PRESENTING SCIENTIFIC INFORMATION

Introduction

There are rules and conventions for presenting scientific and engineering information to which your reports need to conform. At the same time, there are terms and their application that are not always clearly understood.

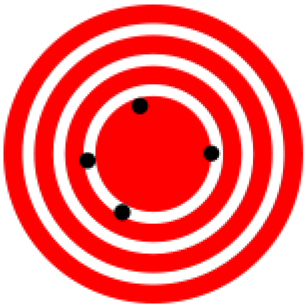
Accuracy and Precision

In the fields of science, engineering, mathematics and statistics *accuracy* and *precision* are often confused and wrongly applied i.e. they are not the same thing.

Precision is a description of *random errors*, a measure of *statistical variability*.
Accuracy is a measure of *closeness* of the result or measurement to the target or true result.

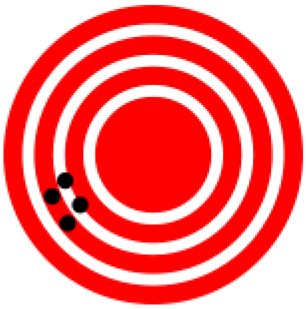
A measurement system may be accurate but not precise, precise but not accurate, neither, or both.

A



Good accuracy, poor precision, good trueness

B



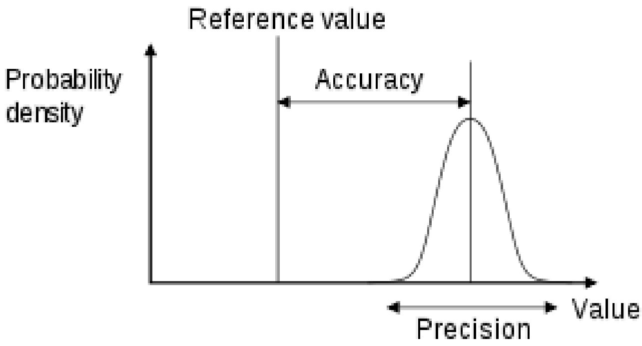
Low accuracy, good precision, poor trueness

In case A, the shooter is quite *accurate* since most of the attempts are within the bulls-eye. There is, however, a fair degree of variability in the shooting; sometimes to the left, sometimes to the right i.e. not consistent. We can say he/she is not *precise*.

In case B, the shooter is *inaccurate*, missing the target by quite a bit; so they are either a bad shot or the sight in the rifle is out of adjustment. Their shooting is, however, very consistent (consistently bad). We can say he/she is *precise*.

If we were asked to state the position on the target for each shooter, in the x- and y- direction; in case A, we would at best say about 5 cm from the left and 4.5 cm from the bottom. This is because of the high degree of scatter or variation. In case B we can define the spot on the target where their shots are landing by 2.55 cm from the left and 1.58 cm from the bottom because the location is much more precisely known.

This is the essence of accuracy and precision. Shooter A was closer to the reference value (the bullseye) but there was lot of scatter. Shooter B was way off the reference value but the location had little spread.



Due to non-random or systematic errors, the readings or measurements may be *precise* but *not accurate* e.g. measuring the mass as 21.3456 kg when it is 50 kg. This may be due to a wrong setting on the instrument, using the wrong formula or equation or the observer not being able to clearly see the dial.

Readings or measurements with many significant figures e.g. 21.3456 are said to be *more precise* than a reading of 21 and require more "precise" instruments e.g. a ruler with millimetre markings rather than only centimetre divisions. The requirement to express an answer to a certain number of significant figures is a measure of precision not accuracy. Precision is affected by random errors and is a measure of scatter e.g. in the case of the shooter, changes in the wind, tension in their arm, minor variations in the explosive cartridge.

There are many references and examples of these terms including the link below;
https://en.wikipedia.org/wiki/Accuracy_and_precision

Significant figures

The population of a town, for example, might only be known to the nearest thousand and be stated as 52,000, while the population of a country might only be known to the nearest million and be stated as 52,000,000. The former might be in error by hundreds and the latter might be in error by hundreds of thousands, but both have two significant figures (figures 5 and 2). This reflects the fact that the significance of the error (its likely size relative to the size of the quantity being measured) is the same in both cases. This is the essence of "significant" figures. Likewise, the size of a bacterium measured with a poor microscope may be 0.0002 mm. The quality of our microscope does not allow us to determine if it is 0.00023 or 0.00018. Relative to the size of the measurement, this is no better than the population example; in fact it is worse and really only 1 significant figure.

