Finding Relationships – A Laboratory Graphing Exercise

Objectives

- To gain experience in using a commercial graphing program such as Microsoft Excel.
- To determine mathematical relationships between variables using graphical methods.
- To practice drawing scientific conclusions based on graphical data.

Introduction

In several laboratory investigations in CHM 111 and CHM 112, a primary objective will be to observe and quantify the mathematical relationship between two variables. An important tool for determining mathematical relationships in laboratory science makes use of graphical methods. For example, CHM 111 students use graphing to determine the mathematical relationship between the density of a salt water solution and the mass percent of salt in the solution. In another experiment, students may be asked to determine the relationship between the volume of a confined gas and the temperature of the gas. At other times, students may use graphical methods to document unusual or unexpected deviations in data sets. In this exercise, students will practice using a commercial graphing program to determine mathematical relationships between variables and draw conclusions based on collected graphical data.

The Parts of a Scientific Graph

A scientific graph should stand by itself. In other words, a reader should be able to see all the important information about the data from the graph. Consequently, scientific graphs must include:

1. Meaningful axes, with well-chosen grid lines, tick marks, and appropriate minimums and maximums
2. A descriptive caption or title
3. A clear legend if there is more than one data set on a single graph
4. A trendline and equation for the slope of the trendline ($y = mx + b$) ***if*** the mathematical relationship between the variables is to be used or reported.

Common Graph Types: Scatter Charts

Commercial graphing applications such as Excel allow users to create a wide variety of graph or chart "types". The **"Scatter"** chart option is usually the best graphing option if data contain both x and y values. ("Line" charts may be used if data contain numerical values only on the *y* axis, or if the data display trends over time; bar, pie, and area graphs are also used in science, but not in this course.) For this course we will almost exclusively prepare **Scatter** charts with data *markers*. Do not draw data *lines* in this course – always select the chart sub-type with data markers, *not* lines.

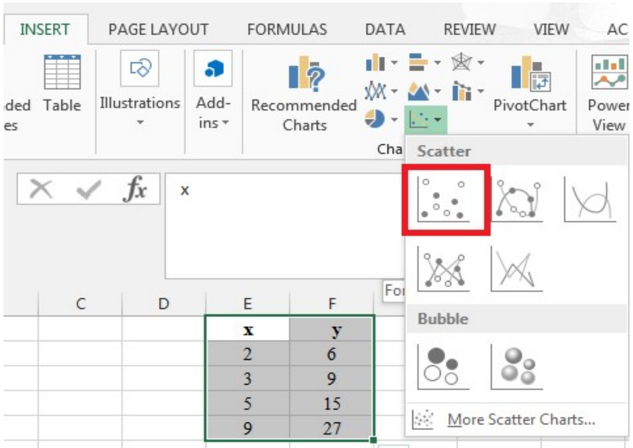
Drawing a Graph

Microsoft Excel is the graphing application that is supported and available on NVCC computers. Many personal computers and tablets have Excel installed, but any NVCC student can also obtain a free copy of Excel as part of the Microsoft Office 365 suite. Other commercial graphing products exist, but are not supported by the College. Note that there are ***many*** versions of Excel in common use – each version has slight variations, so precise instructions for one version may not apply to another version. You may have to experiment a little to produce your desired graph.

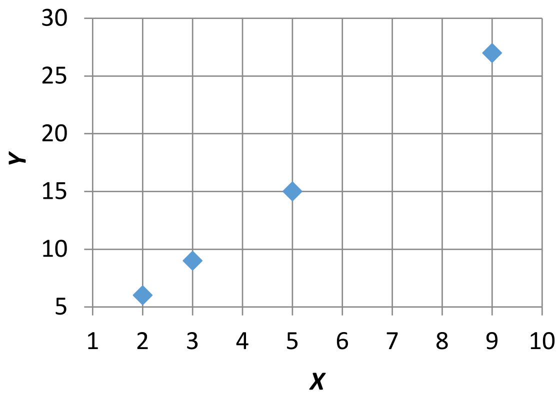
The following is the basic process for preparing a scientific graph – consider ***Example 1***: suppose you have four ordered pairs *x* (in yards) and *y* (in feet): (2, 6), (3, 9), (5, 15), (9, 27) and you want to determine the relationship between them:

x	y
2	6
3	9
5	15
9	27

The first logical step is to make a graph of *y* versus *x*. Input the data in columns in an Excel workbook, with the *x* data placed in a column to the left and adjacent to the y data (see below). Place headings at the top of each column, select the 2 data columns, and **insert a chart that is "scatter type" with data markers** (***not*** lines and markers).



The resultant rough graph for ***Example 1*** is shown below:



Since the shape of the plot appears to be linear it is probably a simple *direct* mathematical relationship, but this cannot be confirmed until the trendline is inserted and the correlation coefficient R^2 is evaluated (see below).

Drawing and Labeling Axes

When drawing bar charts, the *axis* traditionally starts at zero. This rule does **NOT** necessarily apply to scatter or line charts, as can be observed in the graph above. Minimum and maximum values for each axis on a scatter chart should be selected so that the data points are spread over most of the graph. To change a minimum or maximum axis value, right click on the axis and chose "Format Axis" at the bottom of the dialogue box. Set the minimum / maximum to an appropriate fixed value. Always set axis limits so that all data is visible, but there is *not* a lot of empty space. Other adjustments to each axis (grid lines, etc.) can also be made from the same dialogue box. **Add axis titles** to a graph by going to "Chart Tools", "Layout" and "Axis Titles".

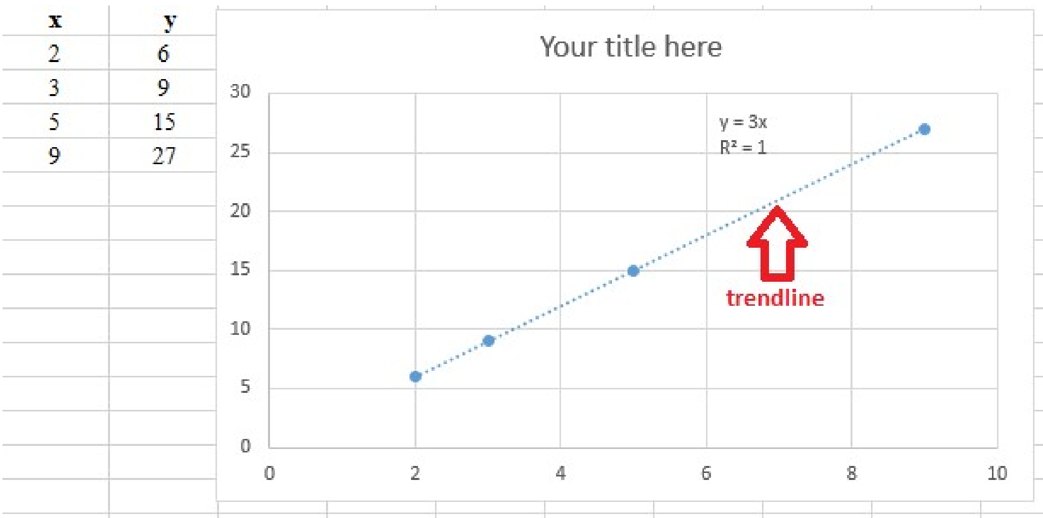
***How do you know which way to orient the axes*?** The *x*-axis should be the *independent variable*, the one that is controlled. The *y*-axis is the *dependent variable*, the variable whose value is to be determined or evaluated. Often, you will need to think about the scientific law or idea that the graph is intended to show, in order to determine what is being controlled (*x*) and what is to be determined (*y*). Remember to always include units for variables that have units.

