Appendix B

Background Quantitative Techniques

B.1 Introduction

Environmental science is essentially concerned with establishing and understanding relationships between phenomena. As you progress through the course you will be introduced to a number of physical relationships. Such relationships may be stated in a number of ways ranging from the verbal to the mathematical. For a variety of reasons relationships that can be expressed in mathematical form are more powerful. In order to help you become familiar with the expression and analysis of simple mathematical relationships a number of component geographic techniques are outlined. It is important that you study and understand this section.

B.2 Physical Relationships as Simple Mathematical Expressions

Let us begin by considering the nature of an expression (more usually called a function) written in its mathematical form. Typically we write:

$$y = f(x) \tag{B.1}$$

Here we are expressing, in formal mathematical notation, the idea that y's dependence on x can be expressed in a mathematical form. A host of other symbols may be used in place of f, including $F, \phi, y,$ and g. As an example, consider the kinetic energy of an object (you will be introduced to this physical relationship in the course). Repeated experiments have identified that KE is a function of (or is dependent on) the velocity (v) and mass (m) of the object. Thus we can write:

$$KE = f(m, v) \tag{B.2}$$

Although the above form tells us that KE depends upon the mass and velocity of an object, we cannot say how much KE results from a given mass moving with a specific velocity. In order to do this we need to define the form of the functional relationship. Fortunately this has been done, so more specifically:

$$KE = \frac{1}{2}mv^2 \tag{B.3}$$

Here we see another aspect of physical relationships – the constant of proportionality. I.e. for objects of different masses and velocities, the resultant KE will be different, but the physical law always holds in the above form.

B.3 Manipulation of Simple Mathematical Equations

The study of environmental science necessarily involves some knowledge of mathematically expressed physical relationships. It is important that you are familiar with them and confident about what they are telling you about the form of the relationship. As a beginning, consider the idea of re-expressing the above equation:

$$KE = \frac{1}{2}mv^2 \tag{B.4}$$

so that it is written in the form:

$$v = f(KE, m) \tag{B.5}$$

Try to do this. The manipulation of equations is a basic science technique. You will be faced with a number of such simple expressions during the course.

B.4 The Graphical Analysis of Physical Relationships

A function can be portrayed graphically as a curve. Drawing curves provides information about how two variables are related. We are concerned with understanding how variables in the environment are interrelated. The relation between dependent variables (i.e. the variable that is responding – KE in equation B.4) and independent variables (i.e. the variable that is influencing – m, v in equation B.4) is usually obtained by fitting a curve to the observed values for those variables. Hopefully the form of the resultant curve is simple enough that we can recognize its mathematical form (the simplest being a linear relationship).

B.5 Linear Relationships

Data are normally plotted so that the theoretical relationship between the pairs of observations (i.e. the x's and y's) is a linear one. A straight line is the curve where it is easiest to see any discrepancies between experimental points and the theoretical curve shape. Also, it is easy to see and analyze systematic departures from linearity. It is simple to write the form of the equation expressing a linear relationship:

$$y = mx + b \tag{B.6}$$

where the parameter m is the slope of the line and parameter b is the value of y when x equals zero. The variable y is the dependent variable, and x is the independent variable. A line with positive slope m and positive intercept b is shown in the following figure. The slope of a line can be found by choosing two points on the line (x_1, y_1) and (x_2, y_2). They should be as far apart as possible to minimize the error. Since:

$$y_1 = mx_1 + b \tag{B.7}$$

and

$$y_2 = mx_2 + b \tag{B.8}$$

subtracting the 2 equations from each other and regrouping

$$(y_2 = mx_2 + b) \quad - \quad (y_1 = mx_1 + b) \tag{B.9}$$

$$(y_2 - y_1) \quad = \quad (mx_2 - mx_1) + (b - b) \tag{B.10}$$

$$(y_2 - y_1) \quad = \quad m(x_2 - x_1) \tag{B.11}$$

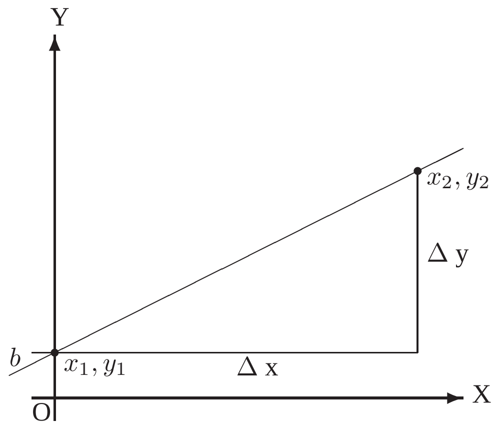
or, rearranged as

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x} = \frac{\text{rise}}{\text{run}} \tag{B.12}$$

The intercept b can be found by substituting a point (x, y) into the equation once the slope is known:

$$b = y_1 - mx_1 \tag{B.13}$$

Or sometimes it can be read directly from the graph by noting the y value where the line crosses the y axis (at $x = 0, y = b$). So, m and b are two defining constants for each linear relationship.



