

---

# Quantitative Skills in the **AP**<sup>®</sup> Sciences

A REFERENCE GUIDE FOR STUDENTS IN:

AP BIOLOGY

AP CHEMISTRY

AP ENVIRONMENTAL SCIENCE

AP PHYSICS

---

# Introduction

## Quantitative Skills and the AP Science Courses

Each AP science course (biology, chemistry, environmental science, and the four physics courses) focuses on the development of quantitative skills such as using measurements, gathering and evaluating data, using quantitative evidence to support claims, and connecting empirical information to scientific theory. In each AP science course you take, you will perform several hands-on college-level investigations and other activities that are designed not only to meet the learning objectives for the course *but also to give you multiple opportunities to acquire, develop, and apply these quantitative skills*. By the end of a course, you will be able to successfully organize and work with data so that you can effectively communicate your conclusions. You will be able to think and act like a scientist!

## How to Use This Guide

This guide is designed to be used as a reference as you navigate various AP science courses; you can use it to refresh your memory on how to perform calculations on the Equations and Formulas sheet for your course as you are completing classroom or at-home assignments. It is not meant to be a comprehensive guide to quantitative analysis, a substitute for your textbook and/or classroom instruction, or a review for the AP Exam.

The first two chapters, Chapter 1: Collecting and Reporting Data and Chapter 2: Data Analysis, contain general information that is relevant to all AP science courses, along with examples. These chapters are followed by additional chapters that contain course-specific quantitative-skills information for AP Biology, AP Chemistry, AP Environmental Science, AP Physics 1, AP Physics 2, AP Physics C: Mechanics, and AP Physics C: Electricity and Magnetism.

Throughout this guide, you will find hyperlinks to videos from Khan Academy that you can use for additional help when you need it. Each hyperlink is also accompanied by a QR code that you can use to access a video with a smartphone or tablet, which will be helpful if you are using a hard copy of this guide. You may use any QR reader app (many are free) to read the codes.

## CHAPTER 1

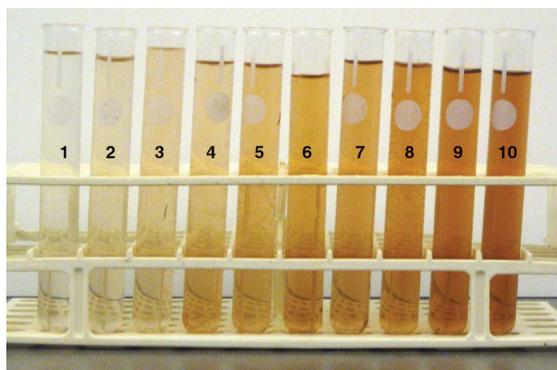
# Collecting and Reporting Data

## Types of Data

Data can be classified as qualitative or quantitative:

- Qualitative data
  - ♦ Are observed rather than measured
  - ♦ Include written descriptions, videos, photographs, or live observations
  - ♦ Examples include observations of appearance, behaviors, smell, taste, etc.
- Quantitative data
  - ♦ Are measured and recorded in numerical form
  - ♦ Examples include absorbance, size, time, height, and mass

Qualitative data and quantitative data are both important and not always used completely separate from each other. Qualitative data can be coded or organized in a quantitative way for the purpose of interpretation or analysis. For instance, in the AP Biology Enzyme Catalysis lab, a color palette (figure 1.1) is used to qualitatively determine the amount of oxygen produced when hydrogen peroxide is degraded by the turnip peroxidase enzyme. By numbering the colors 1–10, the qualitative data obtained from the experiment can be converted to quantitative data. Alternately, quantitative data can be obtained from this experiment by using a spectrophotometer to measure the absorbance or percent transmittance of the samples. The purpose or anticipated outcome of an experiment will determine which type of data you choose to collect and how to organize it.



**Figure 1.1:** Turnip Peroxidase Color Chart

# Measurement

## Units: Use of Metric Prefixes

There are two commonly used systems of measurement in the world, which differ in the units they use for length, mass, and time. The first is the United States Customary System (USCS, formerly called the English system) of feet, pounds, and seconds. The second is the metric system of meters, kilograms, and seconds. In 1960, the metric system was adopted by an international committee in Paris as the worldwide standard for science and is now referred to as the *Système International* or SI. A subset of the metric system is the centimeter-gram-second (cgs) system that is commonly used in atomic physics and chemistry. The meter-kilogram-second (mks) system is another subset commonly used in physics (specifically mechanics). In science, medicine, and government in the United States, the SI system is often used alongside the USCS system. In this guide, we will use the SI system of measurement, which is the preferred measurement system of science.

## Fundamental Units

Most physical quantities, such as velocity, acceleration, force, momentum, and energy can ultimately be expressed in terms of three basic units of length, mass, and time. These three units are referred to as **fundamental units** because they can be used to define all other elements in a particular system of measurement.

Table 1.1 summarizes the fundamental units for the metric and USCS systems of measurement.

**Table 1.1: Fundamental Units of Measurement**

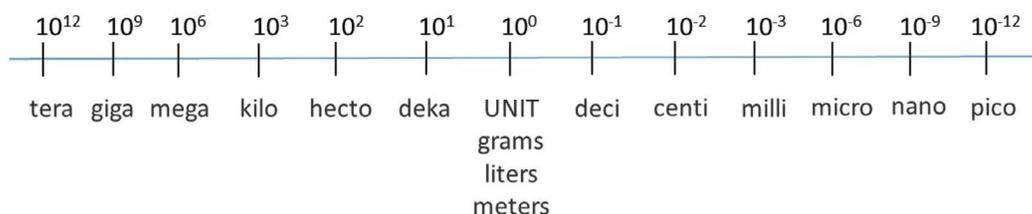
System	Length	Mass	Time
SI (MKS)	meter	kilogram	second
SI (CGS)	centimeter	gram	second
USCS	feet	slug	second

The units of measure in the SI system are often preceded by prefixes to indicate the appropriate size of a measurement. Each prefix represents a power of 10 and has a symbol that is added to the measurement for reporting. For example, the prefix *milli-* indicates 1/1000, which means that there are 1000 milligrams in a gram. So when describing the mass of an object or substance that is very small, it is reported as 3.42 milligrams rather than 0.00342 grams. Table 1.2 lists the prefixes that you will most commonly use in your AP science courses.

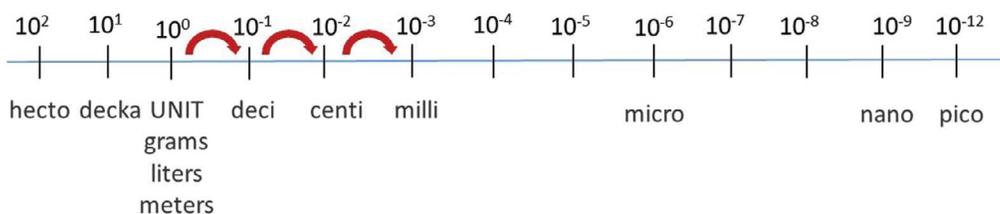
Table 1.2: Common SI Prefixes

Prefix	Symbol	Multiplier	Number
tera	T	$10^{12}$	1 trillion
giga	G	$10^9$	1 billion
mega	M	$10^6$	1 million
kilo	k	$10^3$	1,000
hecto	h	$10^2$	100
deka	da	$10^1$	10
UNIT (grams/ liters/meters)		$10^0$	1
deci	d	$10^{-1}$	0.1
centi	c	$10^{-2}$	0.01
milli	m	$10^{-3}$	0.001
micro	$\mu$	$10^{-6}$	1 millionth
nano	n	$10^{-9}$	1 billionth
pico	p	$10^{-12}$	1 trillionth

Another way of thinking about this is to use a place-value representation:



One way to convert from one unit to another is to use the above representation to count how many decimal places should be used for the adjustment. So, using the example above, if you wanted to convert 0.00342 grams to milligrams, you would start at the base unit then count until you get to the *milli*- prefix, as shown below.



Since we moved three places to the right, we will move the decimal point in our number — 0.00342 — to the right also. By doing this, we find that 0.00342 grams is equal to 3.42 milligrams. If we move the decimal point two places to the right instead of three, we find that

0.00342 grams is equal to 0.342 centigrams, and moving the decimal point one place to the right shows that 0.00342 grams is equal to 0.0342 decigrams.

If you need further review of the metric system, try this tutorial:



[Khan Academy: U.S. customary and metric units](#)

## Dimensional Analysis: Unit Conversions

In AP science courses you will frequently have to analyze relationships between physical quantities. This may require you to convert between units to describe equivalent amounts of the data you are reporting. In doing this, the amounts of data you are describing remain the same. You are only changing the way you report these amounts. Converting units is a type of **dimensional analysis** for which the **factor-label method** is helpful. For example, let's say we want to convert 650 mL to liters.

We know that there are 1000 mL in one liter:

$$1L = 1000mL$$

We first convert this equation to conversion factors:

$$\frac{1L}{1000mL} \quad \frac{1000mL}{1L}$$

Multiplying a quantity by these conversion factors changes the units, but leaves the quantity unchanged. We next choose a conversion factor that will convert our quantity, 650 mL, from units of mL to units of liters:

$$650\cancel{mL} \times \frac{1L}{1000\cancel{mL}} = \frac{650L}{1000} = 0.650L$$

The conversion factor was chosen so that when units are cancelled out (the diagonal lines in the accompanying examples), the desired unit remains. In choosing the conversion factor, we put the mL in the denominator so that it cancels out, and we are left with L. By cancelling out mL, we converted from mL to L.

We can also multiply by a series of conversion factors. For example, consider converting from three miles to meters, given the conversion from miles to feet (there are 5280 feet in a mile) and the conversion from feet to meters (there are 3.28 feet in a meter).

There are 5280 feet in a mile, and 3.28 feet in a meter:

$$1 \text{ mile} = 5280 \text{ feet}$$

$$3.28 \text{ feet} = 1 \text{ meter}$$

This gives us four conversion factors:

$$\frac{1\text{mile}}{5280\text{feet}} \quad \frac{5280\text{feet}}{1\text{mile}} \quad \frac{3.28\text{feet}}{1\text{meter}} \quad \frac{1\text{meter}}{3.28\text{feet}}$$

So the conversion would look like this:

$$2 \cancel{\text{miles}} \times \frac{5280 \cancel{\text{ft}}}{1 \cancel{\text{mile}}} \times \frac{1 \cancel{\text{m}}}{3.28 \cancel{\text{ft}}} = 3219 \text{m}$$

You will often use this method to determine how to make solutions. For example, how many grams of sodium hydroxide (NaOH, molar mass = 40 g/mol) would we need if we wanted to make 500 mL of a 0.40 M (moles per liter) solution? In this case, the molar mass of NaOH (1 mol NaOH = 40 g NaOH) leads to the following conversion factors:

$$\frac{1 \text{mol NaOH}}{40 \text{g NaOH}} \quad \frac{40 \text{g NaOH}}{1 \text{mol NaOH}}$$

Similarly, our target concentration of 0.40 M gives us the following ratios:

$$\frac{0.40 \text{mol NaOH}}{1 \text{L}} \quad \frac{1 \text{L}}{0.40 \text{mol NaOH}}$$

To determine the mass of NaOH needed to make 500 mL of solution we start with 500 mL and multiply as follows:

$$500 \cancel{\text{mL}} \times \frac{1 \cancel{\text{L}}}{1000 \cancel{\text{mL}}} \times \frac{0.4 \cancel{\text{mol}}}{1 \cancel{\text{L}}} \times \frac{40 \cancel{\text{g}}}{1 \cancel{\text{mol}}} = 8 \text{g NaOH needed}$$

These tutorials will help you refresh your memory on how to do unit conversions:



[Khan Academy: Unit conversion within the metric system](#)



[Khan Academy: Converting within the metric system](#)

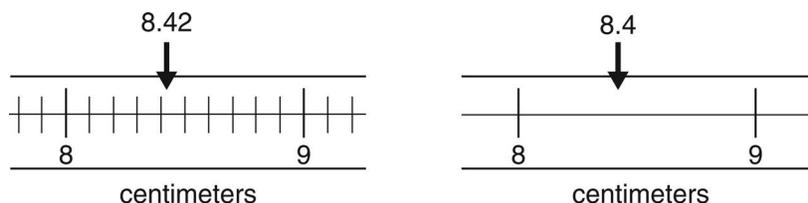
## Significant Digits

To ensure that you are reporting your data to the correct degree of precision, the data you record during an experiment should include only **significant digits** (also called **significant figures**). These are:

- The digits that are meaningful in a measurement or a calculation.
- Determined by the measurement device used during the experiment.
  - ◆ If you use a digital device, record the measurement value exactly as it is shown on the screen.
  - ◆ If you read the result from a ruled scale (such as a ruler or graduated cylinder), the value that you record should include each digit that is certain and one uncertain digit.

