatterplot of Heart Rate and Body Temperature

At first glance, the points in Figure 16 seem to be scattered randomly on the graph. However, take a closer look. Is there a general trend? Does heart rate increase as body temperature increases? If so, what sort of relationship might exist between these variables? By first assuming that this is a linear relationship, we can use the spreadsheet to identify a trend line and perform a regression analysis to help students make a better estimate of the relationship between the two variables. The trend line graphs a best fit line for these points—it summarizes all the points into a single linear equation.
The regression analysis provides a measure of how the two variables are related to each other. Notice that with the trend line drawn in Figure 16, there appears to be a postitive correlation—that is, when body temperature goes up, so does the heart rate, and vice versa. This trend line was created by calculating how far each point is from the mean *x* value and from the mean *y* value. The *r*-value provides an estimate of the degree of correlation. An *r*-value of 1 means a positive 1:1 correlation. This means that, in this example, with every degree that the body temperature increases, there is a regular increase in the heart rate of the individual. An *r*-value of −1 is a negative correlation. In this case, it would mean that as body temperature increases, heart rate decreases. The *r*-value calculated here as 0.25 equals an R^2 value of 0.064. This value suggests that only about 6.4% of the variation in heart rate is attributed to the variation in body temperature—this is

not a strong correlation. The *r*-value does not imply a causal relationship—though it is tempting to think that it does. Rather, the *r*-value provides an explanation or description of the association between two variables. To explore causality we would need to design an experiment that manipulates the variables in a controlled fashion to find evidence to support causality. Finding a significant *r*-value, however, is often a first step in designing such an experiment.

Does this *r*-value indicate a significant association or correlation? (Here, if you are using Excel, there's a bit of a problem—Excel does not calculate a p-value for *r*-values. You will have to go to a website for a p-value calulator or use a table.) The degrees of freedom in this case are N–2 or 128, and you will need to decide whether to use a two-tail or one-tail test. The "tails" refer to the extremes on a normal distribution. In this case, since you are asking if there is a positive relationship between two variables, use a two-tail test. With these parameters, the p-value equals approximately 0.04, which would suggest that students can estimate with about 96% confidence that heart rate and body temperature are correlated, although not strongly so.

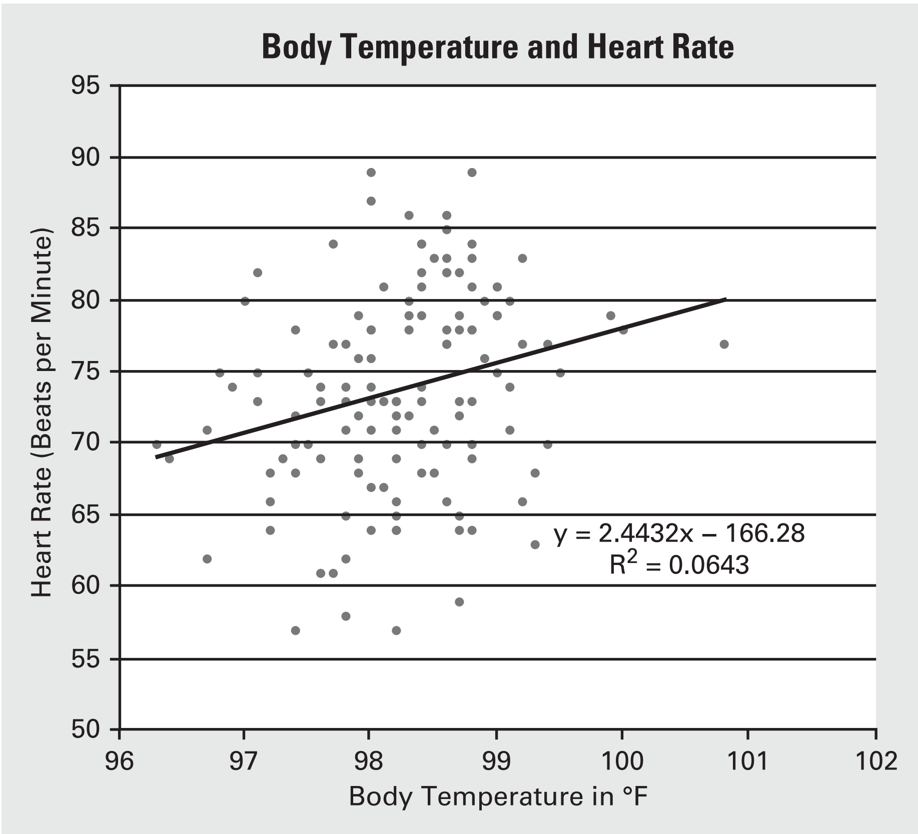


Figure 17. Scatterplot of Heart Rate and Body Temperature Showing the Trend Line

ALIGNMENT TO THE INVESTIGATIONS

Sample Alignment to Investigation 12: Fruit Fly Behavior

What environmental factors trigger a fruit fly response?

Using *Drosophila melanogaster* as a model organism, students investigate behaviors that underlie taxis by designing and conducting experiments to gather information about responses to chemical and other environmental stimuli. After noting patterns and ratios, students can apply statistical methods to test their hypothesis about the relationship between an environmental variable and an observed response.

Alignment to the Curriculum Framework

After making an observation, the student proposes a hypothesis to explain a casual relationship between an underlying variable and the observed phenomenon. However, because absolute proof is not possible, statistical hypothesis testing focuses on trying to reject a null hypothesis, i.e., that the underlying variable is independent of the observed phenomenon—that there is no causal relationship. Through investigation, the student tries to find an alternative to the null hypothesis and, based on statistical tests, the evidence that supports an alternative explanation.

Investigation 12: Fruit Fly Behavior provides an opportunity for students to test a hypothesis about the relationship between an environmental variable and a response, such as a chemical stimulus and chemotaxis. Students apply statistical methods such as Chi-square analysis to determine whether the experimental data differs significantly from the null results.

Science Practices Applied

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

SP 6 The student can work with scientific explanations and theories.

Learning Objectives Addressed

LO 2.24 The student is able to analyze data to identify possible patterns and relationships between a biotic or abiotic factor and a biological system (cells, organisms, populations, communities, or ecosystems) (2.D.1 & SP 5.1).

LO 2.38 The student is able to analyze data to support the claim that responses to information and communication of information affect natural selection (2.E.3 & SP 5.1).

LO 2.39 The student is able to justify scientific claims, using evidence, to describe how timing and coordination of behavioral events in organisms are regulated by several mechanisms (2.E.3 & SP 6.1).

LO 4.14 The student is able to apply mathematical routines to quantities that describe interactions among living systems and their environment, which result in the movement of matter and energy (4.A.6 & SP 2.2).

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance (4.B.3 & SP 5.2).

Example of Sample Data Analysis

This investigation allows students to discover a pattern in fruit fly behavior based on observations that they make and data that they collect using the choice chambers. After making a working hypothesis, students should design an experiment to investigate the relationship between an environmental variable(s) and fruit fly behavior. Conclusions must be supported by evidence; students should present data from a repeated, controlled experiment(s) with graphic representation and quantitative analysis of fruit fly choices. One suggestion is for students to design a data grid reflecting choices for the flies, graph the data, and draw conclusions. Chi-square analysis of the results can determine if the observed distribution of flies in the choice chamber differs significantly from expected ratios. (Students are expected to be able to manipulate the Chi-square formula and interpret the probability table.)

ALIGNMENT TO THE EXAM

Hypothesis Testing and Exam Questions

A hypothesis is a statement explaining that a casual relationship exists between an underlying factor (variable) and an observable phenomenon. Once a hypothesis is made, students can investigate the relationship through experimentation and data analysis. Because absolute proof is not possible, hypothesis testing often uses statistical analysis to reject a null hypothesis. The following sample exam question from the *AP Biology Practice Exam* represents an item that assesses students' ability to quantitatively analyze experimental data and then form a hypothesis that can lead to further investigation, with results subject to testing by using statistical tools. The question is based on a learning objective from the curriculum framework.