B.6 Non Linear Relationships

An initial plot of the data may not appear as a linear relationship. Given the advantages inherent in a linear relationship, it may be possible to manipulate the data, to obtain a straight line. It is often possible to obtain a straight line by applying a logarithmic transformation to either one or both of the variables. This is equivalent to plotting the data on semi-log, or full log paper.

A logarithm applied to only the "y" variable will give a linear plot whenever the data follows a relationship of the form:

$$y = be^{mx} \tag{B.14}$$

which is an exponential relationship. Here y and x are the variables, b and m are constants to be determined (they are analogous to the intercept b, and slope m), and e is a number, 2.71828 (the base of natural logarithms), although any other constant (for example 10) could also be used. To make the curve $y = be^{mx}$ plot as a straight line, take logarithms of both sides as follows:

$$\log y = \log b + mx \log e \tag{B.15}$$

To put the equation more clearly in the form of a straight line ($y = mx + b$), substitute $Y = \log y$, and $M = m \log e$. Then the equation becomes:

$$Y = Mx + \log b \tag{B.16}$$

The slope of the line, $M = m \log e$ gives us the multiplier on the exponent. Similarly, the intercept on the y axis ($x = O$) gives the constant b. [NOTE: if natural logarithms are used $M = m$ since $\ln e = 1$]

When logarithms are applied to both variables, the equation of the form:

$$y = bx^m \tag{B.17}$$

becomes a straight line. This is a "power" relationship because x is raised to the power m. In this equation y and x are the variables, b is a constant (to be determined) and m is a power (positive, negative, whole or fractional).

Taking logarithms of both sides:

$$\log y = \log b + m \log x \tag{B.18}$$

Substituting $Y = \log y$ and $X = \log x$, this becomes:

$$Y = \log b + mX \tag{B.19}$$

which is recognized as the equation of a straight line. The slope of the line gives us the exponent in the power relationship. The intercept on the y-axis gives the constant b when its anti-logarithm is taken. [The term intercept on full log paper is a bit of a misnomer since x, and y never equal zero. However $X = \log x$ does cross zero when $x = 1$ (because $\log 1 = 0$).]

B.7 Graphing Guidelines

It is important to prepare your graphs carefully. The following set of rules should help you construct graphs that are visually appealing and will aid data analysis.

1. Ideally plot the graph before any data collection equipment is dismantled or before leaving the data collection site. This allows gaps where data are missing to be filled, and anomalous (extreme) data points can be checked.

2. Plot the dependent variable on the y-axis unless you are plotting depths or heights. For depths or heights the convention is to always plot these on the y-axis as they represent vertical data regardless of whether they are dependent of independent variables in the relationship.

3. The graph must have a title. Put the title clearly above the graph. Make the title descriptive in order to explain what the graph is showing.

4. The graph must have axes labelled with names and units. Draw the axes and select appropriate scales. Mark the scales at uniform intervals. Make the graph as large as is reasonable, but be sure that the major divisions of the graph paper have a simple relationship to the scales you have selected. Usually the axes are drawn from the origin (O,O), but it is not necessary to do so if all the data points plot far from the origin. If the graph is to demonstrate that y is directly proportional to x, then the origin should be included. Most graphs plot positive values of x and y ($+x$ and $+y$, i.e., the first quadrant), but this is not always the case. Plan ahead!

5. Plot the points clearly. Use a pencil! Use a dot enclosed in a circle or a cross. This allows the data point to be defined precisely. When a more than one curve is to be shown on the same graph, use different symbols for each set of points (e.g. dots in squares, triangles, etc.).

6. For non-linear curves where you are trying to express a mathematical relationship between y and x, draw a neat smooth freehand line through, but not necessarily connecting the data points. (NOTE THAT THIS DOES NOT APPLY TO ALL GRAPHS – see 7). Never join these data using a ruler and the dot-to-dot method. Use a pencil and draw lightly at first, as you may want to erase parts or sections of the curve and try again. Try to get the data points evenly spaced about your curve, with roughly equal numbers above and below your line if the data points do not pass through your curve.