For example, figure 1.2 shows the same measurement made with two different scales, which vary in their precision of measurement. On the left, the digits 8 and 4 are certain because they are shown by markings on the scale and it is clear that the measurement is at least 8.4.

The digit 2 is an estimate of how far the measurement is beyond 8.4, so that is the uncertain digit. This measurement (8.42 cm) has three significant digits. The scale on the right has markings at 8 and 9. The 8 is certain, but you must estimate how far the measurement is beyond 8, so 4 is the uncertain digit. This measurement is 8.4 cm. Even though the measurement on the right is the same as the measurement on the left, it has only two significant digits because the markings are farther apart, and thus there is less precision to the measurement being made.



**Figure 1.2:** Different Significant Digits from Different Scales

**Uncertainties** in measurements should always be rounded to one significant digit. When measurements are made with devices that have a ruled scale, the uncertainty is half the value of the precision of the scale. The markings on the device will show the precision. Looking at the example shown in figure 1.2 above, the scale on the left has markings every 0.1 cm, so the uncertainty is half this, which is 0.05 cm. The correct way to report this measurement is  $8.43 \pm 0.05$  cm. The scale on the right has markings every 1 cm, so the uncertainty is 0.5 cm. The correct way to report this measurement is  $8.4 \pm 0.5$  cm.

Table 1.3 presents the rules you should follow in determining which digits in a number that represents a measured value are meaningful (in the sense described above) and therefore significant.

**Table 1.3: Rules for Significant Figures**

Rule	Examples
Non-zero digits are always significant.	4,735 km has four significant digits. 573.274 in. has six significant digits.
Zeros before other digits are not significant.	0.38 m has two significant digits. 0.002 in. has one significant digit.
Zeros between other digits are significant.	42.907 km has five significant digits. 0.00706 in. has three significant digits. 8,005 km has four significant digits.
Zeros to the right of all other digits are significant if they are to the right of the decimal point.	975.3810 cm has seven significant digits. 471.0 m has four significant digits.
It is impossible to determine whether zeros to the right of all other digits are significant if the number has no decimal point.	8,700 km has at least two significant digits, but the exact number is unknown. 20 in. has at least one significant digit, but the exact number of significant digits is unknown.

Table 1.3: Rules for Significant Figures (*continued*)

Rule	Examples
If a number is written with a decimal point, zeros to the right of all other numbers are significant.	620.0 km has four significant digits. 5,100.4 m has five significant digits. 670. in. has three significant digits.
All digits in the coefficient of a number written in scientific notation are significant.	$6.02 \times 10^4$ cm has three significant digits

Note that it is good scientific practice to use **scientific notation** (see chapter 2). If you use scientific notation, then all digits shown are always significant.

If you need additional review on significant figures, this tutorial can help:



[Khan Academy: Intro to significant figures](#)

## Data Tables

Data tables allow you to gather your data in one place so that it can be organized, compared, or analyzed in a meaningful way for interpretation. When constructing a data table you need to be sure to include both the independent and dependent variables.

- **Independent variable**
  - ◆ Also called the explanatory or controlled variable
  - ◆ The variable that the researcher controls or manipulates
  - ◆ Not changed by the other variable(s) measured in the experiment
  - ◆ Examples: time, distance, velocity, acceleration, concentration, light intensity
- **Dependent variable**
  - ◆ Also called the response or experimental variable.
  - ◆ The response to the independent variable—what is measured.
  - ◆ Example: population growth: The number of individuals in a population will change with time, so the growth of the population is a dependent variable since it is dependent on time (the independent variable).
  - ◆ Example: If you were interested in the velocity of an object as a function of time, then velocity could be a dependent variable, while time would be the independent variable. On the other hand, velocity could be an independent variable if you investigated acceleration as a function of velocity.

## Elements of Effective Data Tables

You may often use computer software to create data tables to communicate the results of an investigation. However, whether you are using software or drawing by hand, you should keep in mind the following elements of effective data tables, shown in figure 1.3:

1. **A meaningful title:** This is a title that informs the reader about the experiment and exactly what is being measured.

- Independent and dependent variables: These are typically with the independent variable on the left side of the data table and the dependent variables on the right.
- Units: Be sure that units are clearly indicated for each variable.
- Data: There should be data for each repeated trial.

**Effect of Glucose Concentration on Plant Growth by Mass** ↙ Meaningful title

Glucose Solution (M)	Mass of Plants (g)						Average
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	
0.0							
0.2							
0.4							
0.6		Data for each repeated trial					
0.8							
1.0							

*Annotations in the original image:*  
 - "Independent variable with units" points to the first column (Glucose Solution (M)).  
 - "Dependent variable with units" points to the top row of the data columns (Mass of Plants (g)).  
 - Red arrows point from the trial columns to the "Data for each repeated trial" text.

**Figure 1.3: Elements of Effective Data Tables**

## Graphs

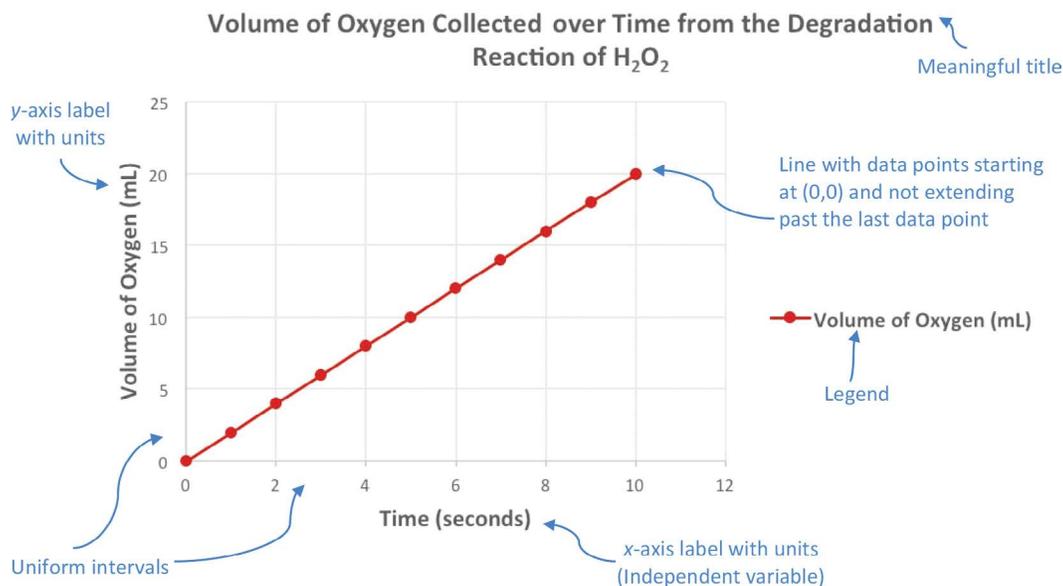
One of the best ways to communicate the results of a scientific investigation is by creating a graph of the data that have been counted, measured, or calculated. Graphs can help you to easily see patterns more easily through a visual display of data and can also help you clearly see how two measured variables affect one another.

### Elements of Effective Graphs

Just as with data tables you may use computer software to create your graphs. However, whether you are using software or graphing by hand, you should keep in mind the following elements required of nearly all effective graphs (illustrated in figure 1.4):

- A meaningful title: This is a title that informs the reader about the experiment and exactly what is being measured.
- Labeled axes with units:
  - ◆ The **x-axis** is the horizontal axis, and it usually denotes the independent variable.
  - ◆ The **y-axis** is the vertical axis, and it usually denotes the dependent variable.
  - ◆ Note that the axes do not always need to denote dependent versus independent variables. In physics, we often choose the axes for straight-line fitting so that the slope or y-intercept provides physical information. For example, for various satellites orbiting the Earth, we might choose to graph period squared ( $T^2$ ) on the y-axis, and radius cubed ( $R^3$ ) on the x-axis in order to see if the orbits obey Kepler's third law.
- Uniform intervals: For example, if one interval on the x-axis corresponds to 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.
- Identifiable lines or bars: Using different colors or patterns and including a legend will help the reader distinguish one line or bar from the others. You can also label each line or bar.

- Origin: The graph should clarify whether the data and any **trend lines** start at the origin (0,0) or not. A trend 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 clearly indicates that this is a prediction about what may (or could) happen if additional data were to be obtained.
- Error bars: For some of the labs you perform in class, you should consider the variability (or confidence) of your data in your analysis and use error bars on your graphical displays when appropriate (see the discussion of standard deviation and standard error later in this chapter).



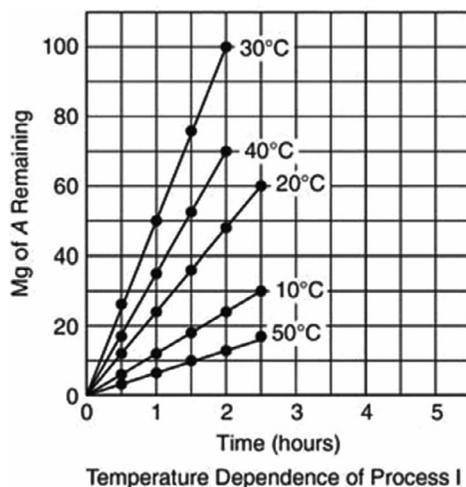
**Figure 1.4:** Example of an Effective Graph

## Types of Graphs

### Line Graphs

Line graphs are plotted on  $x$ - $y$  axes and offer a good visual representation of the relationship between two variables; in other words, how one variable is affected by the other as it increases or decreases. Line graphs can contain one line or multiple lines that represent the data. Clear trends in the data can be seen by the direction of the line(s) on a graph. Line graphs are advantageous because they can sometimes allow you to predict the results of data that have not yet been collected, since the line implies a continuous response of the dependent variable.

Figure 1.5 shows an example of a line graph. It is a type of rate graph called a progress curve, because it shows an amount of a substance on the  $y$ -axis and time on the  $x$ -axis. There are several different curves plotted on the graph and each one is labeled with a different temperature.