Specifically, the rules for identifying significant figures when writing or interpreting numbers are as follows:

- All non-zero digits are considered significant. For example, 91 has two significant figures (9 and 1), while 123.45 has five significant figures (1, 2, 3, 4 and 5).
- Zeros appearing anywhere between two non-zero digits are significant. For example, 101.1203 has seven significant figures: 1, 0, 1, 1, 2, 0 and 3.
- Leading zeros are not significant. For example, 0.00052 has two significant figures: 5 and 2. The zeros are only placeholders to give the digits a place value e.g. tenths or hundredths.
- Trailing zeros in a number containing a decimal point are significant. For example, 12.2300 has six significant figures: 1, 2, 2, 3, 0 and 0. The number 0.000122300 also has only six

significant figures (the zeros before the 1 are not significant). A length of 120.00 m has five significant figures since it has three trailing zeros. The logic here is that the trailing zeros would normally not be written if the measurement was 120 m to the nearest metre. The fact that it was written as 120.00 implies that care was taken with the choice of instrument to improve the precision to hundredths of a metre.

- The significance of trailing zeros in a number not containing a decimal point can be ambiguous. For example, it may not always be clear if a number like 1300 is precise to the nearest unit (and just happens coincidentally to be an exact multiple of a hundred) or if it is only shown to the nearest hundred due to rounding or uncertainty. There are many conventions exist to address this issue which won't be discussed here.

Rounding

The basic concept of significant figures is often used in connection with *rounding*. Rounding to a number of significant figures *n* is a more general-purpose technique than rounding to *n* decimal places, since it handles numbers of different scales in a uniform way.

To round to *n* significant figures:

- Identify the significant figures before rounding. These are the *n* consecutive digits beginning with the first non-zero digit.
- If the digit immediately to the right of the last significant figure is 5 or more, add 1 to the last significant figure. For example, 1.2459 to 3 significant figures should be written 1.25. (There is also the *tie-breaking* rule which we shall not use.)
- Replace non-significant figures in front of the decimal point by zeros.
- Drop all the digits after the decimal point to the right of the significant figures (do not replace them with zeros).

Example for **12.345**:

Precision	Rounded to significant figures	Rounded to decimal places
6	12.3450	12.345000
5	12.345	12.34500
4	12.35	12.3450
3	12.3	12.345
2	12	12.35
1	10	12.3
0	N/A	12

Another example for **0.012345**:

Precision	Rounded to significant figures	Rounded to decimal places
7	0.01234500	0.0123450
6	0.0123450	0.012345
5	0.012345	0.01235
4	0.01235	0.0123
3	0.0123	0.012
2	0.012	0.01
1	0.01	0.0
0	N/A	0

An interesting situation arises when the rounding affects the next position to the left as in the example below:

Precision	Rounded to significant figures	Rounded to decimal places
3	9.96	9.957
2	10.0	9.96
1	10	10.0

Only *measured* quantities are considered in the determination of the number of significant figures in *calculated* quantities. Exact mathematical quantities like π in $A = \pi r^2$ has no effect on the number of significant figures in the final calculated area. Similarly, the ½ in the formula for kinetic energy $KE = \frac{1}{2}mv^2$, has no bearing on the number of significant figures in the final calculated kinetic energy. The constants π and ½ are considered to have an “infinite” number of significant figures.

For quantities created from measured quantities by **multiplication** and **division**, the calculated result should have as many significant figures as the *measured* number with the *least* number of significant figures. For example:

$$1.234 \times 2.0 = 2.468\ldots \approx 2.5$$

with only *two* significant figures. The first factor has four significant figures and the second has two significant figures. The factor with the least number of significant figures is the second one with only two, so the final calculated result should also have a total of two significant figures.

For quantities created from measured quantities by **addition** and **subtraction**, the last significant *decimal place* (hundreds, tens, ones, tenths, and so forth) in the calculated result should be the same as the *leftmost* or largest *decimal place* of the last significant figure out of all the *measured* quantities in the terms of the sum.