7. For graphs where there is no mathematical expression for the relationship, such as profiles where you are trying to follow the trend of one variable with the other, draw a neat smooth freehand line passing through each dot. This method is appropriate as you are interested in the point by point variation. Even though the line must pass through each point, it is a smooth curve (not drawn with a ruler).

8. For histograms, usually a bar-graph style is used. Choose simple class sizes of approximate size (range / $\sqrt{n}$). You may plot the raw count data, and/or the percent, on the y-axis.

B.8 Significant Figures

Every measurement is uncertain to some extent. Suppose, for example, that we wish to measure the mass of an object. If we use a platform balance, we can determine the mass to the nearest 0.1 g.

An analytical balance, on the other hand, is capable of given results correct to the nearest 0.0001 g. The exactness, or precision, of the measurement depends upon the limitations of the measuring device and the skill with which it is used.

The precision of a measurement is indicated by the number of figures used to record it. The digits in a properly recorded measurement are significant figures. These figures include all those that are known with certainty plus one more which is an estimate.

Suppose that a platform balance is used, and the mass of an object is determined to be 12.3 g. The chances are slight that the actual mass of the object is exactly 12.3 g, no more nor less. We are sure of the first two figures (the 1 and the 2); we know that the mass is greater than 12 g. The third figure (the 3), however, is somewhat inexact. At best, it tells us that the true mass lies closer to 12.3 g than to either 12.2 g or 12.4 g. If, for example, the actual mass were $12.28\ldots$ g or $12.33\ldots$ g, the value would be correctly recorded in either case as 12.3 g to three significant figures.

If, in our example, we add a zero to the measurement, we indicate a value containing four significant figures (12.30 g) which is incorrect and misleading. This value indicates that the actual mass is between 12.29 g and 12.31 g. We have, however, no idea of the magnitude of the integer of the second decimal place since we have determined the value only to the nearest 0.1 g. The zero does not indicate that the second decimal place is unknown or undetermined. Rather, it should be interpreted in the same way that any other figure is (see, however, rule 1 that follows). Since the uncertainty in the measurement lies in the 3, this digit should be the last significant figure reported.

The following rules can be used to determine the proper number of significant figures to be recorded for a measurement.

1. Zeros used to locate the decimal point are not significant. Suppose that the distance between two points is measured as 3 cm. This measurement could also be expressed as 0.03 m since 1 cm is 0.01 m.

   3 cm $= 0.03$ m

   Both values, however, contain only *one* significant figure. The zeros in the second value, since they merely serve to locate the decimal point, are not significant. The precision of a measurement cannot be increased by changing units.

   Zeros that arise as a part of a measurement are significant. The number 0.0005030 has four significant figures. The zeros after 5 are significant. Those preceding the numeral 5 are not significant since they have been added only to locate the decimal point.

   Occasionally, it is difficult to interpret the number of significant figures in a value that contains zeros, such as 600. Are the zeros significant, or do they merely serve to locate the decimal point? This type of problem can be avoided by using scientific notation. The decimal point is located by the power of 10 employed; the first part of the term contains only significant figures. The value 600, therefore, can be expressed in any of the following ways depending upon how precisely the measurement has been made.

   6.00×10^2 (three significant figures)
   6.0×10^2 (two significant figures)
   6×10^2 (one significant figure)

2. Certain values, such as those that arise from the definition of terms, are exact. Far example, by definition, there are *exactly* 1000 ml in 1 liter. The value 1000 may be considered to have an infinite number of significant figures (zeros) following the decimal point.

3. At times, the answer to a calculation contains more figures than are significant. The following rules should be used to round off such a value to the correct number of digits.

   (a) If the figure following the last number to be retained is less than 5, all the unwanted figures are discarded and the last number is left unchanged.

   3.6247 is 3.62 to three significant figures.

   (b) If the figure following the last number to be retained is greater than 5 or 5 with other digits following it, the last figure is increased by 1 and the unwanted figures are discarded.

   7.5647 is 7.565 to four significant figures
   6.2501 is 6.3 to two significant figures

   (c) If the figure following the last figure to be retained is 5 and there are only zeros following the 5, the 5 is discarded and the last figure is increased by 1 if it is an odd number or left unchanged if it is an even number. In a case of this type, the last figure of the rounded off value is always an even number. Zero is considered to be an even number.