**Figure 1.5:** Example of a Line Graph with Several Sets of Data

If you need additional review on line graphs, try this tutorial:

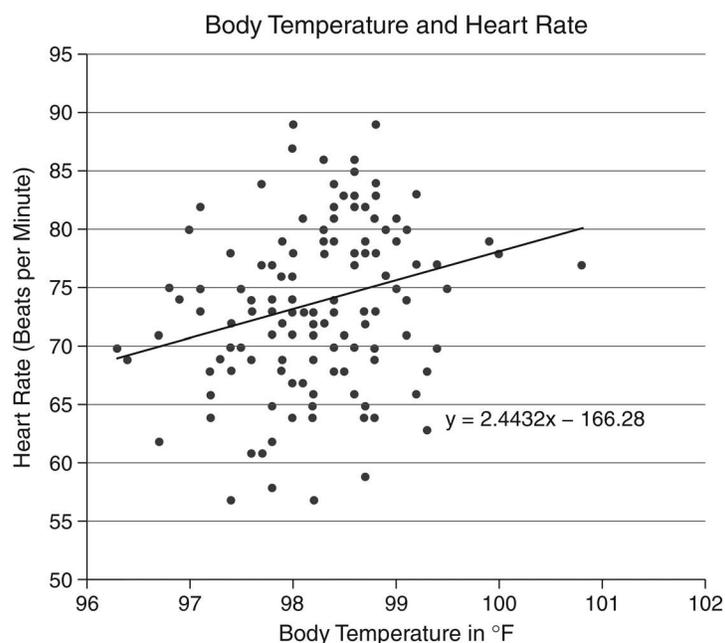


[Khan Academy: Introduction to line plots](#)

## Scatter Plots

Scatter plots are plotted on  $x$ - $y$  axes and are also used to compare two variables. However, in scatter plots, data are presented as an assortment of points that may or may not show one or more of the linear relationships between the two variables that are commonly presented in line graphs. In order to determine whether there is a linear relationship between the two variables, a linear regression (see the Curve Fitting section later in this chapter) can be calculated and plotted to help make the pattern clearer. Keep in mind that the data shown in scatter plots do not have to have a linear relationship.

Figure 1.6 is an example of a scatter plot with a linear regression line. Linear regression lines can indicate a pattern in the data that may not be apparent by looking at the dots alone. We see from the graph that there is a relationship between heart rate and temperature.



**Figure 1.6:** A Scatterplot with a Linear Regression Line

To learn more about scatter plots, review these tutorials:



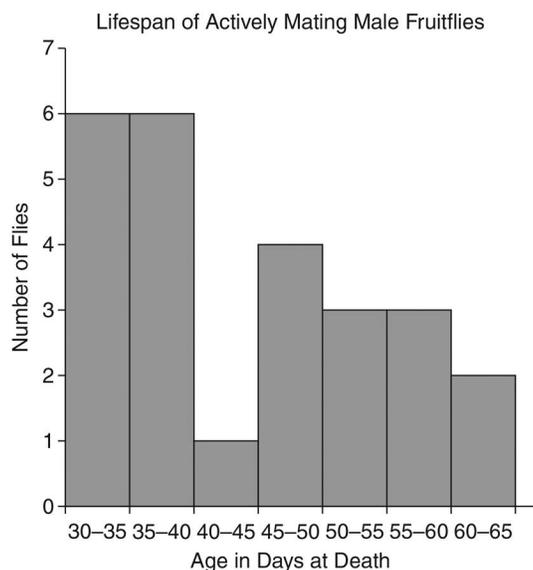
[Khan Academy: Constructing a scatter plot](#)



[Khan Academy: Regression line example](#)

## Histograms

Histograms are plotted on  $x$ - $y$  axes and show the distribution of numerical data. Creating this kind of graph requires setting up grouped intervals called bins for the range of values tested. The range is divided into equal intervals, then the number of measurements that fit into each bin are counted and graphed, which results in a frequency diagram. In figure 1.7, the bins (i.e., the grouped intervals) are the five-day age ranges plotted on the  $x$ -axis: the number of flies in each five-day interval is the measurement represented by the height of each bar as read from the  $y$ -axis.



**Figure 1.7:** A Histogram Showing Nonparametric Data

You can use the tutorials below to review creating and interpreting histograms:



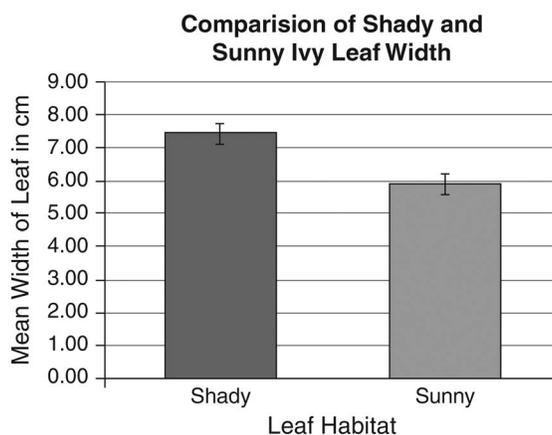
[Khan Academy: Creating a histogram](#)



[Khan Academy: Interpreting a histogram](#)

## Bar Graphs

Bar graphs are plotted on  $x$ - $y$  axes and are used to compare data: typically they represent the summary statistics of a data set, such as the mean or median. Drawn either vertically or horizontally, each bar can represent a category or variable. The relationship between the variable and each group can be determined based on the height or length of the bars, respectively. In figure 1.8, the  $x$ -axis shows the categories of leaf habitat (sunny and shady) that are being compared, and the  $y$ -axis shows the mean width of leaves measured in each environmental condition. Note that when the independent variable is numerical, you should use a histogram (see the previous section) to represent your data instead of a bar graph.



**Figure 1.8:** Example of a Bar Graph

The following tutorials can help you learn more about creating and reading bar graphs:



[Khan Academy: Creating a bar graph](#)



[Khan Academy: Reading bar graphs](#)

## Box and Whisker Plots

Box and whisker plots are plotted on  $x$ - $y$  axes and are used to look at the range of data that has been measured. The advantages of box and whisker plots are that they give you a quick idea of the spread (variability) of your data, the skewness (the amount of skew, or asymmetric distribution), and how different it is from the other data. Box and whisker plots also provide a quick estimate for comparing data sets. A limitation is that you really need at least 10 data points per sample in order to construct an effective box and whisker plot.

To construct a box and whisker plot, gather a five-number summary (or five statistical summary) of data: the (1) minimum, (2) maximum, (3) median, (4) first quartile, and (5) third quartile. You start by numerically ordering your data, and then determine the **median**, which is the number in the middle (or the average of the two middle numbers if the data set contains an even number of observations) of the ordered data set. Once this has been done, the data can be divided into **quartiles** then plotted on a graph.

For example, let's look at the following data set, where researchers measured the percentage of leaf decay in bags containing three different types of leaves.

Bag Number	% Decay		
	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	

First we numerically order each set of data from least to greatest. Then we find the median, which is the number in the middle of this ordered list:

% Decay		
Ash	Sycamore	Beech
32	33	
44	35	11
48	40	15
48	40	19
51	44	21
57	48	32
63	52	34
70	63	

minimum → (points to 32 in Ash column)

→ median (points to the red line between 48 and 51 in Ash column)

maximum → (points to 70 in Ash column)

In each case, since there is an even number of data points, the median is the average of the two middle points: 49.5 for the ash, 42 for the sycamore, and 20 for the beech. The **minimum** in each data set is the smallest number and the **maximum** is the largest number.

Once you have determined the median, you then determine the quartiles by dividing each half of the data in half:

% Decay			
Ash	Sycamore	Beech	
32	33		
44	35	11	
48	40	15	← minimum
48	40	19	
51	44	21	← median
57	48	32	
63	52	34	← maximum
70	63		

1<sup>st</sup> quartile → (points to the row containing 44, 35, 11)

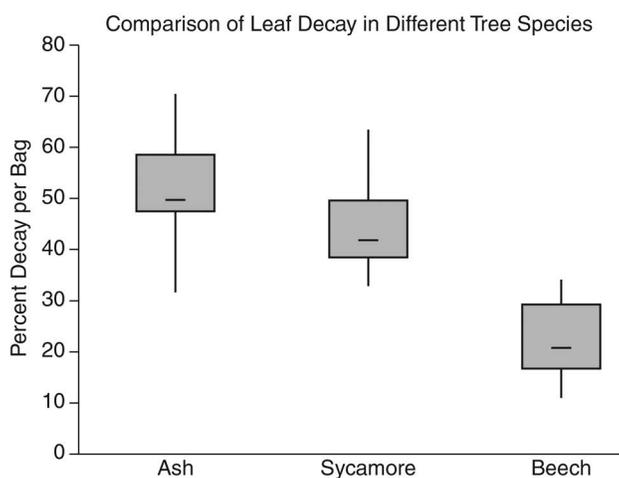
3<sup>rd</sup> quartile → (points to the row containing 63, 52, 34)

So, the five-number summary for this data is as follows:

#### % Decay: Five-Number Summary

	Ash	Sycamore	Beech
<b>Minimum</b>	32	33	11
<b>Quartile 1</b>	46	37.5	13
<b>Median</b>	49.5	42	20
<b>Quartile 3</b>	60	50	33
<b>Maximum</b>	70	63	34

These five values are used to create a box and whisker plot representing the spread of values in a data set. The box and whisker plot for this data is shown in figure 1.9.



**Figure 1.9:** Example of a Box and Whisker Plot

In the graph, the tops and bottoms of the vertical lines typically show the range of the data set. The top of each box shows the upper (third) quartile, the bottom of each box shows the lower (first) quartile, and the horizontal line inside the box represents the median.

The following tutorials can help you learn more about creating and reading box and whisker plots:



[Khan Academy: Constructing a box plot](#)

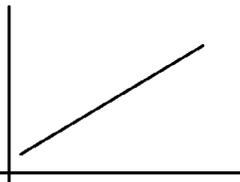
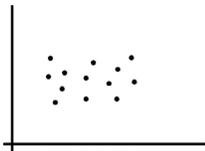
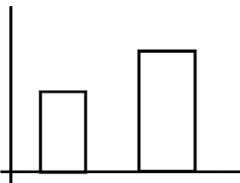
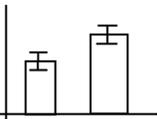
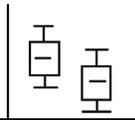


[Khan Academy: Reading box plots](#)

## Summary: Types of Graphs

We have covered several different types of graphs in this chapter. Table 1.4 summarizes all of them and the instances in which you would use each one.

**Table 1.4: Types of Graphs and When to Use Them**

Type of Graph	Example	Examples of When to Use
Line Graph		<ul style="list-style-type: none"> <li>To track changes over time/ concentration, etc.</li> <li>To compare changes over the same time period for multiple groups/ treatments</li> </ul>
Scatterplot		<ul style="list-style-type: none"> <li>To compare two variables that may or may not have a linear relationship</li> </ul>
Histogram		<ul style="list-style-type: none"> <li>To show how values in a data set are distributed across evenly spaced (or equal) intervals</li> <li>To explore the relationship between two or more (in a three-dimensional plot) variables</li> </ul>
Bar Graph		<ul style="list-style-type: none"> <li>To compare multiple groups/ treatments to each other</li> </ul>
Box and Whisker Plots		<ul style="list-style-type: none"> <li>To show the variability in your sample</li> </ul>

For more detailed information on using Microsoft Excel or Google Sheets for graphing, watch the following tutorials:



[Create a chart from start to finish](#)



[Making a graph in Google Spreadsheet and inserting it into a Google Doc](#)

## CHAPTER 2

# Data Analysis

When you complete a laboratory investigation, it is important to make sense of your data by summarizing it, describing the distributions, and clarifying “messy” data. Analyzing your data will allow you to do this.