Sample Multiple-Choice Question Subject to Hypothesis Testing

A student placed 20 tobacco seeds of the same species on moist paper towels in each of two petri dishes. Dish A was wrapped completely in an opaque cover to exclude all light. Dish B was not wrapped. The dishes were placed equidistant from a light source set to a cycle of 14 hours of light and 10 hours of dark. All other conditions were the same for both dishes. The dishes were examined after 7 days and the opaque cover was permanently removed from dish A. Both dishes were returned to the light and examined again at 14 days. The following data were obtained.

	Dish A		Dish B	
	Day 7 Covered	Day 14 Uncovered	Day 7 Uncovered	Day 14 Uncovered
Germinated seeds	12	20	20	20
Greeb-leaved seedlings	0	14	15	15
Yellow-leaved seedlings	12	6	5	5
Mean stem length below first set of leaves	8 mm	9 mm	3 mm	3 mm

Which of the following best supports the hypothesis that the difference in leaf color is genetically controlled?

(A) The number of yellow-leaved seedlings in dish A on day 7

(B) The number of germinated seeds in dish A on days 7 and 14

(C) The death of all the yellow-leaved seedlings

(D) The existence of yellow-leaved seedlings as well as green-leaved ones on day 14 in dish B

RESOURCES

The following resources include ideas for classroom activities and lessons, as well as examples for learning and teaching as you further your exploration into the new AP Biology curriculum. Each resource includes the approximate time involved, a description of the resource and where to find it, the quantitative skills the activity or lesson involves, and where those skills might be applied.

DEVELOPMENTAL PLASTICITY IN OAK LEAVES DAVID WESTMORELAND, 16TH WORKSHOP/CONFERENCE OF THE ASSOCIATION FOR BIOLOGY LABORATORY EDUCATION (ABLE) http://www.ableweb.org/volumes/vol-16/8-westmoreland.pdf	APPROXIMATE TIME 1 class period collecting and teaching methods; 1 class period for analysis
DESCRIPTION OF RESOURCE	
Students collect leaves from an oak tree on a sunny side and a shady side. They determine the surface area of the leaves and sinus cavities. They then run box-and-whisker plots and T-tests to analyze their data and test their hypotheses. These tests will measure the statistical significance of any differences in the measured variable from leaves in shade versus sun.	
QUANTITATIVE SKILLS APPLIED	
SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems. SP 2 The student can use mathematics appropriately. SP 3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course. SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question. SP 5 The student can perform data analysis and evaluation of evidence.	
LEARNING OBJECTIVES ADDRESSED	
LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2] LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]	

HANDBOOK OF BIOLOGICAL STATISTICS http://udel.edu/~mcdonald/statintro.html	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

The Handbook of Biological Statistics gives an excellent overview of the use of statistical tests of biological hypotheses for both you and your students. It begins with a description of the null hypothesis and the difference between the statistical null and biological null hypotheses and then describes how to test it. Probability and significance values are discussed, as well as deeper statistical analysis descriptions.

QUANTITATIVE SKILLS APPLIED

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.2 The student is able to evaluate evidence provided by data to qualitatively and quantitatively investigate the role of natural selection in evolution. [See SP 2.2, 5.3]

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.11 The student is able to design a plan to answer scientific questions regarding how organisms have changed over time using information from morphology, biochemistry and geology. [See SP 4.2]

LO 1.21 The student is able to design a plan for collecting data to investigate the scientific claim that speciation and extinction have occurred throughout Earth's history. [See SP 4.2]

LO 2.23 The student is able to design a plan for collecting data to show that all biological systems (cells, organisms, populations, communities and ecosystems) are affected by complex biotic and abiotic interactions. [See SP 4.2, 7.2]

LO 2.35 The student is able to design a plan for collecting data to support the scientific claim that the timing and coordination of physiological events involve regulation. [See SP 4.2]

LO 3.14 The student is able to apply mathematical routines to determine Mendelian patterns of inheritance provided by data sets. [See SP 2.2]

COLBY COLLEGE BIOLOGY DEPARTMENT http://web.colby.edu/biology/statistics-and-scientific-writing/simple-statistics-and-hypothesis-testing/	APPROXIMATE TIME More than 1 class day depending on the number of tutorials assigned

DESCRIPTION OF RESOURCE

The Colby College Biology Department home site has eight tutorials on statistics and scientific writing. This link focuses on simple statistics and hypothesis testing. The home site includes tutorials on the scientific method, scientific writing, preparing tables and figures, frequency analysis (Chi-square analysis), T-tests, and regression analysis. The focus of the statistics section is the use of statistics to help understand observations. The site uses a simple experiment involving 20 voles and explains how the results can be interpreted, distinguishing between statistics and probability. A second example compares mean and variation for blueberry production with two different sun exposures. The site discusses the significance of the means being equal but variations not being similar. This is connected to the statistic of standard deviation. The T-test chapter uses a small sample of humans to infer information about a large population concerning the heights of men compared with women.

QUANTITATIVE SKILLS APPLIED

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.2 The student is able to evaluate evidence provided by data to qualitatively and quantitatively investigate the role of natural selection in evolution. [See SP 2.2, 5.3]

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time. [See SP 5.3]

LO 1.11 The student is able to design a plan to answer scientific questions regarding how organisms have changed over time using information from morphology, biochemistry and geology. [See SP 4.2]

LO 1.21 The student is able to design a plan for collecting data to investigate the scientific claim that speciation and extinction have occurred throughout Earth's history. [See SP 4.2]

LO 2.23 The student is able to design a plan for collecting data to show that all biological systems (cells, organisms, populations, communities and ecosystems) are affected by complex biotic and abiotic interactions. [See SP 4.2, 7.2]

LO 2.35 The student is able to design a plan for collecting data to support the scientific claim that the timing and coordination of physiological events involve regulation. [See SP 4.2]

BIOQUEST CURRICULUM CONSORTIUM http://bioquest.org/esteem/esteem_result.php	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

The BioQUEST Curriculum Consortium's website opens, "We hope to engage you in a process that is likely to lead to the development of high quality, adoptable, and adaptable curricular materials in biology and mathematics education." The link provides 30 complete modules in various branches of biology, called ESTEEM modules, for Excel Simulations and Tools for Exploratory, Experimental Mathematics. The module on biostatistics is a collection of workbooks with tools for (1) linear regression through the origin, (2) linear regression not through the origin, (3) polynomial fit, and (4) a simple Chi-square analysis to test whether a dataset is in Hardy-Weinberg equilibrium (users may enter their own dataset or use one of the included sets).

QUANTITATIVE SKILLS APPLIED

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations. [See SP 1.4, 2.1]

NUMBERS COUNT PROJECT OF HHMI http://bioquest.org/numberscount/resources/	APPROXIMATE TIME N/A
DESCRIPTION OF RESOURCE	
The Numbers Count Project provides data, tools, and curricular materials for use in the classroom to support student observation, experimentation, and mathematical modeling. Resources include an activity on the use of spatial fern population data in hypothesis testing, an overview of the relationships between statistical concepts and approaches, an Excel module that provides worksheet data and a spreadsheet for investigating the Hardy-Weinberg equilibrium, and a mini-Excel tutorial on calculating variance.	
QUANTITATIVE SKILLS APPLIED	
SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems. SP 2 The student can use mathematics appropriately. SP 3 The student can engage in scientific questioning to extend thinking or to guide investigations within the context of the AP course. SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question. SP 5 The student can perform data analysis and evaluation of evidence. SP 6 The student can work with scientific explanations and theories.	
LEARNING OBJECTIVES ADDRESSED	
LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations. [See SP 1.4, 2.1] LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]	

CHAPTER 4:
Mathematical Modeling

LABS USING MATHEMATICAL MODELING

1: Artificial Selection

2: Mathematical Modeling: Hardy-Weinberg

3: Comparing DNA Sequences to Understand Evolutionary Relationships with BLAST

9: Biotechnology: Restriction Enzyme Analysis of DNA

10: Energy Dynamics

11: Transpiration

Living systems are incredibly complex, and biologists ask questions that can penetrate, explore, and attempt to explain this complexity. Some biological questions are best answered with field or laboratory experiments that involve data analysis and hypothesis testing. Often, though, initial understanding comes from the development of models, or simplifications and approximations of reality that include conceptual, mathematical, and computer models.