   3.250 is 3.2 to two significant figures
   7.635 is 7.64 to three significant figures
   8.105 is 8.10 to three significant figures

The number of significant figures in the answer to a calculation depends upon the numbers of significant figures in the values used in the calculation. Consider the following problem. If we place 2.38 g of salt in a container that has a mass of 52.2 g, what will be the mass of the container plus salt? Simple addition gives 54.58 g. But we cannot know the mass of the two together any more precisely than we know the mass of one alone. The result must be rounded off to the nearest 0.1 g, which gives 54.6 g.

4. The result of an addition or subtraction should be reported to the same number of decimal places as that of the term with the least number of decimal places. The answer for the addition

$$161.032 + 5.6 + 32.4524 = 199.0844 \tag{B.20}$$

should be reported as 199.1 since the number 5.6 has only one digit following the decimal point.

5. The answer to a multiplication or division is rounded off to the same number of significant figures as is possessed by the least precise term used in the calculation. The result of the multiplication

$$152.06 \times 0.24 = 36.4944 \tag{B.21}$$

should be reported as 36, since the least precise term in the calculation is 0.24 (two significant figures).

B.9 Analysis of Measurement Error

All measurement to some extent is uncertain. This uncertainty can arise from several causes: human error in reading the instrument, sampling error (the "population" is incorrectly sampled), and poor instrument precision, accuracy or bias.

Measurement uncertainty is often given by a $\pm$ term following the measured value. For example a measurement of 10.1 ± 0.5m means that the investigator thinks the true value lies between 9.6 and 10.6m. This could be because the instrument used to make the measurement is only accurate to within 1.0 meters, or because several measurements were made which had a mean of 10.1m and a standard deviation of 0.5m. When two or more measurements with error terms are added together, it is too pessimistic to simply add the error terms as well – the errors are equally likely to cancel each other as add to each other. Instead, the rule for the addition of error terms is to square the errors, add them up, and take the square root of the sum. For example, if one were adding 3 lengths; 10.1 $\pm$ 0.5m, 2.6 $\pm$ 0.2m, and 18.2 $\pm$ 1.1m, the answer would be:

$$10.1 + 2.6 + 18.2 = 30.9\text{m} \pm \text{Error} \tag{B.22}$$

with an error term of

$$\text{Error} = \sqrt{0.5^2 + 0.2^2 + 1.1^2} = 1.2\text{m} \tag{B.23}$$

so that the answer would be reported as 30.9 $\pm$ 1.2m. For subtraction, the errors are still the same – just as likely to cancel each other out as to work in the same direction, so the total error follows the same rule as addition.

Estimating the combined error which results when measurements are multiplied or divided is similar to the addition / subtraction method, except fractional errors are used in the calculation. For example, suppose a drainage basin has an area of 5.6 ± 0.1 km^2, and the average annual rainfall over that basin is 678 ± 203 mm yr^{-1}. The annual volume of rain falling on that basin would then be:

$$5.6 \times 10^6\text{m}^2 \times 0.678\text{m} = 3.8 \times 10^6\text{m}^3 \pm \text{error} \tag{B.24}$$

with an error term of:

$$\text{Error (fractional)} = \sqrt{(\frac{.1}{5.6})^2 + (\frac{203}{678})^2} \tag{B.25}$$

$$= \sqrt{0.00032 + 0.0896} = .30 \tag{B.26}$$

$$\text{Error (in m}^3\text{)} = 0.3 \times 3.8 \times 10^6 = 1.1 \times 10^6\text{m}^3 \tag{B.27}$$

(B.28)

so that the answer would be reported as $3.8 \pm 1.1 \times 10^6$m^3. Estimating the combined error in equations involving exponents is similar to multiplication and division, except that the exponent is regarded as a multiplier of the fractional error. For more information, refer to a text such as Haynes (1982).

B.10 Scientific Notation

Scientific notation is a convenient “shorthand” way of depicting very large or very small numbers without the use of many zeros. The notation x^n means that the number x is multiplied by itself n times (e.g. $2^3 = 2 \times 2 \times 2 = 8$) where n is called the exponent. Similarly x^{-n} with a negative exponent, is the reciprocal of x^n, that is $x^{-n} = 1/x^n$ (e.g. $2^{-3} = 1/(2 \times 2 \times 2) = 1/8 = 0.125$). It is often especially convenient to express large or small numbers as powers of 10 (i.e. 10^n or 10^{-n}) and certain of these are given prefixes as listed below.