## Working with Data

Data analysis may involve calculations, such as dividing mass by volume to determine density or subtracting the mass of a container from the total mass to determine the mass of the contents. Using the correct rules for significant digits during these calculations is important to avoid misleading or incorrect results.

When adding or subtracting quantities, the result should have the same number of decimal places (digits to the right of the decimal) as the fewest number of decimal places in any of the numbers that you are adding or subtracting.

Table 2.1 presents examples and explains how the proper results should be written.

**Table 2.1: Writing Your Results When Adding or Subtracting**

Example	Explanation
$3.\underline{7} \text{ cm} + 4.6083 \text{ cm} = 8.\underline{3} \text{ cm}$	The result is written with one decimal place because the number 3.7 has only one significant digit to the right of the decimal.
$48.\underline{3506} \text{ m} - 6.\underline{28} \text{ m} = 42.\underline{10} \text{ m}$	The result is written with two decimal places because the number 6.28 has only two significant digits to the right of the decimal.
$(8 \text{ km} - 4.2 \text{ km}) + 1.94 \text{ km} = 6 \text{ km}$	The result is written with zero decimal places because the number 8 has zero significant digits to the right of the decimal.

Notice that the result of adding and subtracting has the correct number of significant digits if you consider significant digits to the right of the decimal.

When multiplying and dividing a set of numbers, look for the number with the fewest significant digits. Your result should have that number of significant digits. Table 2.2 explains how to apply this concept.

Table 2.2: Writing Your Results When Multiplying or Dividing

Example	Explanation
$5.246 \text{ in.} \times 2.30 \text{ in.} = 12.1 \text{ in.}$	The result is written with three significant digits because 2.30 has three significant digits.
$0.038 \text{ cm} \div 5.273 \text{ cm} = 0.0072 \text{ cm}$	The result is written with two significant digits because 0.038 has two significant digits.
$76.34 \text{ m} \times 2.8 \text{ m} = 2.1 \times 10^2 \text{ m}$	The result is written with two significant digits because 2.8 has two significant digits. [Note that scientific notation had to be used because writing the result as 210 would have an unclear number of significant digits.]

When calculations involve a combination of operations, you must retain one or two extra digits at each step to avoid any round-off errors; at the end of the calculation, you must round to the correct number of significant digits.

An exception to these rules occurs when a calculation involves count data, such as the number of times a ball bounces, or the number of waves that pass a point during a time interval. As shown in the following example, do not consider exact numbers when determining significant digits in a calculation.

### Example

While performing the Millikan oil-drop experiment, you find that a drop of oil has an excess of three electrons. What is the total charge of the drop?

$$\text{Charge} = (\text{number of electrons})(\text{charge per electron})$$

$$q = ne$$

$$= (3 \text{ electrons})(1.6 \times 10^{-19} \text{ C/electron})$$

$$= 4.8 \times 10^{-19} \text{ C}$$

When determining the number of significant digits in the answer we ignore the number of electrons because it is an exact number.

## Scientific Notation

When manipulating data, there will be many times when the numbers that you calculate will be either too large or too small to be conveniently expressed as decimals. To make it easier to work with these very large or very small numbers, scientists use scientific notation. In scientific notation, a number is written as a coefficient multiplied by the base 10 raised to some exponent. Let's look at Avogadro's number to better understand the components:

$$6.022 \times 10^{23}$$

↖ coefficient
↖ exponent  
↖ base

The coefficient must be between 1 and 10, and the exponent must be an integer. Very large numbers will have a positive exponent, while very small numbers will have a negative exponent; for example:

$$10000 = 1 \times 10^4$$

$$1000 = 1 \times 10^3$$

$$100 = 1 \times 10^2$$

$$10 = 1 \times 10^1$$

$$1 = 10^0$$

$$1/10 = 0.1 = 1 \times 10^{-1}$$

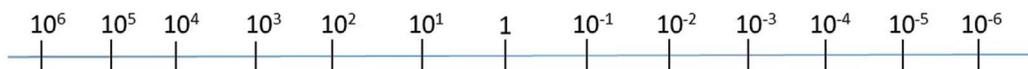
$$1/100 = 0.01 = 1 \times 10^{-2}$$

$$1/1000 = 0.001 = 1 \times 10^{-3}$$

$$1/10000 = 0.0001 = 1 \times 10^{-4}$$

So, a number such as 0.0000000000757 would be written in scientific notation as  $7.57 \times 10^{-12}$ , while a number like 218,000,000 would be written as  $2.18 \times 10^8$ .

Another way of thinking about this is to use the following representation for the place values:



You can rewrite a number in scientific notation by simply using this representation to count how many decimal places to move the decimal point. If the number you are converting is greater than 10, then the decimal point is moved to the left on the line, while if it is less than 1, the decimal point is moved to the right. For instance, if you wanted to convert 0.000436 into scientific notation, you would start at the base unit—the 1—then count the decimal places you would have to move until the coefficient is between 1 and 10, as shown below.



This tells us that we need to move the decimal point four places to the right:



So, 0.000436 would be written in scientific notation as  $4.36 \times 10^{-4}$ .

**For more detailed information on using scientific notation, watch the following tutorial:**



[Khan Academy: Introduction to scientific notation](#)

## Calculations Using Percentages

### Percent Change

When working with data, sometimes we need to compare unequal quantities or scales; in order to do this we normalize the data. One way to do this is to compare the percent change over time. We use the following formula to calculate percent change:

$$\% \text{ change} = \frac{\text{final value} - \text{initial value}}{\text{initial value}} \times 100$$

For example, in the AP Biology Diffusion and Osmosis lab investigation, dialysis bags are first filled with a sucrose solution and then placed in water for 30 minutes. We measure the mass of each bag before and after it sits in the water for 30 minutes, and report this as a percent change in mass. If the mass of a dialysis bag at the beginning of the experiment was 12.2 g and at the end of the experiment it was 16.7 g, the percent change is

$$\frac{16.7 - 12.2}{12.2} \times 100 = 36.9\%$$

Percent change can also be negative. What if in the previous example the mass at the beginning of the experiment was 16.7 g and the mass at the end of the experiment was 12.2 g? Let's look at this new calculation:

$$\frac{12.2 - 16.7}{16.7} \times 100 = -26.9\%$$

In the first calculation the positive result indicates that the dialysis bags gained mass. However, in the second calculation the negative result indicates that the dialysis bag lost mass.

### Percent Difference

There are times when you may need to calculate the percent difference between two experimental results to see how they compare to each other. To calculate percent difference, we use the following formula:

$$\% \text{ difference} = \left| \frac{x_1 - x_2}{(x_1 + x_2) / 2} \right| \times 100$$

where  $x_1$  is the first data point and  $x_2$  is the second data point. The numerator is the difference between the measurements, and the denominator is the average of the measurements. The two vertical lines on either side of the fraction indicate that we are using the absolute value of the calculation.

**Example**

There are two cars traveling at different speeds: one at 25 mph and the other at 33 mph. We want to know the percent difference between the speeds of the two cars. The calculation would be

$$\begin{aligned}\% \text{ difference} &= \left| \frac{x_1 - x_2}{(x_1 + x_2)/2} \right| \times 100 \\ &= \left| \frac{25 - 33}{(25 + 33)/2} \right| \times 100 \\ &= \left| \frac{8}{29} \right| \times 100 \\ &= 0.2759 \times 100 = 27.6\%\end{aligned}$$

This means that there is a 28% difference between the speeds of the two cars.

**Percent Error**

Percent error is a calculation that is done when you want to compare your results to a known or predicted theoretical value. We use the following formula to calculate percent error:

$$\% \text{ error} = \frac{|\text{experimental value} - \text{theoretical value}|}{\text{theoretical value}} \times 100$$

Notice that we are using the absolute value of the difference between the experimental value and the theoretical value.

**Example**

Calculate the percent error of a titration of 3.0% hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) with potassium permanganate ( $\text{KMnO}_4$ ), as in AP Chemistry Investigation 8: Oxidation-Reduction Titration. If we performed this investigation and calculated the concentration of  $\text{H}_2\text{O}_2$  in our sample to be 2.74%, our calculation would be:

$$\begin{aligned}\% \text{ error} &= \frac{|2.74 - 3.0|}{3.0} \times 100 \\ &= \frac{0.26}{3.0} \times 100 \\ &= 8.67\%\end{aligned}$$

This means that our titration yielded data that was in error by 8.67% relative to what was expected.

**Rate Calculations**

You may occasionally have to determine a rate of change when you are processing data from an experiment. Examples include a rate of reaction, growth rate, speed, and acceleration. Each of these describes how a quantity changes over time. The change over time can be expressed as

$$\frac{\Delta Y}{\Delta t} = \frac{\text{the change in the dependent variable}}{\text{the change in time}}$$

where  $\Delta Y$  represents the change on the y-axis and  $\Delta t$  represents the change on the x-axis (time).

**Example**

Suppose you were doing an AP Physics lab and wanted to calculate the magnitude of the average velocity (speed) of an object. You would do this by calculating the displacement traveled during a particular period of time. So, if you pushed a toy car across the floor and it traveled in a straight line from 1.0 meter to 4.0 meters in 8 seconds, you would calculate the speed as follows:

$$\text{average velocity} = \frac{\text{displacement}}{\text{time}} = \frac{(4.0 - 1.0)\text{meters}}{(8 - 0)\text{seconds}} = \frac{3.0\text{meters}}{8\text{seconds}} = 0.375\text{meters/sec}$$

**Example**

Suppose you are doing an AP Chemistry lab and needed to calculate the rate of the decomposition of  $\text{NO}_2$  from 60 to 120 seconds:



Time (seconds)	$[\text{NO}_2]$	$[\text{NO}]$	$[\text{O}_2]$
0	0.0150	0	0
60	0.0085	0.0027	0.0018
120	0.0071	0.0041	0.0024

$$\text{Rate} = \frac{\Delta A}{\Delta t} = \frac{\Delta[\text{NO}_2]}{\Delta t}$$

$$\text{Rate} = \frac{0.0071\text{M} - 0.0085\text{M}}{120\text{sec} - 60\text{sec}}$$

$$\text{Rate} = \frac{-0.0014\text{M}}{60\text{sec}} = -2.33 \times 10^{-5} \text{M/sec}$$

Note that the negative sign indicates that the  $\text{NO}_2$  is being consumed in the reaction. We could also calculate the rate using the one of the products:

$$\text{Rate} = \frac{\Delta A}{\Delta t} = \frac{\Delta[\text{NO}]}{\Delta t}$$

$$\text{Rate} = \frac{0.0041\text{M} - 0.0027\text{M}}{120\text{sec} - 60\text{sec}}$$

$$\text{Rate} = \frac{0.0014\text{M}}{60\text{sec}} = 2.33 \times 10^{-5} \text{M/sec}$$

Notice that the rate of product formation is the same as the rate of consumption of the reactant. The rate is positive in magnitude because product is being formed.

The following tutorials can help you review how to do rate calculations:



[Khan Academy: Intro to rates](#)



[Khan Academy: Introduction to average rate of change](#)

## Linear Relationships and Curve Fitting

### Graphing Data as a Straight Line

When you plot data on  $x$ - $y$  axes, a straight line is the simplest relationship that data might have. Graphing data points as a straight line is useful because you can easily see where data points belong on the line.