Mathematical modeling is the process of creating mathematical or computer-based representations of the structure and interactions of complex biological systems. It is used to build an understanding of the underlying causal factors of a biological phenomenon, to drive prediction, to bring theory and experiment together, and to integrate experimental design and data analysis. Experiments inform models, which in turn inform a series of experiments or the collection of field observations. Mathematical modeling deals with all sorts of questions: *What can we measure? What should we measure? What are the relevant variables? What are the simplest informative models that we can build?*

Some of the main aspects of developing a model are deciding on its assumptions and simplifications, approximating but simplifying real-world conditions, and introducing the limitations of models, while simultaneously introducing the power of models. It doesn't take long to explain or to model using a spreadsheet.

The ability to constantly tweak models and to try out ideas using those models is one of the powerful aspects of working with models. Remember, models, by definition, are simplifications of nature; likewise, by definition, models are very limited. But even with all their limitations, models can still provide students with valuable insights into seemingly impossibly complex problems with a little work and perseverance on the students' part. These insights can further inform their experimental designs when they decide to move from model to investigation.

This chapter introduces some preliminary mathematical modeling skills and ways of thinking about living systems to encourage the use of mathematical models in biological investigations. Some labs in *AP Biology Investigative Labs: An Inquiry-Based Approach* focus specifically on mathematical modeling. Mathematical modeling can inform or actually be the focus of other independent investigations.

COMPONENTS OF MATHEMATICAL MODELING

The mathematical modeling process commonly begins by identifying variables that might account for much of the variation observed in a perturbed biological system. The following series of steps leads to the development of a mathematical model:

1. Examine a biological system to identify which variables seem to affect the system most when they are changed.
2. Develop graphical or physical models (games) to capture the essence of a biological phenomenon.
3. Translate the graphical or physical models into "word equations."
4. Translate word equations into formal equations.
5. Implement the model in a computer environment (program language, spreadsheet, or programming environment).
6. Evaluate, revise, and extend the model.

Models can be deterministic or stochastic. **Deterministic models** are calculated with fixed probabilities. The following are examples of the application of deterministic models to biological inquiry:

http://pubs.acs.org/doi/abs/10.1021/ed086p651

http://www.uvm.edu/rsenr/vtcfwru/spreadsheets/ecologyevolution/EE_Exercises/Exercise10/10%20Donovan%20pages%20112%20%20EE.pdf

http://www.ncssm.edu/courses/math/apcalcprojects/fluactivity

Stochastic models use a random number generator to create a model that has a variable outcome and contingencies, much as normal biological systems do. With proper use of the Random function in a spreadsheet, stochastic models of biological phenomena can be created and explored, and data from a number of simulations can be collected and analyzed. The following is an example of random number generation applied to genetic cross outcomes:

http://academic.pgcc.edu/~ssinex/excelets/genetic_odds.xls

(Stochastic models are the kind of model featured in Investigation 2: Mathematical Modeling: Hardy-Weinberg.)

USING MATHEMATICAL MODELING

What should students consider when developing a mathematical model? Some of the first steps include deciding on assumptions and simplifications, approximating but simplifying real-world conditions, and introducing limitations of the models. Because models by definition are simplifications or approximations of reality, so are their results. Students must remember that models in biology are limited in their accuracy or precision, but they prove their value by providing insights into the ways in which various parameters can affect a model or a system. In other words, students should understand that a model is false even before they begin working with it, because a model is a consciously produced, simplified version of some natural phenomenon.

The focus, then, is on whether the students can predict and explain variation in their data with a simple causal mechanism that can be modeled. The process of building a mathematical model, testing it, and exploring how it responds to various inputs helps to inform the students' understanding of complex biological processes.

There is really no learning tool for mathematical modeling other than working through the process. Example 1 at the end of this chapter will explain how students can build mathematical models in a spreadsheet using simple, discrete time steps, difference equations, and iteration. Spreadsheets offer several advantages as a tool for introducing mathematical modeling. They are ubiquitous and familiar to you and your students. They offer a fast and easy way to construct graphs. Finally, spreadsheets are particularly well suited for discrete, iterative model construction.

You may want to create your own spreadsheet model. You will be better able to teach the modeling process if you create the spreadsheets from scratch yourself. As with any lab procedure, you need experience in it to teach it.

Most students will have had some sort of spreadsheet introduction, but many have not had to create their own spreadsheets from scratch. It is important that each student have a chance to make missteps while constructing this spreadsheet. Valuable learning takes place when students recognize and correct their errors. Most often, the first student-constructed spreadsheets will have some fundamental error. Use probing questions to help the students recognize these errors, but generally let the students propose their own solutions. Try to achieve a mindset in the students whereby they propose possible solutions, enter them in the spreadsheet, and then use the results from the spreadsheet calculations to evaluate their original proposed solution—in essence, scientific thinking. Students will struggle a bit but will still succeed. This success is key to building the tenacity needed to solve problems and gain essential skill sets.

Here are some of the common problems. Remember to ask questions to help the students see these problems:

1. Students generally just fill in the numbers from the board and fail to create formulas in the cells.
2. Students copy cells incorrectly.
3. Later, students fail to recognize and use the differences between fixed references versus relative cell references. (This is a valuable skill necessary to build accurate spreadsheet models. Look up "fixed references" and "relative references" in the spreadsheet help menu.)
4. Students have a difficult time evaluating their graphs. Students undertaking mathematical modeling must understand that every step of the model building should be checked and validated.

EXAMPLES OF MATHEMATICAL MODELING

Example 1: Simple Model of Population Growth

In mathematical modeling, it is often productive to start on a simple level to capture the fundamentals of a problem, then expand it to develop a more powerful model. As a first exercise in mathematical modeling, have students consider a very simple model of population growth. Later, you will ask them to develop a more sophisticated model.

A popular high school biology text from the 1960s (*Biological Science Curriculum Studies,* Green version) introduced mathematical modeling with an imaginary population of sparrows. Students may be surprised and amused to learn that this exercise was developed before the advent of desktop computers or even hand-held calculators, so the calculations and graphing were done by hand. They will learn that exploring this same model today with a spreadsheet does not take long.

The exercise introduces some of the main points of developing a model—deciding on assumptions/simplifications, approximating but simplifying real-world conditions, and introducing the limitations of models—while introducing the power of models.This exercise also provides for many students a first introduction to semi-log plots.

Begin by asking questions and making some simplifying assumptions. These assumptions are not valid in reality, but they are useful to start building a model. First, ask, *What four factors determine whether or not a population grows or declines?* Answers should include the following:

- Rate of births (growth)
- Rate of deaths (decline)
- Rate of migration out (decline)
- Rate of migration in (growth)

Explain to students that they can estimate the growth of a particular population if they know all these factors. Have them imagine the following scenario, which keeps these factors in mind and describes a particular, simplified population:

- An island with unlimited resources and no limiting factors. (This assumption immediately places the model in a hypothetical world and not the real world. Remind students that they are trying to tease out the nature of population growth, not describe a real population in detail.)
- Ten sparrows are introduced to the island: five males and five females. (Start small and use a simple hypothetical population.)
- Each year, each pair of sparrows produces 10 offspring. (Obviously, in a real population some pairs would not be successful, and others might produce more offspring. However, this assumption balances out the successful and unsuccessful breeding outcomes.)
- All offspring survive to reproduce the next year.
- The parents all die before the next year. Every sparrow reproduces and then dies. (Taken together, these two assumptions—that all offspring survive, and that parents only survive one year—tend to balance each other out.)
- No new sparrows immigrate to or emigrate from the island. (Another simplification—stating that there is no migration—eliminates two of the ways in which populations grow or decline from our model.)