Standard form for scientific notation is to express numbers so that in the decimal part, there is one digit to the left of the decimal point:

$$a.bc \times 10^d$$

where a, b, c, d are all numbers from 0-9. Scientific notation is also useful, since it can be immediately clear how many significant figures there are in a measurement. In the above example, there would be 3 significant figures, since 3 digits are displayed. For example, the number 123 in standard scientific form would be 1.23×10^2. If the number 123 represented a measurement, that we know accurately to 2 decimal places, then it should be written 123.00, or in standard scientific form: 1.2300×10^2.

Prefix			Scientific notation	Decimal notation
T	tera-	one trillion	10^{12}	$1,000,000,000,000$
G	giga-	one billion	10^9	$1,000,000,000$
M	mega-	one million	10^6	$1,000,000$
k	kilo-	one thousand	10^3	$1,000$
h	hecto-	one hundred	10^2	100
da	deka-	ten	10	10
d	deci-	one tenth	10^{-1}	0.1
c	centi-	one hundredth	10^{-2}	0.01
m	milli-	one thousandth	10^{-3}	0.001
μ	micro-	one millionth	10^{-6}	0.000001
n	nano-	one billionth	10^{-9}	0.000000001
p	pico-	one trillionth	10^{-12}	0.000000000001

B.11 Greek Alphabet

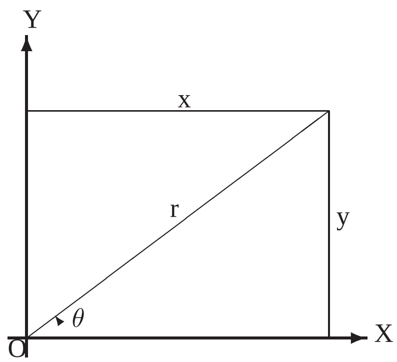
Lower case	Capital	Name	Lower case	Capital	Name
α	A	alpha	ν	N	nu
β	B	beta	ξ	Ξ	xi
γ	Γ	gamma	o	O	omicron
δ	Δ	delta	π	Π	pi
ϵ	E	epsilon	ρ	P	rho
ζ	Z	zeta	σ	Σ	sigma
η	H	eta	τ	T	tau
θ	Θ	theta	υ	Υ	upsilon
ι	I	iota	ϕ	Φ	phi
κ	K	kappa	χ	X	chi
λ	Λ	lambda	ψ	Ψ	psi
μ	M	mu	ω	Ω	omega

B.12 Mathematical Signs and Symbols

$=$	equals	$\sim$	is similar to
$\approx$	equals approximately	$\neq$	does not equal
$\equiv$	is identical to, is defined as	$>$	is greater than
$<$	is less than	$\geq$	is greater than or equal to
$\leq$	is less than or equal to	$\pm$	plus or minus ($\sqrt{4} = \pm 2$)
$\propto$	is proportional to	$\sum$	the sum of
$\overline{x}$	the average value of x		

B.13 Useful Constants and Formulae

Gravity	$g = 9.82$ m s^{-2}	trigonometric	$\sin\theta = \frac{y}{r}$
Stephan-Boltzmann constant	$\sigma = 5.6697 \times 10^{-8}$ W m^{-2} K^{-4}	functions	$\cos\theta = \frac{x}{r}$
rectangle circumference	= 2(length + width)		$\tan\theta = \frac{y}{x} = m$ (slope)
rectangle area	= length $\times$ width		$\csc\theta = \frac{r}{y} = 1/\sin\theta$
circumference of a circle	$= 2\pi r$ (r is radius)		$\sec\theta = \frac{r}{x} = 1/\cos\theta$
area of a circle	$= \pi r^2$		$\cot\theta = \frac{x}{y} = 1/\tan\theta$
area of a sphere	$= 4\pi r^2$	Pythagorean	$x^2 + y^2 = r^2$
volume of a sphere	$= \frac{4}{3}\pi r^3$	theorem	
quadratic formula	if $ax^2 + bx + c = 0$, then $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$		



B.14 References

Haynes, R.M., J.G. Harvey and T.D. Davies, 1982. *Measurement, In Environmental Science Methods*, R. Haynes (ed.), Chapman and Hall, London, 1-16.
Sumner, G.N., 1978. *Mathematics for Physical Geographers*, Edward Arnold, London, 1-16.