For example:

$100.0 + 1.234 = 101.234\ldots \approx 101.2$

with the last significant figure in the *tenths* place. The first term has its last significant figure in the tenths place and the second term has its last significant figure in the thousandths place. The leftmost of the decimal places of the last significant figure out of all the terms of the sum is the tenths place from the first term, so the calculated result should also have its last significant figure in the tenths place.

The rules for calculating significant figures for multiplication and division are opposite to the rules for addition and subtraction.

- For multiplication and division, only the number of significant figures in each of the factors matter; the decimal place of the last significant figure in each factor is irrelevant.
- For addition and subtraction, only the decimal place of the last significant figure in each of the terms matters; the total number of significant figures in each term is irrelevant.

To see the logic in the degree of rounding, examine the following example;

$$\frac{3.45 \times 3.678 - 4.2}{4}$$
$$= 2.1222275$$

Here, we have performed the calculation "precisely" with 7 significant figures but are they meaningful? The "4" in the denominator is only to 1 significant figure. The "4" may have actually been "3.55" and the overall result 2.391295775 or "4.45" and the result 1.907662921. We see a range in the possible answers summarised below with a difference of $\approx$ 0.4836, so the 2nd significant figure onwards are quite meaningless. The answer should be stated as "2" (1 sf).

4.45 →	1.907662921
4 →	2.1222275
3.55 →	2.391295775

Presentation of results

It is important to set your working out in a neat and logical manner with equal signs dividing the steps in your calculations. Wherever applicable, neat annotated diagrams may also be required. Your work should line up and for this reason lined calculation sheets preferably with a grid, should be used. An example may be found in the Appendix to this document.

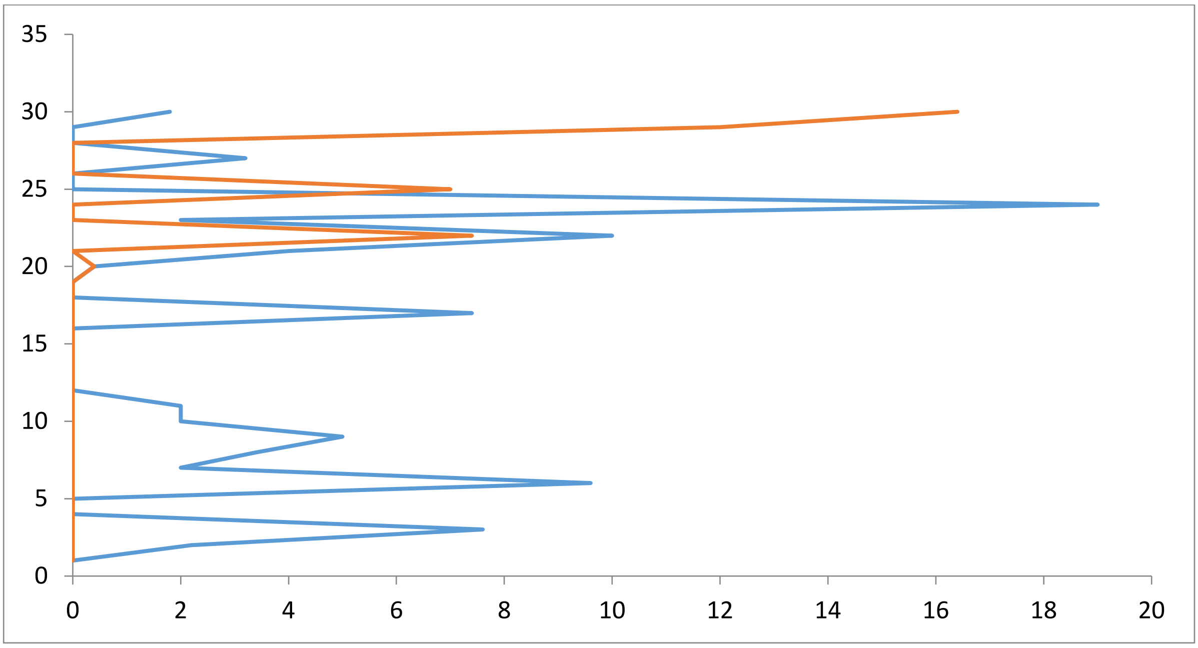
Most engineering quantities have units which are sometimes shown with a prefix. The right unit and prefix must be used and correctly represented i.e. in the right form and either upper or lower case. Refer to the table below for examples of correct and incorrect usage.