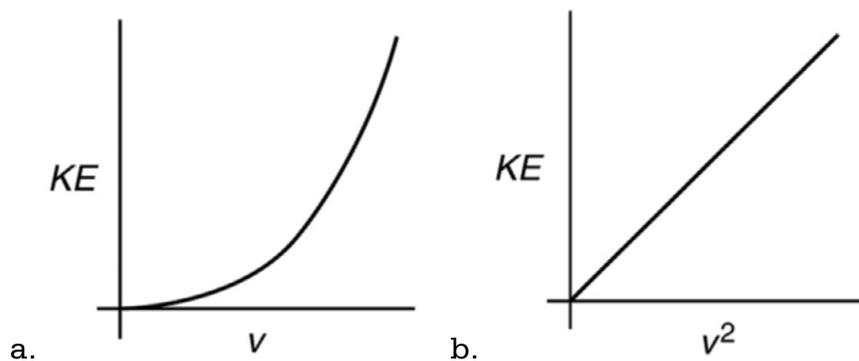
You can represent data as a straight line on a graph as long as you can identify its slope ( $m$ ) and its  $y$ -intercept ( $b$ ) in a linear equation:  $y = mx + b$ . The slope is a measure of how  $y$  varies with changes in  $x$ ,  $m = \Delta y / \Delta x$ . The  $y$ -intercept is where the line crosses the  $y$ -axis (where  $x = 0$ ).

### Linearizing Data

Even if the data you measure do not have an apparent linear relationship, you may be able to represent the data as a straight line by revising the form of the variables in your graph. One method is to transform the equation to represent the relationship so that it has the linear form of  $y = mx + b$  by substitution. For powers of  $x$ , the data would be in the form  $y = Ax^c + b$ . To linearize this data, substitute  $x^c$  for the  $x$  in the linear equation. Then you can plot  $y$  vs.  $x^c$  as a linear graph. For example, graphing kinetic energy,  $KE$ , and velocity,  $v$ , for the function

$KE = \frac{1}{2}mv^2$ , yields a parabola, as shown in figure 2.1a. But if we set the horizontal axis variable

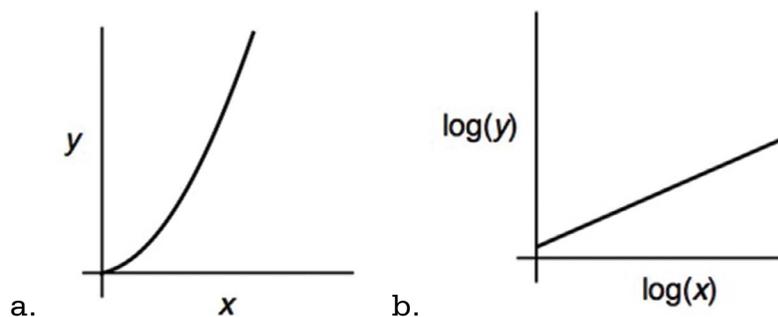
equal to  $v^2$  instead, the graph is linear, as shown in figure 2.1b, and the slope is equal to  $1/2m$  [Note that “ $1/2$ ” should be a built-up fraction, with  $m$  setting next to it.].



**Figure 2.1:** Changing the Variable on the  $x$ -axis to Produce a Linear Graph

If the data is exponential, as in  $y = Ae^{bx}$ , or is a power of  $x$ , as in  $y = ax^n$ , taking the log of both sides of the equation will linearize them. For exponential data, the equation you obtain is  $\ln(y) = \ln(A) + bx$ . The data will approximate a line with  $y$ -intercept  $\ln(A)$  and slope  $b$ .

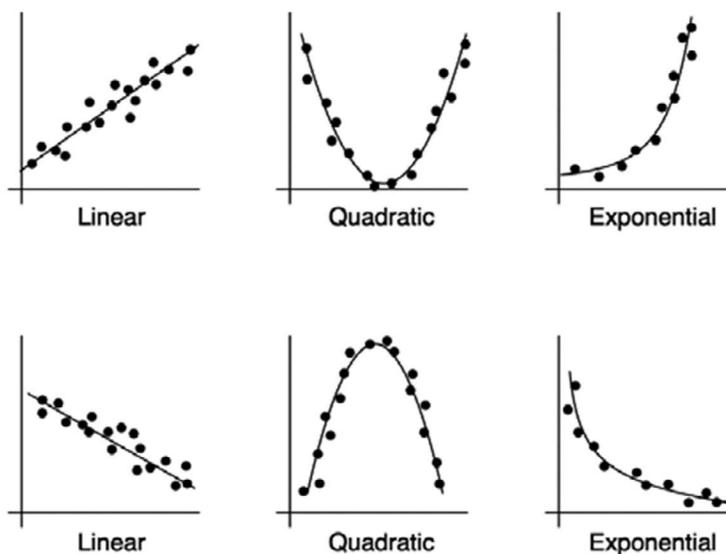
Similarly, for an equation with a power of  $x$ , taking the log of both sides of  $y = ax^n$  results in  $\log(y) = \log(a) + n\log(x)$ . If you plot  $\log(y)$  versus  $\log(x)$ , the data will approximate a line with  $y$ -intercept  $\log(a)$  and slope  $n$ , as shown in figures 2.2a and 2.2b.



**Figure 2.2:** Linear Graphs of Equations with a Power of  $x$

## Curve Fitting

A useful way to analyze data is to determine whether it corresponds to a certain mathematical model. A mathematical relationship or function will allow you to make a prediction if you know the function and an initial condition. The first step is to plot the points and see if they follow a recognizable trend, such as a linear, quadratic, or exponential function. Figure 2.3 shows examples of each of these types.



**Figure 2.3:** Common Mathematical Models

The general equation of a **linear function** is  $y = mx + b$ , as noted above, in which  $m$  is slope and  $b$  is the  $y$ -intercept. For example, a linear function in physics is the time dependence of the velocity of an object undergoing constant acceleration,  $v = v_0 + at$ , where the acceleration,  $a$ , is the slope and the initial velocity,  $v_0$ , is the  $y$ -intercept. An example of a linear function in biology is the amount of oxygen consumption by an endotherm over time at a constant temperature. In chemistry, an example of a linear function is the relationship between the concentration of a solution and the amount of light that is transmitted through the solution.

The general equation of a **quadratic function** is  $y = ax^2 + bx + c$ , where  $a$ ,  $b$ , and  $c$  are constants.

An example of a quadratic function in physics is the potential energy of a spring,  $U = \frac{1}{2}kx^2$ , where  $x$  is the distance the spring is stretched from equilibrium,  $k$  is the spring constant, and in this case the constants  $b$  and  $c$  are zero. Another example of a quadratic function is the position as a function of time for a constantly accelerating object,  $x = \frac{1}{2}at^2 + v_0t + x_0$ , where  $a$  is acceleration,  $v_0$  is initial velocity, and  $x_0$  is initial position.

The general equation of an **exponential function** is  $y = Ae^{bx}$ , where  $A$  and  $b$  are arbitrary constants. An example of the exponential function in physics is the number of radioactive particles left after a certain time of radioactive decay:  $N = N_0e^{-\lambda t}$ , where  $N_0$  is the original number of particles, and  $\lambda$  is the decay rate. Population growth is an example of an exponential function in biology and environmental science (see the section on population growth later in this chapter).

If the pattern of the data is clearly linear, or if you can plot the data using linearization, you can use a straightedge to draw a **best-fit line** that has approximately the same number of data points above and below the line. You can then determine an equation of the line by identifying the slope and  $y$ -intercept of the best-fit line.

If a more exact equation is desired, or if the data do not clearly follow a linear pattern, you can use a graphing calculator or a computer to fit the data to a mathematical model. In this case, you input the data and choose the model that you think will best fit the data. This is called **regression analysis**. Regression analysis is a common curve-fitting procedure. An analysis using this procedure provides parameters for the equation you have chosen for the fit, as well as parameters that describe how well the data fit the model. Figure 2.4 shows the same data using a linear model and a quadratic model. The value  $r^2$  is the **coefficient of determination**. It indicates how well the model fits the data. A value closer to 1 indicates a better fit. In the examples in figure 2.4, both models are a good fit for the data, but the  $r^2$  values show that the quadratic model is better because 0.9826 is closer to 1 than 0.95492 is.

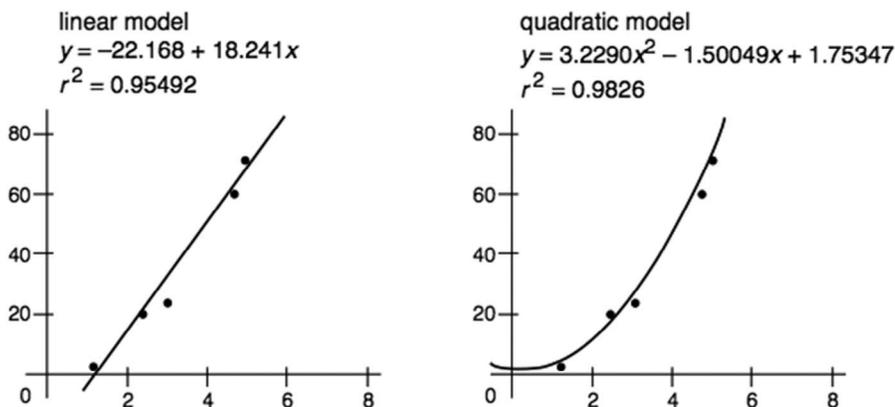


Figure 2.4: Different Models Have Different  $r^2$  Values

For more detailed information on linear functions, watch the following tutorial:



[Khan Academy: Comparing linear functions word problem](#)

## Descriptive Statistics

### Mean, Standard Deviation, and Standard Error

You can describe the uncertainty in data by calculating the mean and the standard deviation. The **mean** of a set of data is the sum of all the measurement values divided by the number of measurements. If your data is a sample of a population, then the mean you calculate is an estimate of the mean of a population. The mean,  $\bar{x}$ , is determined using the formula

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + x_3 + \dots}{n}$$

where  $x_1$ ,  $x_2$ , etc. are the measurement values and  $n$  is the number of measurements.

**Standard deviation** is a measure of how spread out the data values are. If your measurements have similar values, then the standard deviation is small: each value is close to the mean. If your measurements have a wide range of values, then the standard deviation is high: some values may be close to the mean, but others are far from it. In general, if you make a large number of measurements, then the majority of them are within one standard deviation above or below the mean. (See the section on confidence intervals for a graph of the standard deviation ranges later in this chapter.)

Since standard deviations are a measure of uncertainty, they should be standard using only one significant digit. Standard deviation is commonly represented by the letter  $s$ . You calculate sample standard deviation using this formula:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

When you make multiple measurements of a quantity, the **standard error** (SE) of the data set is an estimate of the precision to which you know the mean of the quantity. The standard error is related to how spread out the data is, but it also includes the fact that when you measure a quantity many times and take the average of your measurements, you get a more precise value than if you only measure the quantity a few times. You calculate standard error using this formula:

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

Because the number of measurements,  $n$ , is in the denominator, the more measurements you take, the smaller the standard error.