Adding Trendlines

In lab, you will usually (but *not* always) be drawing graphs based on a scientific law. Consequently, the next step usually is to determine the mathematical relationship between the data points. There are two different types of lines that can be drawn on graphs; **data lines** and **trendlines**. Data lines "connect the dots" between data points. They can be useful *if* the data set is huge, but *you will almost never draw data lines in this course.* A **trendline** is a mathematical fit to the data, and in this course is usually based on scientific law. Linear trendlines are common, but exponential or logarithmic trendlines are also sometimes necessary, depending on the experiment. To **insert a trendline** on a graph, right click a data point and choose "Insert Trendline" from the dialogue box. Display the equation for the slope of the trendline and the R^2 value by selecting the appropriate checkboxes in the Trendline dialogue box. R and R^2 are correlation coefficients – they give a sense of whether the trendline is a good match to the data points. What makes an R or R^2 value "good" depends on the type of data. A value of 1 is a perfect match, and for analytical chemistry 0.99999 is a good R^2, but for some business analyses $R^2 = 0.8$ is considered an acceptable fit. Note that "**Forecast Forward**" and "**Forecast Backward**" can also be used if the trendline needs to extend beyond the data you are fitting.

Sometimes, graphical data do **not** follow a pattern or mathematical relationship that allows for insertion of a meaningful trendline. Even in such cases, conclusions and predictions drawn from the graphical data can be dramatic and scientifically significant. Some common examples include the variation in life expectancy throughout U.S. history and documented increases in worldwide CO_2 emissions with increasing population and industrialization.

An example of the "rough draft" graph from ***Example 1*** is shown below:

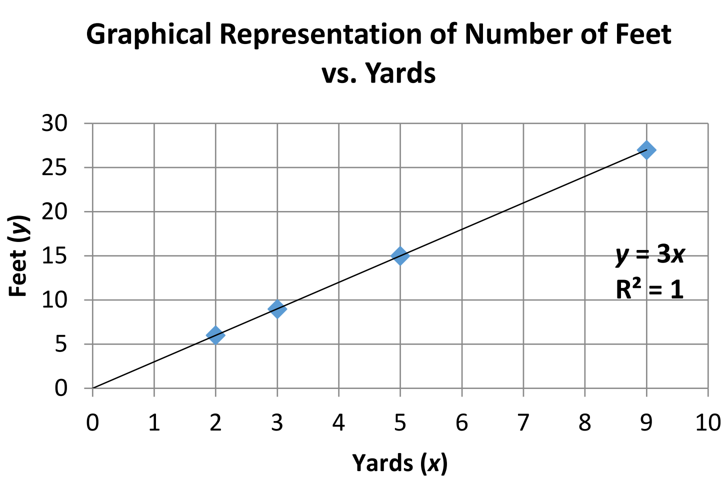


Graph Titles and Legends

All graphs need descriptions that explain the graph. Scientific writing usually uses captions, complete sentences that explain the relevant experimental details needed to understand the graph. For visual presentations however, scientists often use titles. Titles should NOT just repeat the axis labels – the reader can see the axes, so there's no point repeating them. Think about why you are drawing the graph and then describe it to the reader in the title. Go to "Chart Tools", "Layout", "Chart title", and change the title to a meaningful summary.

A graph should have a legend only if there are multiple data sets being graphed – it should not have a legend for one data set and a trendline. If there are multiple data sets, make sure that the data markers are distinct and that the legend can distinguish the data sets.