Next, ask students, *How many birds will be alive on the island at the end of each year for the next 10 years?* Have them use the information that has been supplied to create a model using a spreadsheet that calculates the population of sparrows on the hypothetical island each year, over a 10-year period. Then ask them to graph the results using a scatterplot.

This exercise introduces a way of thinking about biological problems mathematically and helps to build a more intuitive sense of exponential processes that are fundamental to biology specifically and to life skills in general. There is no single way to organize the spreadsheet, but one possibility is described here and shown in Table 11.

Table 11. Possible Spreadsheet Structure for Population Growth Model

Year	Sparrows	Pairs	Offspring

First, have the students decide on labels (Table 11). Next, have them simply calculate the first two rows in their heads so that they can be sure that they are programming their spreadsheet correctly (Table 12).

Table 12. Spreadsheet with First Rows Calculated Manually

Year	Sparrows	Pairs	Offspring
0	10	5	50
1	50	25	250

With these numbers in mind, have students enter the correct formulas in the appropriate cells. Tell them to expand their model by adding more rows until they have 10 years modeled. The best and easiest way to do this is with the Copy tools in the spreadsheet.

Figures 18 and 19 show two graphs of the output from this model. Figure 18 is a standard *x-y* scatterplot. Have students determine what the shape of this graph would be if the data points were connected. This simple model of population growth is known as *exponential growth*—exponential because it is described by an exponential function in which the growth rate for each step is proportional to the population size at each step. Here, the students have plotted an exponential function without actually defining it.

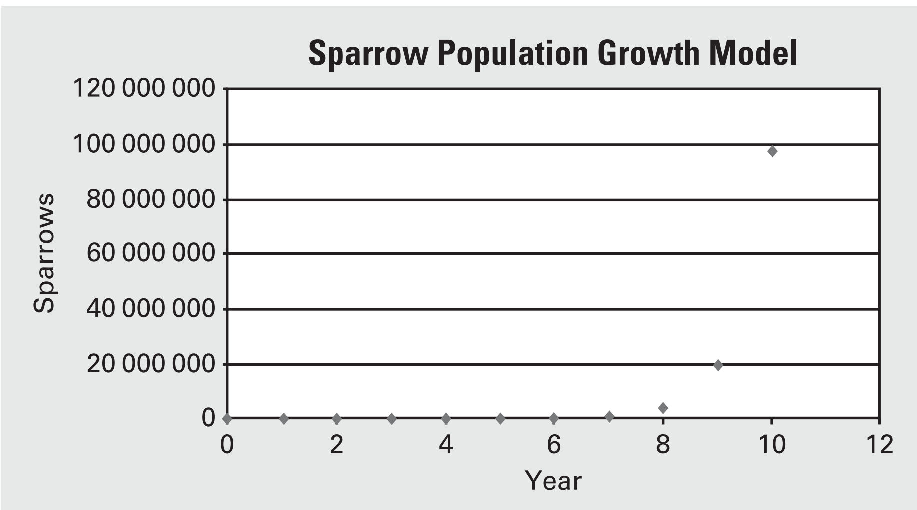


Figure 18. Standard *x– y* Scatterplot for Sparrow Population Growth Model

Students should have several questions as they explore the output of their model. Is this growth pattern plausible? Can any population continue to grow in an unlimited manner, as this model suggests? Does this model represent some fundamental capacity for growth in all populations? Does this type of population growth occur in the real world? If so, under what conditions? What is missing in the current model?

Based on the curve in Figure 18, students could predict how many sparrows there will be in this population after 20 more years, but making such a prediction is difficult with this kind of plot. It is easier with a graph of the log of the *y*-values versus *x*. In a semi-log plot shown in Figure 19, the *y*-axis is "logged"; that is, the *y* values are expressed as logs of each year's population rather than as the large population numbers in the previous graph. Note that by plotting the logs of the populations for each year in Figure 18, the plot now appears to be a straight line rather than a curve.

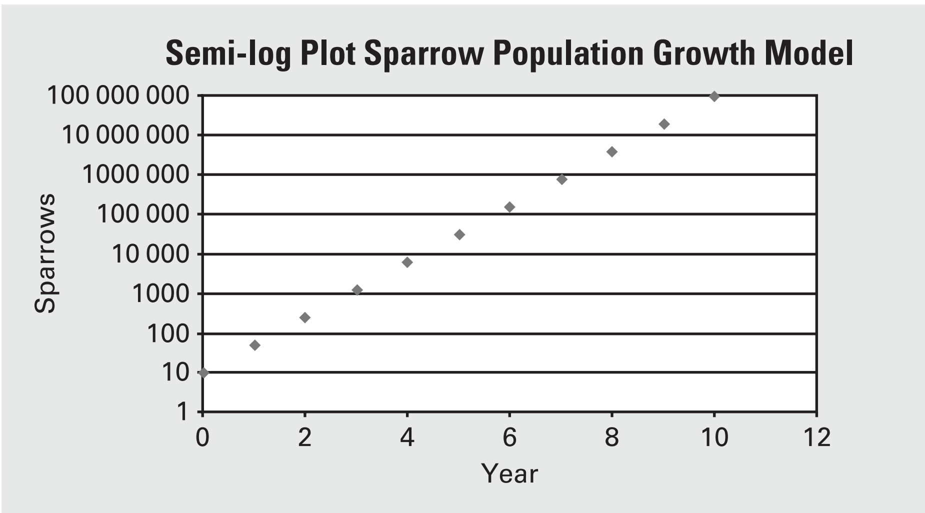


Figure 19. Semi-log Plot for Sparrow Population Growth Model

What makes the semi-log plot easier to use when predicting the population in 20 years? To have students explore this model, ask them to add new factors to the population growth. Simplifications in this model, as listed above, are that there is no migration and that the birth rate is established at 10 offspring per each pair of sparrows. Discuss how the death rate is taken into account. There is no explicit cell with a death rate parameter, but only the offspring in each row move to the next year—death is taken care of by omission. It should not be too challenging to reconstruct the model with birds migrating to or emigrating from the island or changing the rates of survival in the parents or offspring. Ask students to build a new model, and compare growth in the new model with growth in the old model. Ask them whether the overall curve changes much. Now ask them to try changing how many parents live to the next generation or how many birds are born to each pair. Ask, *Do any of these changes alter the overall outcome of the model—unlimited population growth?*

Example 2: Using Single Equations for Population Growth Models

The overall nature of the population growth modeled in the previous example essentially stays the same, so perhaps the essence of this population growth can be captured in a single equation. Doing so might make exploring the model easier, and it might open pathways to a more general model of population growth. Thus, explain to students that they are going to consolidate terms into a single equation and model that in the spreadsheet.

The four factors that directly influence population growth can be combined into one term—a percentage of increase, or *r*. A variable for the population size at any particular time interval also can be named: N_t. Now ask students to represent the exponential population growth in a single equation with the variables *r* and N_t. Rather than describe the unit of time as *years,* as in the island sparrow model, they should use a generic time unit, *interval,* (±1).

A good strategy when building mathematical models is to try to verbalize the equation first. Here's a possibility:

The new population is equal to the previous population + rate of increase × the previous population.