**Example**

Suppose you measure the following values for the temperature of a substance:

Trial	1	2	3	4
Temperature (°C)	20.5	22.0	19.3	23.0

The mean of the data is

$$\bar{x} = \sum_{i=1}^4 \frac{x_i}{4} = \frac{20.5 + 22.0 + 19.3 + 23.0}{4} = 21.2^\circ\text{C}$$

The standard deviation of the data is

$$\begin{aligned} s &= \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^4 (x_i - \bar{x})^2}{4-1}} \\ &= \sqrt{\frac{(20.5 - 21.2)^2 + (22.0 - 21.2)^2 + (19.3 - 21.2)^2 + (23.0 - 21.2)^2}{3}} \\ &= 1.63 = 2 \text{ (rounded to one significant digit)} \end{aligned}$$

The SE is

$$SE = \frac{s}{\sqrt{n}} = \frac{1.63}{\sqrt{4}} = 0.8 \text{ (rounded to one significant digit)}$$

Using one standard deviation, we would report the temperature of the substance as  $21.2 \pm 2^\circ\text{C}$ , meaning the typical temperature is in a range that is  $2^\circ$  above or  $2^\circ$  below the mean temperature. Since we only have a few data values, a standard deviation of  $2^\circ\text{C}$  shows that most of the data values were close to the mean. However, if we had taken a large number of measurements the standard deviation would show that the majority (specifically, 68%; see the Confidence Intervals section later in this chapter) of the data values were between  $19.2^\circ\text{C}$  and  $23.2^\circ\text{C}$ . Alternatively, the data could be reported using the standard error as  $21.2 \pm 0.8^\circ\text{C}$ . This tells us how our data compare to the true population mean with 95% confidence. In other words, because we took four measurements we have 95% confidence that the average temperature is within  $0.8^\circ\text{C}$  of  $21.2^\circ\text{C}$ . The standard error tells us how confident we are in our determination of the mean, while the standard deviation tells us how far we expect any individual measurement to be from the mean.

A graph of your data showing the statistics can clearly summarize the data in a way that is easy to understand and interpret.

**Example**

Suppose you conducted an investigation to determine if English ivy leaves in a shady area have a greater width than English ivy leaves in a sunny area. Table 2.3 shows the raw data from your experiment.

Table 2.3: 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

Table 2.3: Leaf Measurement Data (*continued*)

Shady Leaves (in cm)	Sunny Leaves (in cm)
9.9	7.9
9.9	8.0
10.4	8.2

The statistics from each experimental group can be calculated and shown in a table such as table 2.4.

Table 2.4: Descriptive Statistics

	Shady Leaves	Sunny Leaves
Mean	7.43	5.88
Standard Deviation	1.63	1.32
<i>N</i>	30	30
Standard Error	0.30	0.24

Using this information, you can graph your data to visually compare the means of the two groups of leaves:

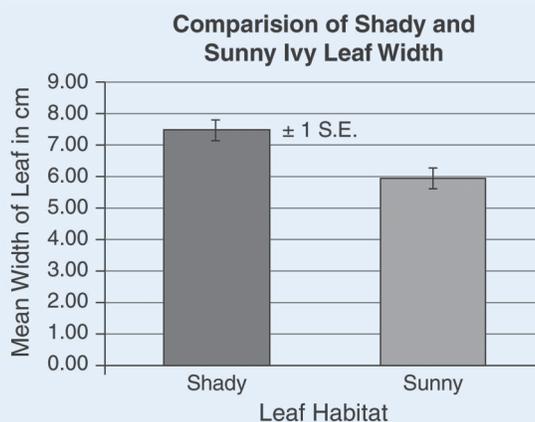


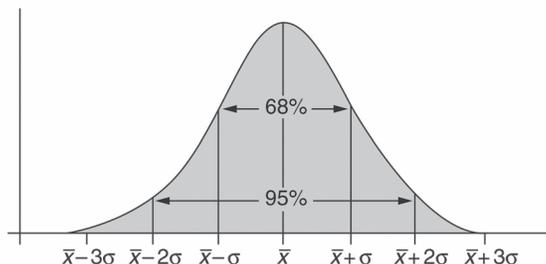
Figure 2.5: Comparison of Shady and Sunny Ivy Leaf Width

We see that the mean width of leaves grown in a shady environment is greater than the mean width of the leaves grown in a sunny environment. We also see that the error bars ( $\pm 1$  SE) for the two means do not overlap. This supports our claim that the two populations are different, in other words, that English ivy leaves grown in a shady environment have a greater width than English ivy leaves grown in a sunny environment.

## Confidence Intervals

A **confidence interval** is a range of values that the true value of the population has a probability of being within. If you measure a single quantity such as the mass of a certain isotope multiple times, you would expect a small standard deviation compared to the mean: the confidence intervals would be narrow. A wide confidence interval in this case would indicate the possibility of random errors in your measurements.

Confidence intervals can be presented in different ways. Figure 2.6 illustrates a commonly used method.



**Figure 2.6:** Confidence Intervals for a Normal Distribution

This method applies only to data that has a normal (bell-shaped) distribution. The mean lies at the peak of the distribution. Confidence intervals on either side of the peak describe multiples of the standard deviation from the mean. The percentage associated with each confidence interval (68%, 95%) has been determined by calculating the area under the curve.

A wide variety of data types in various subjects follow a **normal distribution** (i.e., a bell curve). In science, normal distributions apply to repeated measurements of a single value, such as multiple measurements of fluorescence decay time. A normal distribution is not appropriate when more than one central value is expected, or when only a few measurements are made.

**If you need more information, the following tutorials can help to further explain these concepts:**



[Khan Academy: Measures of spread: range, variance & standard deviation](#)



[Khan Academy: Standard error of the mean](#)



[Khan Academy: Confidence interval 1](#)

## Accuracy, Precision, and Experimental Error

Communication of data is an important aspect of every experiment. You should strive to analyze and present data that is as accurate as possible. Keep in mind that in the laboratory neither the measuring instrument nor the measuring procedure is ever perfect. Every experiment is subject to experimental error. Data reports should describe the experimental error for all measured values.

**Experimental error** affects the accuracy and precision of data.

- **Accuracy:** how close a measurement is to a known or accepted value. For example, suppose the mass of a sample is known to be 5.85 g. A measurement of 5.81 g would be more accurate than a measurement of 6.05 g because 5.81 g is closer to actual value of the measurement.
- **Precision:** how close several measurements are to each other. The closer measured values are to each other, the higher their precision.

Measurements can be precise even if they are not accurate. Consider again a sample with a known mass of 5.85 g. Suppose several students each measure the sample's mass, and all of the measurements are close to 8.5 g. The measurements are precise because they are close to each other, but none of the measurements are accurate because they are all far from the known mass of the sample.

**Systematic errors** are errors that occur every time you make a certain measurement.

- They result in measurements that can be inaccurate or incorrect by making measurements that are consistently either higher or lower than they would be if there were no systematic errors.
- Examples include errors due to the calibration of instruments and errors due to faulty procedures or assumptions; for example, using a balance that is not correctly calibrated.

**Random errors** are errors that cannot be predicted.

- This includes errors of judgment in reading a meter or a scale and errors due to fluctuating experimental conditions.
- If the random errors in an experiment are small, the experiment is said to be *precise*.
- For example, when having numerous groups of students making temperature measurements of a classroom at the same time, they will have random variations due to local variation and instrument fluctuation.

## CHAPTER 3

# Quantitative Skills in AP Biology

This chapter focuses on some of the quantitative skills that are important in your AP Biology course. These are not all of the skills that you will learn, practice, and apply during the year, but these are the skills you will most likely encounter as part of your laboratory investigations or classroom experiences, and potentially on the AP Biology exam.

## Surface Area and Volume

Cells are small because a high surface area to volume ratio allows them to function more efficiently. As a cell increases in size, its volume increases faster than its surface area, which does not allow it to transport substances across the membrane fast enough to get rid of what is unnecessary or obtain substances that are needed. The transport of substances across the membrane is largely driven by diffusion, which puts the size limit on the cell. To accommodate this, cells have adaptations to increase the surface area to volume ratio, such as folding of the cell membrane (e.g., villi in the small intestine).

### Example

To calculate the surface area to volume ratio for a spherical cell with a radius of  $5\ \mu\text{m}$ , first we calculate the surface area:

$$SA = 4\pi r^2 = 4\pi(5\ \mu\text{m})^2 = 100\pi\ \mu\text{m}^2$$

Next we calculate the volume:

$$V = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(5\ \mu\text{m})^3 = \frac{4}{3}\pi(125\ \mu\text{m}^3) = 166.67\pi\ \mu\text{m}^3$$

Now that we know the surface area and the volume, we can calculate the ratio:

$$\frac{SA}{V} = \frac{100\pi\ \mu\text{m}^2}{166.67\pi\ \mu\text{m}^3} = 0.6\ \mu\text{m}^{-1}$$

Alternately, we can simplify the formula first and then calculate the ratio:

$$\frac{SA}{V} = \frac{4\pi r^2}{\frac{4}{3}\pi r^3} = \frac{12\pi r^2}{4\pi r^3} = \frac{3}{r} = \frac{3}{5\ \mu\text{m}} = 0.6\ \mu\text{m}^{-1}$$

**Example**

To calculate the surface area to volume ratio of a cuboidal cell where each side is  $5\mu\text{m}$ , again, let's first calculate the surface area:

$$A = 6a^2 \quad (\text{where } a = \text{the length of one side of the cube})$$

$$A = 6(5)^2 = 150\mu\text{m}^2$$

Then we'll calculate the volume:

$$V = lwh = 5\mu\text{m} \times 5\mu\text{m} \times 5\mu\text{m} = 125\mu\text{m}^3$$

Now let's calculate the surface area to volume ratio:

$$\frac{SA}{V} = \frac{150\mu\text{m}^2}{125\mu\text{m}^3} = 1.2\mu\text{m}^{-1}$$

The spherical cell in these examples is smaller than the cuboidal cell and has a higher surface area to volume ratio, making it more efficient at diffusion than the cuboidal cell.

If you need more information, the following tutorial can help to further explain these concepts:



[Khan Academy: Cell size](#)

## Water Potential

**Water potential** measures the tendency of water to move by osmosis from one area to another, and is calculated from two major components, **pressure potential** and **solute potential**. The equation for water potential is

$$\Psi = \Psi_s + \Psi_p$$

where  $\Psi$  is the water potential,  $\Psi_s$  is the solute potential, and  $\Psi_p$  is the pressure potential.

Water will always move from an area of high water potential to an area of low water potential. The value of water potential can be positive, zero, or negative.

The water potential of pure water in an open beaker is zero ( $\Psi = 0$ ) because both the solute potential and pressure potential are equal to zero ( $\Psi_s = 0$  and  $\Psi_p = 0$ ). Increasing the amount of solute in water will increase solute potential, which in turn will decrease water potential. An increase in pressure potential (pushing) will increase water potential, while a decrease in pressure potential (pulling) will decrease water potential.

**Example**

Suppose you wanted to know whether water would go from an area where  $\Psi = -4.2$  to an area where  $\Psi = 0$ , or vice versa.

Because water flows from areas of high water potential to areas of low water potential, water would go from an area where  $\Psi = 0$  to an area where  $\Psi = -4.2$ .

In an open system, the pressure potential is zero, so the water potential of a solution can be calculated from the solute potential,  $\Psi_s$ , using the following formula:

$$\Psi_s = -iCRT$$

where  $i$  is the ionization constant,  $C$  is the molar concentration of the solution,  $R$  is the pressure constant ( $R = 0.0831 \text{ liter bars / mole K}$ ), and  $T$  is the temperature in kelvins.

**Example**

Suppose you wanted to know what the water potential is of a 1.5M sucrose solution at 22°C under standard atmospheric conditions. (Note: The ionization constant for sucrose is 1.0 because it does not ionize in water.)