An example of the "finished" graph from ***Example 1*** with a backward forecast trendline is shown below:



As noted earlier, since the shape of the plot is a straight line that passes through the origin (0,0), it is a simple *direct* relationship, with the y-intercept = 0. The equation is written showing this relationship: $y = m \bullet x$. This is done by writing the variable from the vertical axis (dependent variable) on the left side of the equation, and then equating it to a proportionality constant, m, multiplied by x, the independent variable. The constant, m, can be determined either by displaying the slope of the graph or by solving the equation for m ($m = y/x$), and finding m for one of the ordered pairs. In this simple example, $m = 6/2 = 3$. If it is the correct proportionality constant, then you should get the same m value by dividing any of the y values by the corresponding x value. Consequently, the equation for the relationship is written as:

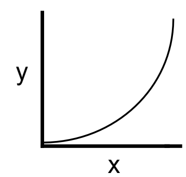
$y = 3 \bullet x$ (y varies directly with x)

A few examples of *non*-linear fit graphs are briefly considered below:

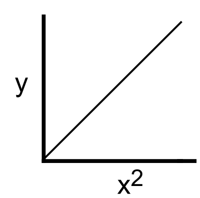
Example 2: Consider these ordered pairs:

x	y
1	2
2	8
3	18
4	32

If one plots y versus x, the graph appears as shown on the following page:



It appears that y increases as x increases, but since the graph is not a straight line, the increase is not proportional (direct). Rather, y varies *exponentially* with x. Thus y might vary with the *square* of x or the *cube* of x. The next logical plot would be y versus x^2. The resultant graph looks like this:



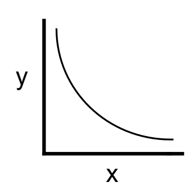
Since this plot is a straight line passing through the origin, y varies with the square of x, and the equation is: **$y = m \bullet x^2$**
Determine m either by displaying the slope of the graph or by dividing y by x^2 for any data pair: $m = y/x^2 = 8/(2)^2 = 2$
This value will be the same for any of the four ordered pairs, and yields the equation:

$y = 2 \bullet x^2$ (y varies directly with the square of x)

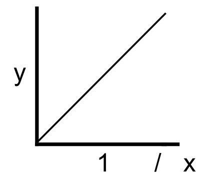
Example 3: Consider these ordered pairs:

x	y
2	24
3	16
4	12
8	6
12	4

A plot of y versus x gives a graph that looks like this:



A graph with this curve always suggests an *inverse* relationship. To confirm an inverse relationship, plot the *reciprocal* of one variable versus the other variable. In this case, y is plotted versus the reciprocal of x, or $1/x$. The graph looks like this:



Since this graph yields a straight line that passes through the origin (0,0), the relationship between x and y is inverse. Using the same method we used in Examples 1 and 2, the equation would be: $\boldsymbol{y = m(1/x)}$ **or** $\boldsymbol{y = m/x}$
To find m, display the equation for the slope or solve for m($m = y \cdot x$). Using the ordered pair (2,24): $m = 2 \cdot 24 = 48$

Thus the equation would be: $\boldsymbol{y = 48/x}$ **(*y* varies inversely with *x*)**

Note that in some cases, the best fit of the data might be an *inverse* square relationship, or even a logarithmic relationship.

Standard Curves

A standard curve (or *calibration curve*) is a graph relating a measured quantity (radioactivity, fluorescence, or optical density, for example) to the concentration of a substance of interest in "known" samples.

The data obtained from a set of known samples is used to make a standard curve, by plotting the independent variable on the *x*-axis, and dependent variable on the *y*-axis. If the data exhibit a linear relationship, the standard curve will be a straight line. The same experiment is then performed with samples of *unknown* concentration. To analyze the data, one can visually locate the measurement on the *y*-axis that corresponds to the unknown substance, and follow a line to intersect the standard curve. The corresponding value on the *x*-axis is the concentration of substance in the unknown sample.

Alternatively, the value on the y-axis and the equation for the best fit line can be used to solve for x mathematically. When a line in a two-variable system is defined by the equation $y = mx + b$, the y variable can be expressed in terms of a **constant** (b) and a **slope** (m) times the x variable. The constant is also referred to as the **intercept** (or more specifically the *y*-intercept), and the slope as the regression coefficient. Again, a correlation coefficient (R) or coefficient of determination (R^2) may be computed. The closer R is to ±1 the better the approximation. Typically an R-value of at least ±0.97 (or an R^2 value of 0.95 or higher) is needed for a good linear approximation in a scientific application.