Quantity	**Correct**	**Incorrect**
20 kilometres	20 km	20 Km / 20 KM
30 milligrams	30 mg	30 Mg / 30 MG
15 microfarads	15 μF	15 uF
8 nanofarads	8 nF	8 ηF
9 megapascals	9 MPa	9 MPA / 9 mPa / 9 mPA

When using graphs, whether hand drawn or in Excel;

- Show the data without distortion
- Avoid displaying a large set of data in a small space
- Label all axes with the name and units of the variables
- Place independent vs dependent variables on the correct axes
- Insert captions as required
- Use a combination of colour and symbols (hand drawn graphs should be done in pencil)
- Units must be consistent with the text of the report

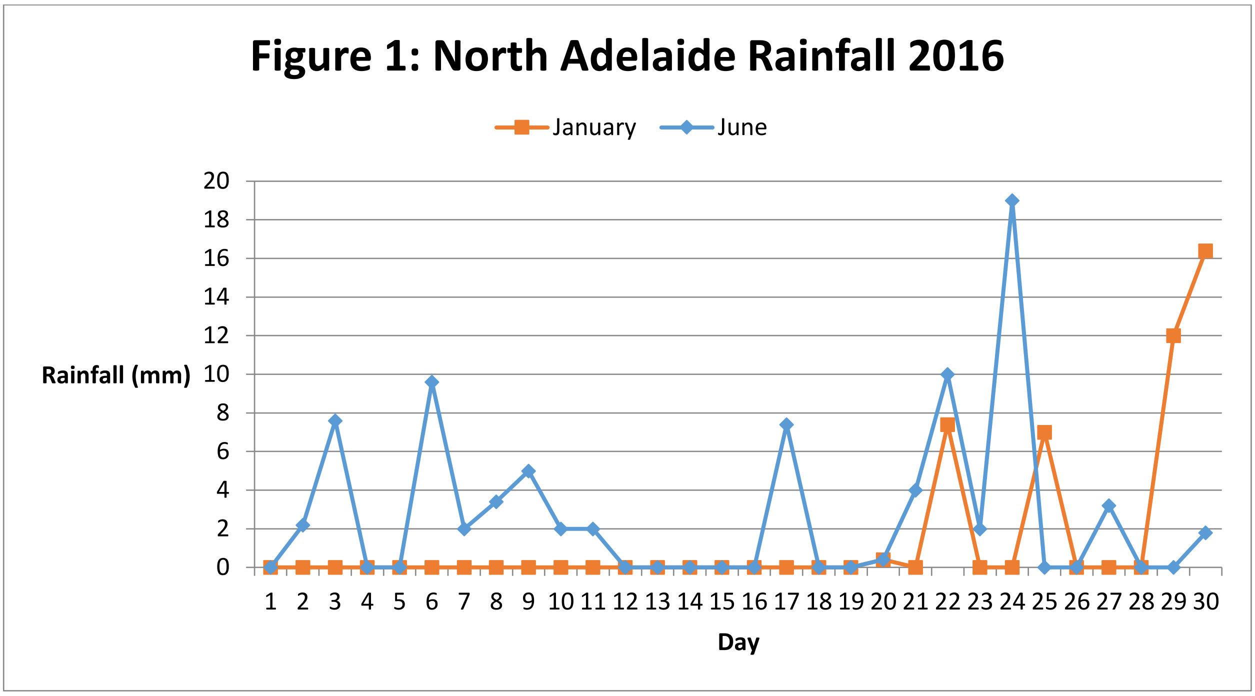
Graph of North Adelaide Rainfall in Jan and Jun 2016



In the graph above;

- The dependent and independent variables need to be switched around and the peaks going up and down
- The axes need to be correctly labelled including the correct units