First you would need to calculate the solute potential:

$$\begin{aligned}\Psi_s &= -iCRT \\ &= -(1.0)(1.5 \frac{\text{mol}}{\text{L}})(0.0831 \frac{\text{L} \cdot \text{bars}}{\text{mol} \cdot \text{K}})(295 \text{ K}) \\ &= -36.77 \text{ bars}\end{aligned}$$

Since the pressure potential is zero, the water potential is equal to the solute potential. Therefore,  $\Psi = -36.77 \text{ bars}$ .

If you need more information, the following tutorials can help to further explain these concepts:



[Khan Academy: Osmosis](#)



[Bozeman Science: Water potential](#)

## Hypothesis Testing

Hypothesis testing refers to procedures used by scientists to reject or fail to reject statistical hypotheses. It helps us attend to uncertainty and deal with random error in data collected during an investigation. There are two types of statistical hypotheses:

- **Null hypothesis ( $H_0$ ):** the hypothesis that there is no difference between two groups of data in an investigation and that the experimental observations are the result of chance

- **Alternative hypothesis ( $H_A$ ):** one of several hypotheses that experimental observations are the result of some nonrandom cause

## Chi-Square Goodness of Fit Test

The **chi-square test** is a statistical method that is used to determine if there is a significant relationship between two groups of data: observed values are compared to expected (or theoretical) values to determine if any variance from the expected data could be due to chance. It is called a goodness of fit test because we are checking to see if our results match up to a known or theoretical outcome of proportions for multiple categories. The chi-square test tests the null hypothesis. The null hypothesis is the claim against which we are looking for evidence in an investigation, specifically that the population proportions are what we would expect given random chance. For example, if we were rolling a standard six-sided die, our null hypothesis would be that the proportion of 1's, 2's, 3's, 4's, 5's, and 6's would all be 1/6. The alternative hypothesis would be that at least one of these proportions is not equal to 1/6.

The formula for chi-square is

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

where  $o$  is the value of the observed data, and  $e$  is the value of the expected data. Sigma ( $\Sigma$ ) is an indication that the repeated calculations that follow it should be added together.

The chi-square test is used for count data (i.e., data that is not measured) and is often used to analyze the results of genetic crosses.

### Example

If you crossed two pea plants that were heterozygous for yellow seed color ( $Y$  = yellow,  $y$  = green), you would expect the phenotypic ratio of the offspring in the next generation to be 3:1. This means that if you collected 100 seeds, you would expect 75 yellow seeds and 25 green seeds. This is our null hypothesis: that there is no difference between the phenotypic ratios in our cross and what is expected by Mendelian theory.

If you actually collected 69 yellow seeds and 31 green seeds, is that close enough to what is expected that you can say your results are consistent with the theory—that the deviation in your results is likely due to chance? What if you collected 65 yellow and 35 green seeds?

We can calculate the chi-square value for these data to answer the question:

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

Categories	$o$	$e$	$(o-e)^2/e$
Yellow	69	75	0.48
Green	31	25	1.44

$$\chi^2 = 0.48 + 1.44 = 1.92$$

Now that we know what our chi-square value is, we can compare it to the critical value found in table 3.1.

Table 3.1: Critical Values of the Chi-Square Distribution

Probability ( <i>p</i> )	Degrees of Freedom				
	1	2	3	4	5
0.05	3.84	5.99	7.82	9.49	11.1
0.01	6.64	9.21	11.3	13.2	15.1
0.001	10.8	13.8	16.3	18.5	20.5

- If the chi-square value is lower than the critical value, we will fail to reject our null hypothesis: the data we observed fits well with the data we expected.
- If the chi-square value is higher than the critical value, we will reject our null hypothesis: our observed data does not fit well with our expected data.
- The probability, or *p* value, is used to tell us how often results like ours could occur under the null hypothesis if only chance is at play.

In order to use the critical value table, we need to know our degrees of freedom, which is the number of categories minus one. For our example, we have 1 degree of freedom because we have two categories. Using the table, we find that the critical value for 1 degree of freedom at a *p*-value of 0.05 is 3.84. Our chi-square value of 1.92 is less than this critical value, so we fail to reject our null hypothesis. This means that our observed data for this cross is within a 3:1 phenotypic ratio.

We can follow the same logic for the second cross:

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

Categories	<i>o</i>	<i>e</i>	$(o-e)^2/e$
Yellow	65	75	1.33
Green	35	25	4

$$\chi^2 = 1.33 + 4 = 5.33$$

Our chi-square value of 5.33 is higher than the critical value of 3.84 at *p* = 0.05 and our null hypothesis is rejected. This means that the offspring of the cross in this experiment did not yield a 3:1 phenotypic ratio. This does not mean that if we were to do this cross again we would get the same result: we might get a different result. If that were to happen we would need to repeat our experiment several times so that we have an aggregate of data to analyze. The more data we have, the more confident we can be in our results.

If you need more information, the following tutorial can help to further explain this concept:



[Khan Academy: Chi-square distribution introduction](#)

## Mathematical Modeling

Mathematical models can be used to investigate the relationships that occur in nature. The Hardy–Weinberg theorem helps us to understand the relationship between allele frequencies in populations of organisms and evolutionary change, specifically how populations can change over time and what might happen to a population in the future.

### Hardy–Weinberg Equilibrium

The **Hardy–Weinberg theorem** states that the allele frequencies of a gene (at a specific locus) in a population will stay the same from one generation to the next as long as the following requirements are met:

- No mutations occur.
- No movement into or out of the population occurs.
- The population is large.
- Mating is random.
- No natural selection occurs.

This means the following factors can disrupt Hardy–Weinberg equilibrium:

- Mutations
- Gene flow
- Genetic drift
- Nonrandom mating
- Natural selection

The changes in the allele frequencies of a population over time can be studied mathematically. Let's start with a gene pool of  $A$ 's and  $a$ 's, where  $A$  is the dominant allele and  $a$  is the recessive allele.

We can define  $p$  as the frequency of  $A$  and  $q$  as the frequency of  $a$ . Since the total frequencies of all alleles must equal 1, then  $p + q = 1$ . This means, for example, that if 60% of the alleles in a population are  $A$ , then the frequency of  $A$  is 0.6 and the frequency of  $a$  is 0.4.

The following are all the possible combinations of  $A$  and  $a$  in the population:  $AA$ ,  $Aa$ ,  $aA$ ,  $aa$ . We can use a Punnett square to determine the frequencies of these combinations (i.e., the genotypes in the next generation):

	$A$ ( $p$ )	$a$ ( $q$ )
$A$ ( $p$ )	$AA$ ( $p^2$ )	$Aa$ ( $pq$ )
$a$ ( $q$ )	$aA$ ( $qp$ )	$aa$ ( $q^2$ )

The frequencies of the genotypes in the population must also equal 1. The equation that describes this is

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

This can be simplified to

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

where  $p^2$  is the frequency of the dominant genotype,  $q^2$  is the frequency of the recessive genotype, and  $2pq$  is the frequency of the heterozygous genotype. This is the **Hardy–Weinberg equation**.

Using the previous example, if 60% of the alleles in a population are  $A$ , we can calculate all of the genotypic frequencies:

$$p = \text{frequency of } A = 0.6$$

$$q = \text{frequency of } a = 0.4$$

$$p^2 = \text{frequency of } AA = (0.6)^2 = 0.36$$

$$q^2 = \text{frequency of } aa = (0.4)^2 = 0.16$$

$$2pq = \text{frequency of } Aa = 2(0.6)(0.4) = 0.48$$

### Example

If we assume that a population is in Hardy–Weinberg equilibrium for a particular locus, we can apply the Hardy–Weinberg equation to other questions, such as, if an autosomal recessive disease affects three in 100 people of a population, what is the frequency of carriers in the population?

We know that three out of 100 people are affected by an autosomal recessive condition—this is the frequency of  $aa$ , which is  $q^2$ :

$$q^2 = \frac{3}{100} = 0.03$$

If  $q^2$  is 0.03, then we can calculate the frequency of the recessive allele:

$$q = \sqrt{0.03} = 0.173$$

Since  $p + q = 1$ , we can now use this information to calculate the frequency of the dominant allele:

$$p + q = 1$$

$$p + 0.173 = 1$$

$$p = 1 - 0.173$$

$$p = 0.827$$

Since the disease is autosomal recessive, this means that the heterozygous individuals in the population are carriers of the disease. Now that we know the frequency of each allele in the population, we can calculate the frequency of the carriers:

$$\begin{aligned} \text{frequency of } Aa &= 2pq \\ &= 2(0.827)(0.173) \\ &= 0.286 \end{aligned}$$

If you need more information, the following tutorial can help to further explain this concept:



[Khan Academy: Allele frequency](#)

## Population Growth

Populations can experience unlimited growth when resources are abundant, but this is rarely the case in nature.

The change in the size of a population over a given amount of time mostly depends on the number of births and the number of deaths that occur within it. This can be written mathematically as

$$\frac{\Delta N}{\Delta t} = B - D$$

where  $\Delta N$  is the change in the population,  $\Delta t$  is the change in time,  $B$  is the number of births, and  $D$  is the number of deaths.

In a given population, the per capita rate of increase,  $r$ , can be written as

$$r = \frac{B - D}{N}$$

The equation for the change in population size can now be written as

$$\frac{\Delta N}{\Delta t} = rN$$

Two basic models of population growth are the exponential growth and the logistic growth model:

- **Exponential growth model:** A population increases in size when  $r$  is constant and greater than zero:

$$\frac{\Delta N}{\Delta t} = r_{\max} N$$

where  $r_{\max}$  is the maximum per capita rate of increase and  $N$  is the size of the population. In other words, the population (usually a small one) grows at a fixed rate.

- **Logistic growth model:** A population increases in size until  $r$  approaches zero as the size of the population reaches its carrying capacity:

$$\frac{\Delta N}{\Delta t} = r_{\max} N \frac{(K - N)}{K}$$

where  $K$  is the carrying capacity of the population. The **carrying capacity** is the maximum population size that can be supported by a particular environment at a certain point in time without disruption to the habitat.

**Exponential growth example**

Suppose a population of 100 bacteria is experiencing exponential growth: The current birth rate is 22 bacteria per day and the current death rate is 15 individuals per day. We want to calculate the growth of this population over 30 days.