In **Part A** of this exercise, students will learn how to create standard curves using Excel and then use them to determine the concentration level of unknown samples. The focus will be on linear functions and graphs. Students will:

- **Enter Data into an Excel Spreadsheet**
- **Create and Modify Graphs**
  - Using Chart Wizard to Plot the Data
    - Select Chart Type
    - Add Data
    - Format Graph

- **Create Trendlines to find Standard Curves in Excel**
  - Display equation
  - Display R^2 value

In **Part B** of this exercise, students will use some of the skills practiced in Part A to graph and assess the trends and variations in the average mass of U.S. pennies produced over a 25 – 30 year period.

PART A – Preparation of a Standard Curve/ Determination of Concentration from a Standard Curve

General Instructions:

1. The assignment is to be completed by groups of two students, but **each student must submit a copy of the assignment for grading**.
2. All data should be recorded and graphed using all available significant figures.
3. Your instructor will assign either **Exercise 1** or **Exercise 2** to each student group. Each student should work on and submit work for the assigned Exercise.

Exercise 1: Standard Curve for Protein Measurements:

A standard curve for protein concentration is often created using known concentrations of bovine serum albumin (protein). This process is called the Bradford Assay; it is a colorimetric assay. A special reagent turns blue when it binds to amino acids present in protein. The intensity of the color is best measured with a spectrophotometer (a device for comparing two light radiations, wavelength by wavelength). In the case of the Bradford Assay the greater the absorbance, the higher the protein concentration.

A series of tests were performed on some samples and the following measurements were obtained using a spectrophotometer:

Protein Concentration (mg/ml)	Absorbance (A)
0.26	0.098
0.56	0.213
0.84	0.383
1.12	0.473
1.40	0.527

TASKS:

1. Enter the data into Excel – the protein concentration is the independent variable.
2. Create a scatter graph of the data that is appropriately titled and labeled (see Introduction).
3. Use the Trendline feature of Excel to find the best linear fit to the data.
4. Display the equation and the R^2 value of the trendline on the graph and print out the graph.
5. Print the graph, scaling so that it fits on an 8½ x 11 in. sheet.
6. Use the equation for the standard curve to determine the protein concentrations of the following unknown samples:

Absorbance (A)
0.023
0.201
0.642
0.760

Exercise 2: Standard Curve for DNA Measurements:

Pure, dissolved DNA forms a transparent, colorless solution. DNA will react with a chemical called diphenylamine (DPA), and the reaction produces a blue color. The DPA reacts with the deoxyribose in the DNA and is quantitative; that is, the more DNA present, the more intense the blue color. The intensity of the blue color is measured using a spectrophotometer as described in Exercise 1 above.

A series of tests were performed and the following measurements were obtained using a spectrophotometer:

µg DNA in the Sample	**Absorbance (A)**
0	0
100	0.064
200	0.132
400	0.253
600	0.415
800	0.512
1000	0.694

TASKS:

1. Enter the data into Excel – the amount of DNA is the independent variable.
2. Create a scatter graph of the data that is appropriately titled and labeled (see Introduction).
3. Use the Trendline feature of Excel to find the best linear fit to the data.
4. Display the equation and the R^2 value of the trendline on the graph and print out the graph.
5. Print the graph, scaling so that it fits on an 8½ x 11 in. sheet.
6. Use the equation for the standard curve to determine the DNA concentrations of the following unknown samples:

Absorbance (A)
0.015
0.280
0.356
0.480
0.763

Completed assignment for **Exercise 1** or **Exercise 2** includes Part A Data Sheet, showing results for unknown concentrations and *sample calculation* for at least one unknown concentration, and the titled and labeled graph.