The graph below is a better representation of the data

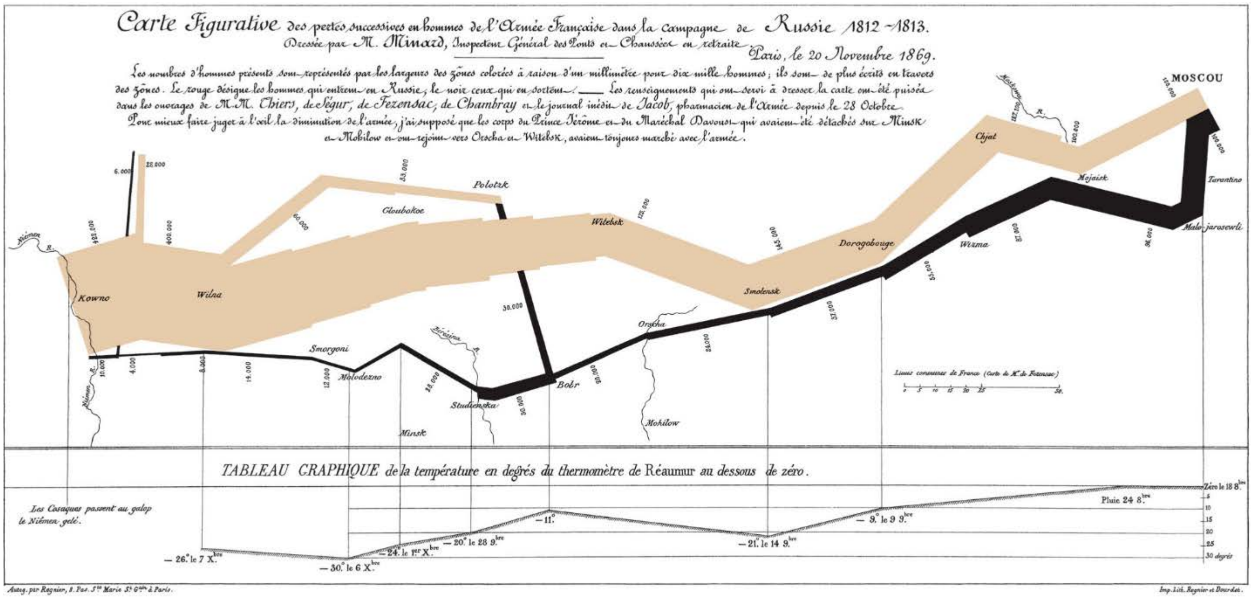


In the rainfall graph, note the use of different shapes for each set of data in the legend rather than simply a different colour. This is because the graph may be printed black and white and the two graphs would not be distinguished if we used the same symbol.

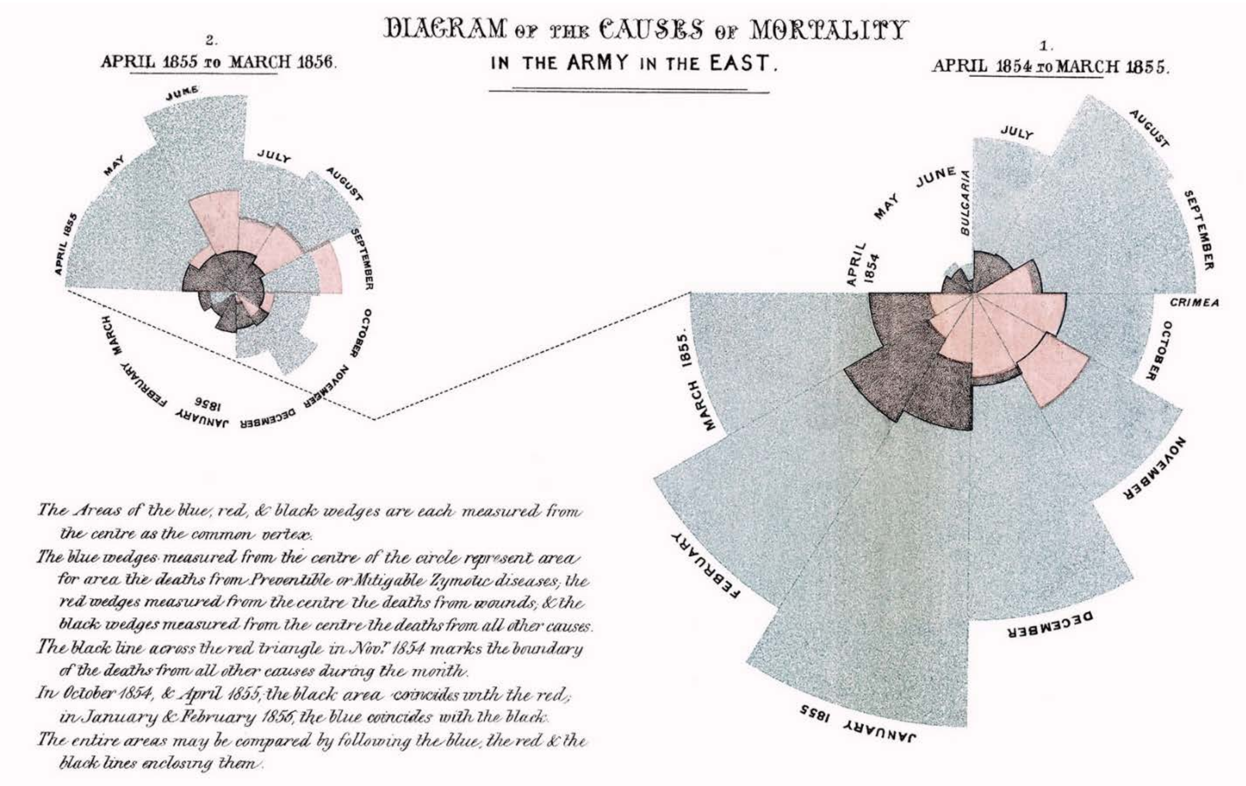
It is also important to avoid displaying too much information on a graph or chart making it difficult to read. Thought needs to be given to the scale so that important information can be easily read and interpreted.