This verbalization can be expressed as an equation:

$$N_t = N_{(t-1)} + r \cdot N_{(t-1)}$$

Now have students enter the formula into the spreadsheet. This is a bit difficult, because the spreadsheet requires a different format than standard algebraic notation. Have students start by defining the initial value of their variables, *r* and *N*. Table 13 shows one possible structure.

Table 13. Initial Value of Population Parameters

r =	0.1
N =	10
Time Interval	Population
0	10
1	11

In time interval 0, in the population cell, have students enter a reference to the cell with the starting population. Tell them to use formulas in each of the cells in the next row—a formula to calculate the time interval and a formula to calculate the new population using the previous population and the value for *r.*

There is one spreadsheet technique that students will need to consider and understand in order to make this model. Students need to know the difference between, and how to apply, relative cell references and fixed cell references in a cell's formula. Because the reference to *r* always points to the same cell, it should not change—that requires a *fixed* reference. The default in a spreadsheet formula reference is *relative*, which changes as it is copied. A cell reference can be *fixed* by adding a dollar sign in front of both components of a cell's address. For example, referring to cell **B1** is a relative reference, but referring to cell **B1** is a fixed reference.

By iterating the formula (using the results from one time interval as the basis for the next), students can explore and create models that, without computers, would require a familiarity with calculus. Have students expand this model to 300 generations or time intervals, and graph the results. If they have correctly built this model, then a scatterplot of the population as a function of the time interval will be shaped much like the graph in Figure 20. Ask students to determine what happens when the value of *r* or *N* is changed: Does the growth curve just change, or does it get less or more steep?

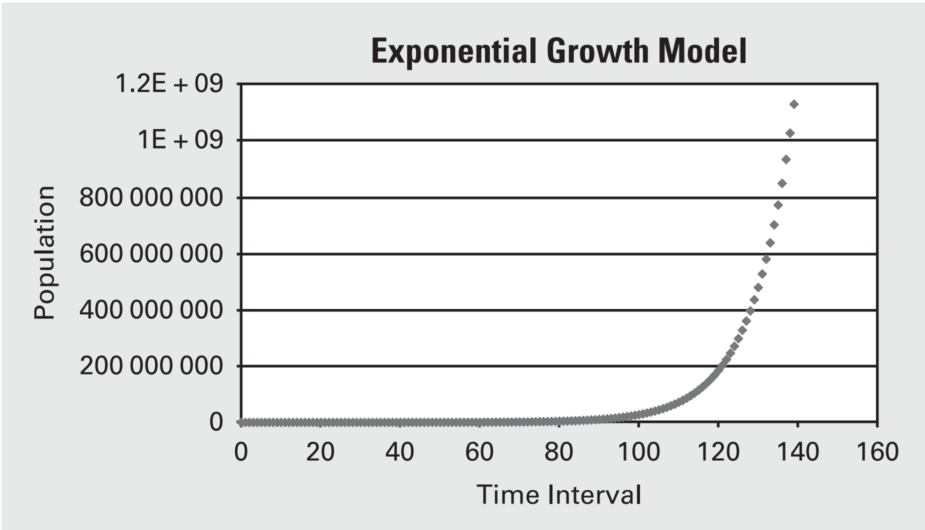


Figure 20. Exponential Growth Model

Mathematical modeling is an additive process, and this exercise so far has only begun to touch on the possibilities. Note that the procedure thus far has only added a bit of complexity at each step—with only rudimentary math operations. The next step will be more complex, but the math will not. Students will try to build a model that is a more realistic representation of how real-world populations grow. They must remember, though, that it will still be a simplification compared with the natural world.

Ask students, *In the real world, do populations continue to grow exponentially?* Point out that from year to year, the number of squirrels in the park seems to be fairly constant. The number of milkweed plants in the meadow seems to change very little. What is the explanation? Explain that the first model was based on constant rates. In real life, however, the different factors that affect population growth actually change. Ask, *What factors might limit population size?* Discuss that most populations tend to stabilize—the environment has a carrying capacity. Multiple factors contribute to the carrying capacity, but the carrying capacity of the environment can be represented as a variable, *K,* in the equation. As the population approaches the carrying capacity of the environment, the growth rate of the population declines. When the population is at carrying capacity, the rate of growth is zero—creating an S-shaped curve. Have students build a model of this growth pattern, as follows.

First, ask students to consider the variables in the earlier model: *N* and *r*. Ask, *Which of the two terms changes as the exponential equation is recalculated? Which term is constant?* Discuss what has to happen to *r* if the students want to modify the exponential growth curve into an S-shaped curve. The answer is that *r* can no longer be constant. Remember, this is the algebraic form of the first exponential model:

$$N_t = N_{(t-1)} + r \cdot N_{(t-1)}$$

Now, explain that modeling a population with limits necessitates introducing a changing *r* into the model, along with *K*, the variable for carrying capacity. Discuss what has to happen to the rate of growth in a population as the population grows toward the carrying capacity. If carrying capacity represents a limit to population size, then once a population reaches the carrying capacity, there is no growth. What about a population that is just starting to grow in a new area? Would the rate of growth, *r*, be maximized or small? Are there any limits on growth if the size of the population is not near carrying capacity? If the current spreadsheet model has a constant value of *r*, how might that value be changed during each iteration to maximize growth early on when the population is small and to minimize growth later as the population approaches *K*? All of these questions are good discussion points.

Ask whether there is some mathematical expression that could be added to the first exponential equation that maximizes *r* in early generations but minimizes *r* in later generations. Have students try to think of an expression that includes just the *N* variable and the *K* variable that can be multiplied times the *r* term to fill the needs of the model. In other words, ask them to think of an expression that is approximately equal to 1 when *N* (the population size) is small but is approximately equal to 0 when *N* approaches *K* in size—an expression that generates a graph like Figure 21.

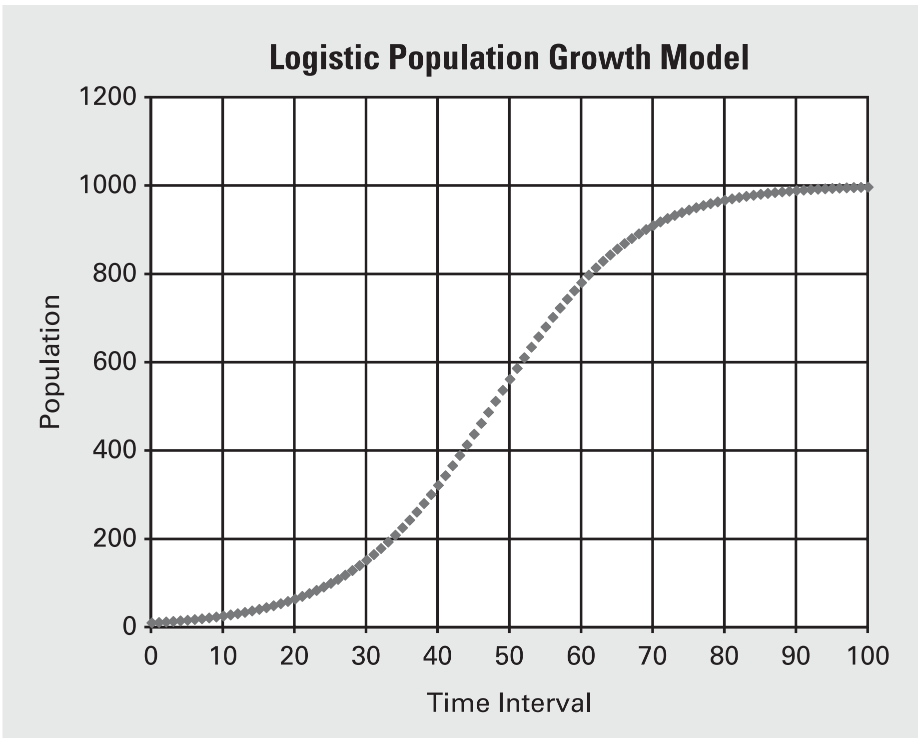


Figure 21. A Discrete Logistic (S-shaped) Growth Model

Figure 21 is a logistic curve, and the expression students are looking for is called the *logistic*. Have them try out different expressions they think will work in their spreadsheets. To evaluate their proposed expressions, have students put each one in the spreadsheet and use the graph produced to evaluate whether the expression works as planned.