First, let's calculate the per capita rate of increase:

$$r = \frac{B-D}{N}$$

$$r = \frac{22-15}{100}$$

$$r = 0.07$$

Now, let's calculate the growth rate for day one:

$$\frac{\Delta N}{\Delta t} = rN$$

$$\frac{\Delta N}{\Delta t} = (0.07)(100)$$

$$\frac{\Delta N}{\Delta t} = 7$$

This means that 7 bacteria are added the first day. So, for the second day, we start with 107 bacteria. Using this same equation, we can calculate the size of the population for the next day:

$$\frac{\Delta N}{\Delta t} = rN$$

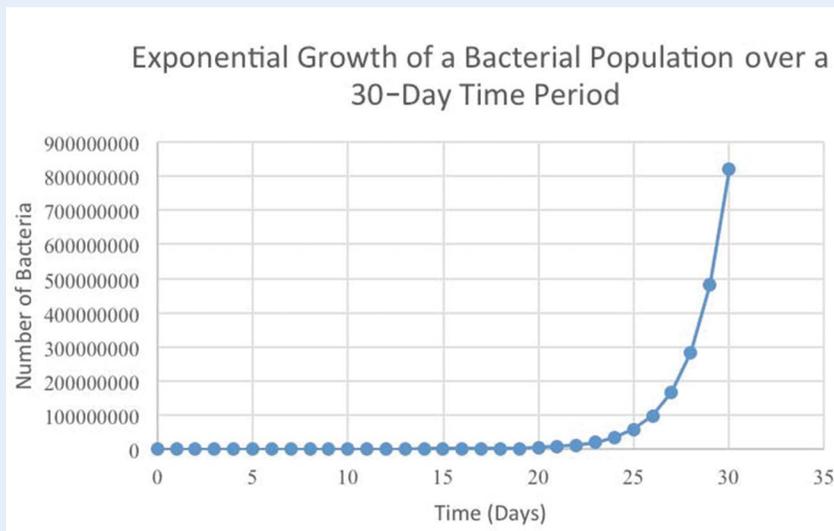
$$\frac{\Delta N}{\Delta t} = (0.07)(107)$$

$$\frac{\Delta N}{\Delta t} = 7.49$$

If we repeat this calculation until day 10, we will see the following for our population growth:

Day	Number of Bacteria
0	100
1	107
2	115
3	123
4	132
5	141
6	150
7	160
8	170
9	180
10	191

If we continued this for the next 30 days, the growth of our population would look like this:



**Figure 3.1:** Exponential Growth of a Bacterial Population

### Logistic growth example

Say that the same population of bacteria from the previous example is experiencing logistic growth at the same rate, but the carrying capacity of the environment is 2000 bacteria. Let's calculate the growth rate of the population under these circumstances:

$$\frac{\Delta N}{\Delta t} = r_{\max} N \frac{(K - N)}{K}$$

$$\frac{\Delta N}{\Delta t} = (0.7)(100) \left[ \frac{(2000 - 100)}{2000} \right]$$

$$\frac{\Delta N}{\Delta t} = (70)(0.95)$$

$$\frac{\Delta N}{\Delta t} = 66.5$$

The growth rate is slower under these circumstances due to the limited availability of resources.

If you need more information, the following tutorial can help to further explain these concepts:



[Khan Academy: Exponential and logistic growth in populations](#)

# Appendix A

## AP Biology Equations and Formulas

Statistical Analysis and Probability								
<b>Mean</b>			<b>Standard Deviation*</b>					
$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$			$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$					
<b>Standard Error of the Mean*</b>			<b>Chi-Square</b>					
$SE_{\bar{x}} = \frac{S}{\sqrt{n}}$			$\chi^2 = \sum \frac{(o-e)^2}{e}$					
<b>Chi-Square Table</b>								
<i>p</i>	Degrees of Freedom							
value	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
<p><b>Laws of Probability</b></p> <p>If A and B are mutually exclusive, then:  <math>P(A \text{ or } B) = P(A) + P(B)</math></p> <p>If A and B are independent, then:  <math>P(A \text{ and } B) = P(A) \times P(B)</math></p> <p><b>Hardy-Weinberg Equations</b></p> <p><math>p^2 + 2pq + q^2 = 1</math>     <math>p</math> = frequency of the dominant allele in a population</p> <p><math>p + q = 1</math>                 <math>q</math> = frequency of the recessive allele in a population</p>								
<b>Metric Prefixes</b>								
<b>Factor</b>			<b>Prefix</b>			<b>Symbol</b>		
10 <sup>9</sup>			giga			G		
10 <sup>6</sup>			mega			M		
10 <sup>3</sup>			kilo			k		
10 <sup>-2</sup>			centi			c		
10 <sup>-3</sup>			milli			m		
10 <sup>-6</sup>			micro			μ		
10 <sup>-9</sup>			nano			n		
10 <sup>-12</sup>			pico			p		
<p>Mode = value that occurs most frequently in a data set</p> <p>Median = middle value that separates the greater and lesser halves of a data set</p> <p>Mean = sum of all data points divided by number of data points</p> <p>Range = value obtained by subtracting the smallest observation (sample minimum) from the greatest (sample maximum)</p>								
<p>* For the purposes of the AP Exam, students will not be required to perform calculations using this equation; however, they must understand the underlying concepts and applications.</p>								

<p style="text-align: center;"><b>Rate and Growth</b></p> <p><b>Rate</b> <math>\frac{dY}{dt}</math></p> <p><b>Population Growth</b> <math>\frac{dN}{dt} = B - D</math></p> <p><b>Exponential Growth</b> <math>\frac{dN}{dt} = r_{\max} N</math></p> <p><b>Logistic Growth</b> <math>\frac{dN}{dt} = r_{\max} N \left( \frac{K - N}{K} \right)</math></p> <p><b>Temperature Coefficient <math>Q_{10}</math><sup>†</sup></b> <math>Q_{10} = \left( \frac{k_2}{k_1} \right)^{\frac{10}{T_2 - T_1}}</math></p> <p><b>Primary Productivity Calculation</b> <math>\frac{\text{mg O}_2}{\text{L}} \times \frac{0.698 \text{ mL}}{\text{mg}} = \frac{\text{mL O}_2}{\text{L}}</math> <math>\frac{\text{mL O}_2}{\text{L}} \times \frac{0.536 \text{ mg C fixed}}{\text{mL O}_2} = \frac{\text{mg C fixed}}{\text{L}}</math> (at standard temperature and pressure)</p>	<p><math>dY</math> = amount of change</p> <p><math>dt</math> = change in time</p> <p><math>B</math> = birth rate</p> <p><math>D</math> = death rate</p> <p><math>N</math> = population size</p> <p><math>K</math> = carrying capacity</p> <p><math>r_{\max}</math> = maximum per capita growth rate of population</p> <p><math>T_2</math> = higher temperature</p> <p><math>T_1</math> = lower temperature</p> <p><math>k_2</math> = reaction rate at <math>T_2</math></p> <p><math>k_1</math> = reaction rate at <math>T_1</math></p> <p><math>Q_{10}</math> = the factor by which the reaction rate increases when the temperature is raised by ten degrees</p>	<p><b>Water Potential (<math>\Psi</math>)</b> <math>\Psi = \Psi_p + \Psi_s</math></p> <p><math>\Psi_p</math> = pressure potential</p> <p><math>\Psi_s</math> = solute potential</p> <p>The water potential will be equal to the solute potential of a solution in an open container because the pressure potential of the solution in an open container is zero.</p> <p><b>The Solute Potential of a Solution</b> <math>\Psi_s = -iCRT</math></p> <p><math>i</math> = ionization constant (this is 1.0 for sucrose because sucrose does not ionize in water)</p> <p><math>C</math> = molar concentration</p> <p><math>R</math> = pressure constant (<math>R = 0.0831</math> liter bars/mole K)</p> <p><math>T</math> = temperature in Kelvin (<math>^{\circ}\text{C} + 273</math>)</p>
<p style="text-align: center;"><b>Surface Area and Volume</b></p> <p><b>Volume of a Sphere</b> <math>V = \frac{4}{3} \pi r^3</math></p> <p><b>Volume of a Rectangular Solid</b> <math>V = lwh</math></p> <p><b>Volume of a Right Cylinder</b> <math>V = \pi r^2 h</math></p> <p><b>Surface Area of a Sphere</b> <math>A = 4\pi r^2</math></p> <p><b>Surface Area of a Cube</b> <math>A = 6s^2</math></p> <p><b>Surface Area of a Rectangular Solid</b> <math>A = \sum</math> surface area of each side</p>	<p><math>r</math> = radius</p> <p><math>l</math> = length</p> <p><math>h</math> = height</p> <p><math>w</math> = width</p> <p><math>s</math> = length of one side of a cube</p> <p><math>A</math> = surface area</p> <p><math>V</math> = volume</p> <p><math>\Sigma</math> = sum of all</p>	<p><b>Dilution (used to create a dilute solution from a concentrated stock solution)</b> <math>C_i V_i = C_f V_f</math></p> <p><math>i</math> = initial (starting)      <math>C</math> = concentration of solute <math>f</math> = final (desired)      <math>V</math> = volume of solution</p> <p><b>Gibbs Free Energy</b> <math>\Delta G = \Delta H - T\Delta S</math> <math>\Delta G</math> = change in Gibbs free energy <math>\Delta S</math> = change in entropy <math>\Delta H</math> = change in enthalpy <math>T</math> = absolute temperature (in Kelvin)</p> <p><math>\text{pH}^* = -\log_{10} [\text{H}^+]</math></p>
<p>* For the purposes of the AP Exam, students will not be required to perform calculations using this equation; however, they must understand the underlying concepts and applications.</p>		
<p><sup>†</sup> For use with labs only (optional).</p>		

---

# Appendix B

## AP Calculator Policy

Calculators can be used on all or some parts of the AP Exams listed here (and on no others).

- Students should check the list of approved graphing calculators located at <https://apstudent.collegeboard.org/takingtheexam/exam-policies/calculator-policy>; bring only approved models.
- Students may bring two permitted calculators. They should bring calculators they are familiar with that are in good working order. **Calculators may not be shared.**
- Calculator memories do not need to be cleared before or after the exam.

---

Courses and Calculator Use			
Course	Type of calculator allowed	Exam Section I	Exam Section II
Biology	Four-function (with square root), scientific, or graphing calculator	Allowed	Allowed
Chemistry	A scientific or graphing calculator is recommended; a four-function calculator is allowed but not recommended.	Not Allowed	Allowed
Physics 1	Four-function, scientific or graphing calculator	Allowed	Allowed
Physics 2	Four-function, scientific or graphing calculator	Allowed	Allowed
Physics C: Mechanics	Four-function, scientific or graphing calculator	Allowed	Allowed
Physics C: Electricity & Magnetism	Four-function, scientific or graphing calculator	Allowed	Allowed

---

---

# Appendix C

## Helpful Links

“12 – Help for Students in Mental Math.” Chemistry & Cognition: Support For Cognitive-Based First-Year Chemistry. Accessed September 14, 2017. <http://chemreview.net/blog/?p=409>.

“Analyzing and Visualizing Data with Excel.” EdX. Accessed September 14, 2017. <https://www.edx.org/course/analyzing-visualizing-data-excel-microsoft-dat206x>.

“Averaging, Errors, and Uncertainty.” Department of Physics & Astronomy. University of Pennsylvania. Accessed September 14, 2017. [http://avntraining.hartrao.ac.za/images/Error\\_Analysis.pdf](http://avntraining.hartrao.ac.za/images/Error_Analysis.pdf).

“Excel 2013 Training.” Microsoft Office. Accessed September 14, 2017. <https://support.office.com/en-us/article/Excel-video-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb?ui=en-US&rs=en-US&ad=US>.

“Functions and Formulas.” Docs Editors Help. Google. Accessed September 14, 2017. [https://support.google.com/docs/topic/1361471?hl=en&ref\\_topic=2811806](https://support.google.com/docs/topic/1361471?hl=en&ref_topic=2811806).

“Excel Tutorial.” Department of Physics & Astronomy. University of Pennsylvania. Accessed September 14, 2017. <http://physics.unc.edu/files/2012/10/exceltutorial.pdf>.

“Sampling and Normal Distribution.” HHMI BioInteractive. Accessed September 14, 2017. <http://www.hhmi.org/biointeractive/sampling-and-normal-distribution>.

“Spreadsheet Data Analysis Tutorials.” HHMI BioInteractive. Accessed September 14, 2017. <http://www.hhmi.org/biointeractive/spreadsheet-data-analysis-tutorials>.

“Useful Excel Commands for Lab.” Department of Physics. Randolph College. Accessed September 14, 2017. <http://physics.randolphcollege.edu/lab/IntroLab/Reference/exchint.html>.