Note that a line chart is not the only way to display information. We can customise our plots to highlight the important ideas such as the following chart of the Napolean Russia campaign in 1812 by Charles Minard. In the map, Minard has used the width of the line to indicate the size of the army and the length for the distance travelled. He also incorporates notes on the line to indicate location and date, colour for direction and a subgraph on the bottom to indicate temperature.

Napolean Russia campaign 1812 map by Charles Minard includes: size, distance, temperature, location, direction, date



Florence Nightingale's Rose Diagram



The second Rose Diagram is a clever use of colour and size to indicate the importance of sanitation. In the Rose Diagram, the blue wedges are death from preventable causes, red are from wounds and black from other causes. Hence, it is clear that preventable causes are the biggest factor for the mortality rate. One problem with the chart is the absence of legends to clearly indicate what the colours represent.

Dependent and independent variables

The *independent variable (cause variable)* is customarily plotted along the horizontal (x-)axis. The measured or *dependent variable (effect variable)* is customarily plotted along the vertical (y-)axis. If no dependent variable exists, either type of variable can be plotted on either axis.

The independent variable is the factor being adjusted and the dependent variable is the result or outcome of this adjustment. For example if one is testing the amount of weight gain with increase of food intake, food intake would be the independent variable and weight gain would be the dependent variable since it depends on the food intake. The table below shows some other examples;

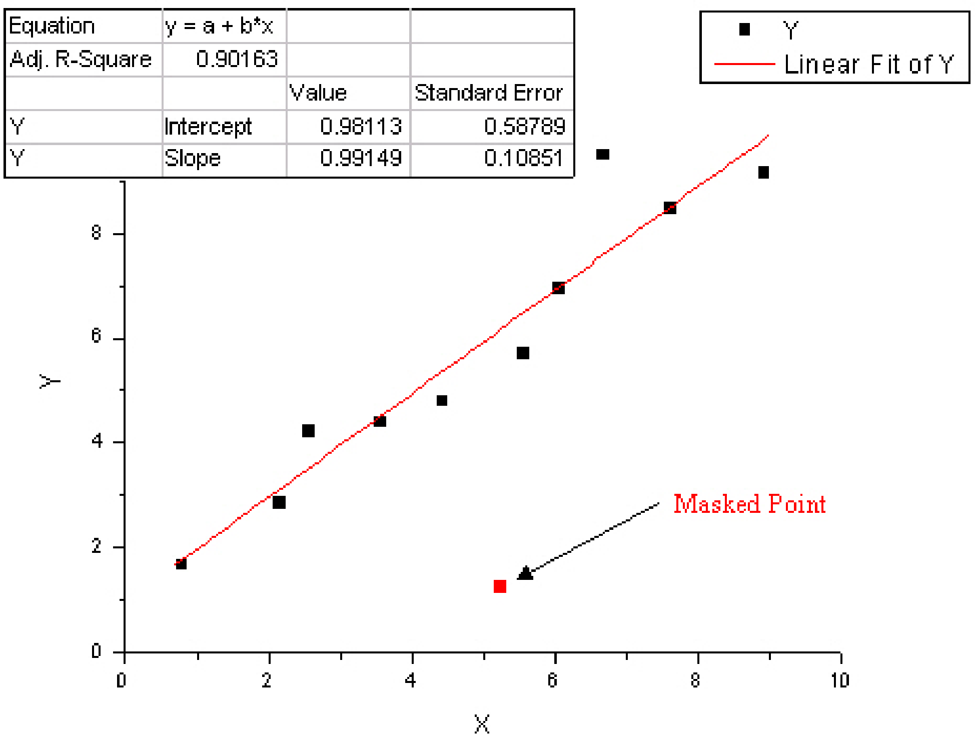
Independent variable	**Dependent variable**
Time in spray booth	Paint thickness
Amount of fertilizer	Plant growth
Study time	Exam results
Alcohol consumed	Road accidents