One benefit of spreadsheets is that they can quickly provide feedback as to whether or not formulas have been entered correctly or even if a proposed formula works in the predicted way. In other words, making mistakes and fixing them is a critical part of these exercises. Don't let students bypass a learning opportunity by taking the shortcut of looking up the completed spreadsheet on the Web. This is different from the usual way they might try to discover a formula. Here they can just try it out. The model is a development environment. The model informs the user as to the correctness of a proposed expression.

Once students have correctly constructed a model, the next step is for them to test different values for the various population parameters they have modeled. Ask, *Does changing* K *change the shape of the graph? If so, in what way? How does changing* r *change the graph?*

Real-world populations tend to have small, measured *r*-values (less than 10%). Using a spreadsheet to model population growth with this strategy creates what is known as a *discrete model*, a model in which the time steps are small, discrete intervals. A similar model that had been developed with differential equations would be a *continuous model.* The behaviors of these two models are similar, but there are important differences. Students can explore some of those differences by trying unrealistic growth rates in their logistic model. Have them try out values greater than 1, 2, or even 3 (the equivalent of 100%, 200%, or 300% growth per time interval).

Finally, ask students how this model might be applied to study laboratory or field populations. You can find links to examples of these spreadsheets (already completed) and an extension exercise at http://www.nabt.org/blog/2009/06/26/an-extension-to-the-logistic-model/.

ALIGNMENT TO THE INVESTIGATIONS

Sample Alignment to Investigation 2: Mathematical Modeling: Hardy-Weinberg

How can mathematical models be used to investigate the relationship between allele frequencies in populations of organism and evolutionary change?

Most lab situations in which students try to manipulate an evolving population are flawed because the sample population size is small and subject to genetic drift. However, the complexity of evolution in a population can be illuminated by mathematical equations, several of which are based on Hardy-Weinberg equilibrium. In this investigation, students build a spreadsheet that models how a hypothetical gene pool changes from one generation to the next. This model allows for the exploration of parameters that affect allele frequencies, such as selection, mutation, and migration.

Alignment to Curriculum Framework

Math and computers provide tools to explore the complexity of biological systems. Although some biological questions are best answered through experiments that involve data analysis and hypothesis testing, often initial understanding of complex living systems comes from the development of models (conceptual, mathematical, and computer), or simplifications and approximations of real-world conditions. Deterministic models are calculated with fixed probabilities, whereas stochastic models use a random number generator with variable outcomes and contingencies, much like normal biological systems. Using the random function in a spreadsheet, students can create stochastic models of biological phenomena, and data from various simulations can be collected and analyzed.

Stochastic models are featured in Investigation 2: Mathematical Modeling: Hardy-Weinberg. In this investigation, students develop a spreadsheet-based Hardy-Weinberg genetics model. Throughout the process, students make claims and evaluate evidence about parameters that influence population genetics, thus engaging in both reasoning and rebuttal.

Quantitative Skills Applied

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 The student can use mathematics appropriately.

SP 4 The student can plan and implement data collection strategies appropriate to a particular scientific question.

SP 5 The student can perform data analysis and evaluation of evidence.

Learning Objectives Addressed

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change (1.A.1 & 1.5, 2.2).

LO 1.2 The student is able to evaluate evidence provided by data to qualitatively and quantitatively investigate the role of natural selection in evolution (1.A.1 & 2.2, 5.3).

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future (1.A.1 & 2.2).

LO 1.4 The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time (1.A.2 & 5.3).

LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effect of selection in the evolution of specific populations (1.A.3 & 1.4, 2.1)

LO 1.7 The student is able to justify data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and the effects of selection in the evolution of specific populations (1.A.3 & 2.1).

LO 1.26 The student is able to evaluate given data sets that illustrate evolution as an ongoing process (1.C.3 & 5.3).

Example of Sample Mathematical Modeling

Because models are simplifications or approximations of reality, students need to recognize how various parameters can affect a model or systems. The focus is on whether students can explain and predict variation in their data with a simple causal mechanism that can be modeled. In Investigation 2: Mathematical Modeling: Hardy-Weinberg, the computer spreadsheets that students build and manipulate enable them to investigate allele inheritance patterns in a theoretically infinite population with inherent randomness. This model allows for the exploration of parameters that affect allele frequencies, such as selection, mutation, and migration. By creating larger and larger populations to minimize fluctuations from expected probabilities, students build their knowledge toward an inference on allele inheritance patterns in a theoretically finite population.

There are dozens of computer models already built and freely available, and most students are familiar with constructing spreadsheets. To gain maximum benefit from this exercise, students should not do too much background preparation; by building and exploring their own models, students develop a more thorough understanding of how genes behave in populations. Once students are confident in their ability to build a model to explore aspects of Hardy-Weinberg equilibrium, they can move to more sophisticated models available on the Internet to answer their own questions about parameters that influence population genetics.

ALIGNMENT TO THE EXAM

Mathematical Modeling and Exam Questions

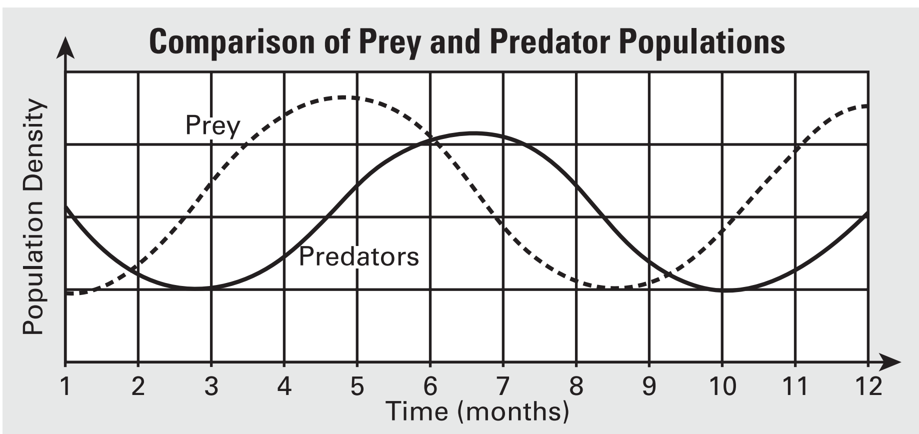
Because biological systems are complex, initial understanding comes from the development of conceptual, mathematical, and computer models that simplify and approximate real situations. Modeling allows students to grapple with the underlying casual factors of a biological phenomenon and interactions, make predictions, bring theory and experiment together, and integrate experimental design and data analysis. One example of mathematical modeling is the development of a spreadsheet model using simple and discrete time steps, difference equations, and iteration; spreadsheets also offer an easy way to construct graphs. Modeling also can be used to convert words to equations and graphs. (The interaction of a predator and a prey population is a straightforward biological system to model because students can begin with a simple accounting of the birth and death of each species.) Other examples of mathematical modeling include population growth curves and the application of the Hardy-Weinberg equilibrium equation. The following sample questions from the *AP Biology Practice Exam* represent items that assess students' ability to glean information about biological phenomena through modeling. The questions are based on the learning objectives from the curriculum framework.

Sample Grid-In Question Using Mathematical Modeling

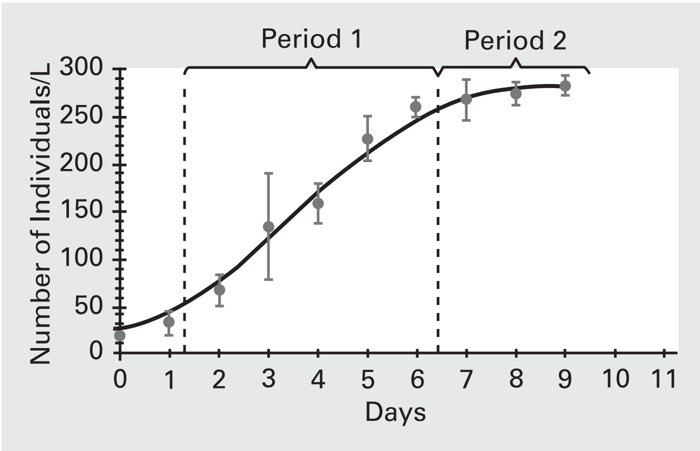
In a population of certain frogs in which the allele for brown skin is dominant to the allele for green skin, a drought leads to selection against green-skinned frogs. When the drought ends, 12 percent of the remaining frogs exhibit the green-skin phenotype. If the population is now in Hardy-Weinberg equilibrium, what will be the frequency of the green-skin allele in the next generation? Provide your answer to the nearest hundredth.

Sample Grid-In Question Using Mathematical Modeling

Use the graph below to calculate the lag time in months between the change in the densities of the prey and the predator populations. Give your answer to the nearest tenth of a month.



Sample Free-Response Question Using Mathematical Modeling
A population of microscopic eukaryotic organisms growing in a large flask had the growth pattern shown.



In one paragraph, **explain** the biological factors that determine the shape of the growth pattern shown above in both period 1 and period 2.

Sample Free-Response Question Using Mathematical Modeling
In fruit flies (*Drosophila melanogaster*), straight wing shape is dominant to curly wing shape. A particular population of fruit flies is in Hardy-Weinberg equilibrium with respect to the alleles for wing shape.

The Hardy-Weinberg equation, given below, is useful in understanding population genetics:

$$p^2 + 2pq + q^2 = 1$$

(a) **Explain** what the terms (p^2, $2pq$, and q^2) represent in the population of fruit flies.

(b) **Describe** one condition that is necessary for the population to be in equilibrium.

RESOURCES

The following resources include ideas for classroom activities and lessons, as well as examples for learning and teaching as you further your exploration into the new AP Biology curriculum. Each resource includes the approximate time involved, a description of the resource and where to find it, the quantitative skills the activity or lesson involves, and where those skills might be applied.

MATHBIOLOGY: HOW TO MODEL A DISEASE http://www.uh.edu/honors/honors-and-the-schools/houston-teachers-institute/curriculum-units/pdfs/2002/new-developments-understanding-body/collins-02-body.pdf	APPROXIMATE TIME 1–2 or more class periods

DESCRIPTION OF RESOURCE

The Houston Teachers Institute (HTI) helps teachers by offering seminars taught by university faculty on subjects teachers themselves request. One area addressed by HTI is the need for students to be able to apply mathematics to their exploration of concepts studied in biology. In this curriculum module, students use a basic mathematical model to study disease in an idealized population of rabbits. The primary model that is used is an SIR (susceptible, infectious, or recovered) model. Students critically examine solutions to problems the teacher will introduce about the rabbit population. The curriculum module includes objectives, suggestions for assessment, and a list of skills that students will develop as they work through the module. This curriculum is an effective means of introducing the creation and use of Excel spreadsheets in mathematical modeling.

QUANTITATIVE SKILLS APPLIED

SP 2 The student can use mathematics appropriately. [Specifically 2.2]

SP 5 The student can perform data analysis and evaluation of evidence. [Specifically 5.2]

LEARNING OBJECTIVES ADDRESSED

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

MATHEMATICAL MODELING IN THE PRE-COLLEGE CLASSROOM NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS AT THE UNIVERSITY OF ILLINOIS http://www.newtrier.k12.il.us/page.aspx?id=16929	APPROXIMATE TIME 1–2 class periods

DESCRIPTION OF RESOURCE

This resource contains several biological modeling projects to download, and chemistry and physics models are also included. The projects represent only one way to introduce students to mathematical modeling; another approach is to allow students to model problems/systems of their own choosing. Although the mathematical modeling environment that is most recommended is Stella, other models include Mathematica, spreadsheets, traditional computer languages, and graphing calculators. Projects especially appropriate for AP Biology include Succession of Grasses, Bird Population, Wolves and Rabbits, Enzyme Reactions, Epidemic Modeling, and Biorhythms. The resources also contain numerous related links.

QUANTITATIVE SKILLS APPLIED

SP 1 (1.1) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1-2.2) The student can use mathematics appropriately.

SP 5 (5.2) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

MATHEMATICAL MANIPULATIVE MODELS: IN DEFENSE OF "BEANBAG BIOLOGY" http://www.lifescied.org/content/9/3/201	APPROXIMATE TIME N/A

DESCRIPTION OF RESOURCE

This paper describes two projects developed by the BioQUEST Curriculum Consortium to demonstrate the utility of mathematical models that involve physical and kinesthetic learning, as well as the usage of authentic data and interactive simulations in the study of biology. The two projects are The Biological Excel Simulations and Tools in Exploratory, Experiential Mathematics (ESTEEM) and Numerical Undergraduate Mathematical Biology Education (Numbers Count). Mathematical manipulative models range from genetics and epidemiology to ecology, photosynthesis, and cancer and help students develop an appreciation for how mathematical reasoning informs problem solving, inference, and communication while enhancing quantitative biology education.

QUANTITATIVE SKILLS APPLIED

SP 1 (1.4) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1-2.2) The student can use mathematics appropriately.

SP 5 (5.2) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.3 (2.2) The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future.

LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations. [See SP 1.4, 2.1]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 3.14 The student is able to apply mathematical routines to determine Mendelian patterns of inheritance provided by data sets. [See SP 2.2]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

THE BIOLOGICAL ESTEEM COLLECTION: EXCEL SIMULATIONS AND TOOLS FOR EXPLORATORY, EXPERIENTIAL MATHEMATICS BIOQUEST CURRICULUM CONSORTIUM http://bioquest.org/esteem/esteem_result.php	APPROXIMATE TIME 1–2 or more class periods

DESCRIPTION OF RESOURCE

Developed by BioQUEST in response to the National Research Council's recommendation for more mathematics in undergraduate biology education, this resource is a collection of downloadable lesson modules designed to integrate quantitative applications in the life sciences. Microsoft Excel was chosen as a general development environment for the ESTEEM project because most biologists and mathematicians have it on their desktops or personal computers, use it for data collection, and find it easy to manipulate. With parameters in Excel easy to change, the modules are adaptable and flexible for students who are engaged in a variety of biology courses, including AP Biology, with module activities ranging from botany and biochemistry to phylogenetics and bioinformatics.

QUANTITATIVE SKILLS APPLIED

SP 1 (1.4) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1, 2.2) The student can use mathematics appropriately.

SP 5 (5.2) The student can perform data analysis and evaluation of evidence.

LEARNING OBJECTIVES ADDRESSED

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations. [See SP 1.4, 2.1]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams, or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

NUMBERS COUNT http://bioquest.org/numberscount/resources/	APPROXIMATE TIME 1–2 or more class periods depending on which aspects of the module the teacher chooses to incorporate into the course

DESCRIPTION OF RESOURCE

Developed by BioQUEST, the Numbers Count Project provides data, tools, and curricular materials for use in the classroom to support student observation, experimentation, statistical analysis, and mathematical modeling. The website focuses on biological problem solving using data. Resources such as an Excel worksheet model for testing Hardy-Weinberg are accessible, in addition to data such as the classic Mendel pea traits or West Nile virus sequences for the envelope protein. Also available is a conceptual model for statistical analysis.