Correlation and scatter

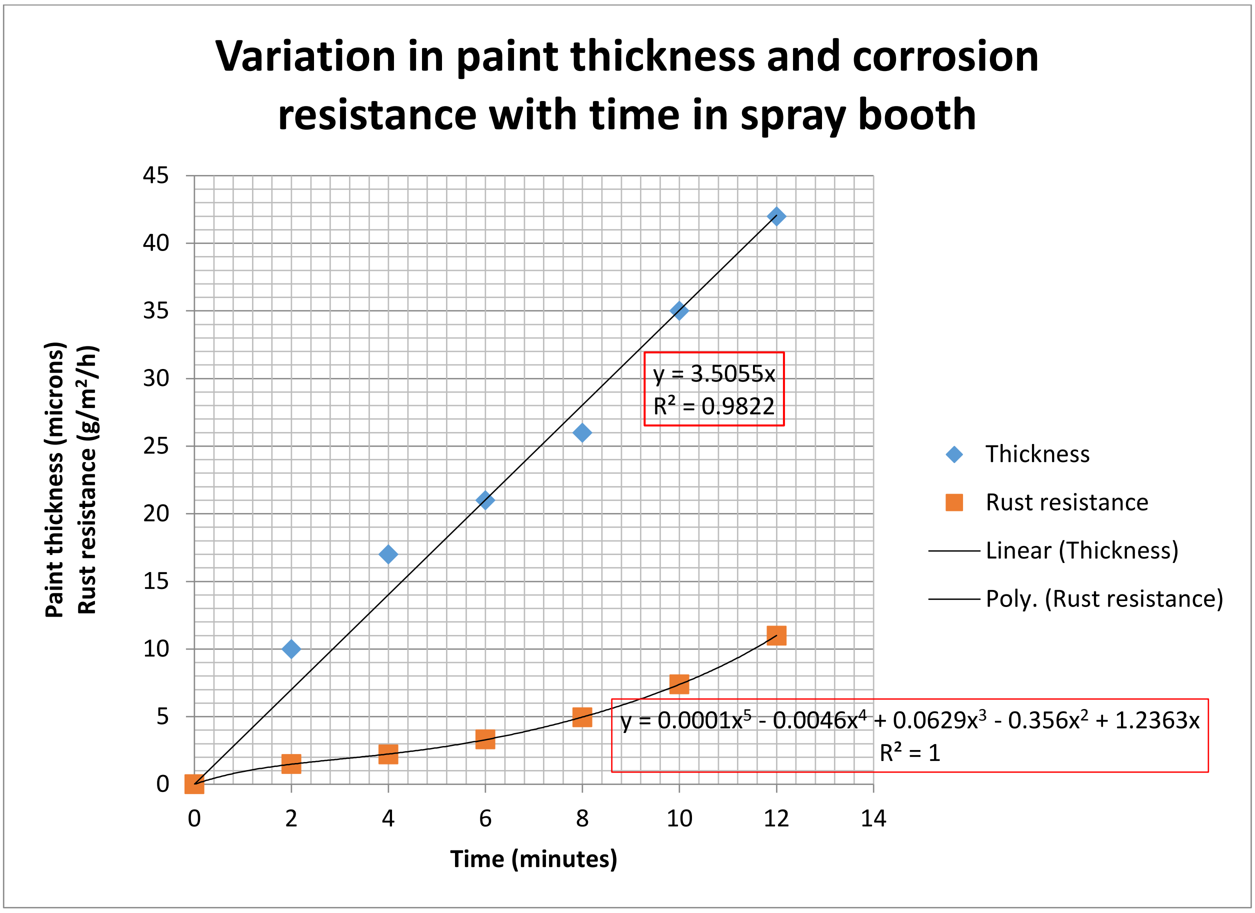
Correlation involves looking for a relationship between different sources of information and implies that there is a relationship between the variables. It means that as one variable changes, the other also changes e.g. increased rates of heart attack with increased cholesterol levels. It is important to vary only one control variable, otherwise it would be very difficult to determine correlation. The better the correlation, the tighter the points will hug a *trendline*. For a *linear* relationship, a strong correlation is present when the points are tight, resembling a straight line. A weak correlation is present when the pattern still resembles a straight line but the points are scattered or spread out and may mean that there are other causes or factors. The relationship between the two variables may, however, be *non-linear* e.g. population growth, with the *trendline* resembling a curve.