QUANTITATIVE SKILLS APPLIED

SP 1 (1.1, 1.4) The student can use representations and models to communicate scientific phenomena and solve scientific problems.

SP 2 (2.1-2.2) The student can use mathematics appropriately.

SP 5 (5.2) The student can perform data analysis and evaluation of evidence.

SP 6 (6.4) The student can work with scientific explanations and theories.

LEARNING OBJECTIVES ADDRESSED

LO 1.1 The student is able to convert a data set from a table of numbers that reflect a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change. [See SP 1.5, 2.2]

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations. [See SP 1.4, 2.1]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 1.22 The student is able to use data from a real or simulated population(s), based on graphs or models of selection, to predict what will happen to the population in the future. [See SP 6.4]

LO 3.14 The student is able to apply mathematical routines to determine Mendelian patterns of inheritance provided by data sets. [See SP 2.2]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

MATHMODELS http://www.mathmodels.org/problems/	APPROXIMATE TIME 1–2 or more class periods

DESCRIPTION OF RESOURCE

MATHmodels is a mathematical modeling forum that allows you to select from a wide range of mathematical modeling problems. Students can work on problems based on math topic and application area. Practitioners/mentors read and react to solutions or partial solutions and provide hints, if appropriate, and guidance. You can use the resource to enrich a curriculum, discover how mathematical modeling resources can challenge students and enhance the classroom experience, and submit problems to the mathematical modeling database to challenge students globally. MATHmodels allows students to work not only with peers in their own class but also with students around the world.

QUANTITATIVE SKILLS APPLIED

SP 1 The student can use representations and models to communicate scientific phenomena and solve scientific problems. [Specifically 1.4]

SP 2 The student can use mathematics appropriately. [Specifically 2.1, 2.2]

SP 5 The student can perform data analysis and evaluation of evidence. [Specifically 5.2]

LEARNING OBJECTIVES ADDRESSED

LO 1.3 The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future. [See SP 2.2]

LO 1.6 The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations. [See SP 1.4, 2.1]

LO 1.13 The student is able to construct and/or justify mathematical models, diagrams or simulations that represent processes of biological evolution. [See SP 1.1, 2.1]

LO 4.19 The student is able to use data analysis to refine observations and measurements regarding the effect of population interactions on patterns of species distribution and abundance. [See SP 5.2]

References

Jungck, John R., Holly Gaff, and Anton E. Weisstein. "Mathematical Manipulative Models: In Defense of 'Beanbag Biology.' " *Cell Biology* Education 9, no. 3 (2010): 201–211.

Mackowiak, P. A, S. S. Wasserman, and M. M. Levine. "A Critical Appraisal of 98.6°F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich." *Journal of the American Medical Association* 268, no. 12 (1992): 1578.

National Research Council. *America's Lab Report: Investigations in High School Science*. Washington, DC: The National Academies Press, 2005.

National Research Council. *BIO2010: Transforming Undergraduate Education for Future Research Biologists*. Washington, DC: The National Academies Press, 2003.

Shoemaker, A. L. "What's Normal? Temperature, Gender, and Heart Rate." *Journal of Statistics Education* 4, no. 2 (1996).

Stewart, Timothy W. "Measuring Animal Movements in a Natural Ecosystem: A Mark-Recapture Investigation Using Stream-Dwelling Snails." *The American Biology Teacher* 69, no. 1 (2007): e6-e16.

This page intentionally left blank.

Appendix

AP BIOLOGY EQUATIONS AND FORMULAS

STATISTICAL ANALYSIS AND PROBABILITY	
Standard Error	Mean
$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$ ***NOTE:*** *For the purposes of the AP Exam, students will not be asked to manipulate or derive this equation; however, they must know the underlying concepts and applications.*	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Standard Deviation	Chi-Square
$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ ***NOTE:*** *For the purposes of the AP Exam, students will not be asked to manipulate or derive this equation; however, they must know the underlying concepts and applications.*	$\chi^2 = \sum \frac{(o-e)^2}{e}$

s = sample standard deviation (i.e., the sample based estimate of the standard deviation of the population)

$\bar{x}$ = mean

n = size of the sample

o = observed individuals with observed genotype

e = expected individuals with observed genotype

Degrees of freedom equals the number of distinct possible outcomes minus one.

CHI-SQUARE TABLE

Degrees of Freedom								
p	1	2	3	4	5	6	7	8
0.05	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51
0.01	6.64	9.21	11.34	13.28	15.09	16.81	18.48	20.09

LAWS OF PROBABILITY

If A and B are mutually exclusive, then P (A or B) = P(A) + P(B)

If A and B are independent, then P (A and B) = P(A) x P(B)

HARDY-WEINBERG EQUATIONS

$p^2 + 2pq + q^2 = 1$ — p = frequency of the dominant allele in a population

$p + q = 1$ — q = frequency of the recessive allele in a population

METRIC PREFIXES

Factor	Prefix	Symbol
10^9	giga	G
10^6	mega	M
10^3	kilo	k
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	µ
10^{-9}	nano	n
10^{-12}	pico	p

Mode = value that occurs most frequently in a data set

Median = middle value that separates the greater and lesser halves of a data set

Mean = sum of all data points divided by number of data points

Range = value obtained by subtracting the smallest observation (sample minimum) from the greatest (sample maximum)

RATE AND GROWTH

Rate dY/dt	dY= amount of change t = time
Population Growth dN/dt=B-D	B = birth rate D = death rate
Exponential Growth $\frac{dN}{dt} = r_{max} N$	N = population size K = carrying capacity
Logistic Growth $\frac{dN}{dt} = r_{max} N \left(\frac{K-N}{K} \right)$	r_{max} = maximum per capita growth rate of population
Temperature Coefficient Q_{10} ***NOTE:*** *For use with labs only (optional)* $Q_{10} = \left(\frac{k_2}{k_1} \right)^{\frac{10}{t_2 - t_1}}$	t_2 = higher temperature t_1 = lower temperature k_2 = metabolic rate at t_2 k_1 = metabolic rate at t_1 Q_{10} = the *factor* by which the reaction rate increases when the temperature is raised by ten degrees
Primary Productivity Calculation mg O_2/L x 0.698 = mL O_2/L mL O_2/L x 0.536 = mg carbon fixed/L	

Water Potential (Ψ)

$\Psi = \Psi p + \Psi s$

Ψp = pressure potential

Ψs = solute potential

The water potential will be equal to the solute potential of a solution in an open container, since the pressure potential of the solution in an open container is zero.

The Solute Potential of the Solution

$\Psi s = -iCRT$

i = ionization constant (For sucrose this is 1.0 because sucrose does not ionize in water).

C = molar concentration

R = pressure constant (R = 0.0831 liter bars/mole K)

T = temperature in Kelvin (273 + °C)

SURFACE AREA AND VOLUME

Volume of a Sphere $V = 4/3\, \pi r^3$	r = radius
Volume of a Cube (or Square Column) $V = l\, w\, h$	l = length h = height w = width
Volume of a Column $V = \pi r^2 h$	A = surface area V = volume
Surface Area of a Sphere $A = 4 \pi r^2$	Σ = Sum of all a = surface area of one side of the cube
Surface Area of a Cube $A = 6\, a$	
Surface Area of a Rectangular Solid $A = \Sigma$ (surface area of each side)	

Dilution – used to create a dilute solution from a concentrated stock solution

$C_i V_i = C_f V_f$

i = initial (starting)

C = concentration of solute

f = final (desired)

V = volume of solution

Gibbs Free Energy

$\Delta G = \Delta H - T\Delta S$

ΔG = change in Gibbs free energy

ΔS = change in entropy

ΔH = change in enthalpy

T= absolute temperature (in Kelvin)

pH = – log [H+]

NOTE: *For the purposes of the AP Exam, students will not be asked to manipulate or derive this equation; however, they must know the underlying concepts and applications.*