A *line of best fit* or *trendline* attempts to have a balance between the points above and below the line, both in number and deviation or degree of scatter. The strength of the relationship is represented by the *correlation coefficient*, sometimes referred to as the “r-value”. In Microsoft Excel the goodness of fit of a model is represented by another statistical measure, the *coefficient of determination* R^2 or r^2. In regression analysis, the R^2 is a statistical measure of how well the regression line approximates the real data points. An R^2 of 1 indicates that the regression line perfectly fits the data. When drawing graphs in Excel, different “Trendline Options” can be selected and each “Trend/Regression Type” tested to achieve a R^2 value closest to 1.

An *outlier* is an observation point that is distant or stands apart from other observations and the overall pattern and sometimes excluded from the data set. An outlier may be due to measurement errors, may indicate experimental error; systematic errors, faulty data, erroneous procedures, or areas where a certain theory might not be valid. The Excel chart below illustrates the application of a trendline and the presence of an outlier.



There are situations where (0, 0) may be a data point, but if it is a value precisely defined, the line of best fit needs to go through (0, 0) not simply balanced around it. For example, the graphs showing paint thickness and corrosion resistance for different times car panels have been placed in a heated spray booth will pass through (0, 0) precisely. In Excel, this can be achieved by ticking the “Set Intercept = 0.0” box. The graphs below illustrate all the above principles. Note that the relationship between time and rust resistance is non-linear.



Scale and gridlines

It is also important to determine the scale and the subdivisions (called Major and Minor gridlines in Excel) so that the labelling is clear and uncluttered. In the graph above, only the major divisions are labelled but intermediate values such as 6, 12 etc. on the vertical axis can be easily determined from the minor gridlines. Adding the 6, 7, 8 etc. would clutter the axis label making it difficult to read. At the same time avoid major divisions using 3, 7 etc. because it makes it very difficult to interpolate or read intermediate values; stick to 2, 5 and 10. The graph showing North Adelaide Rainfall at the start could have been more clearly represented marking all days but labelling only every 5th day e.g. 0, 5, 10, 15 etc.

Appendix

Date: Subject: Lecturer:

Assessment: Student:

Page of

Significant Figures – Exercises

A. State the number of significant figures in each of the following.

1. 1000
2. 0.00035
3. 0.000350
4. 1006
5. 560
6. 560.0
7. 25.3 x 4
8. 25.3 x 4.0
9. 25.3 x 1.654 / 0.0004
10. 23.7 + 654.189

B. Round the following to the required number of significant figures.

1. 742,396	4 s.f.
2. 742,396	3 s.f.
3. 742,396	2 s.f.
4. 0.07988	4 s.f.
5. 0.07988	3 s.f.
6. 0.07988	2 s.f.
7. 5.00 x 4.001	4 s.f.
8. 5.00 x 4.001	3 s.f.
9. 5.00 x 4.001	2 s.f.
10. 13.5 + 0.5	1 s.f

Graphing – Exercises

On the blank graph sheet below, label the variables on the correct axis and mark (with a value) the main divisions (major gridlines). Plot the data points and draw the line of best fit (trendline), taking note of the significance of (0, 0).

Time in spray booth (seconds)	Paint thickness (microns)
20	10
45	17
65	28
80	42
120	78
195	135