11 Units, Best Values, Formulas, and Other Good Stuff

True Value

Statements about the accuracy of numbers imply that an accurate, or true value, is known. Measurements that are based essentially on counts, where the items counted are people, money, pulses in digital systems, and so on, can be considered accurate. A statement that a new computer costs $1499 means exactly 1499 dollars.

On the other hand, for measurements of analog variables, and this includes nearly all process measurements, such as temperature and pressure, the exact value is known only to the Great Creator. The value ultimately obtained will only be the best approximation that we humans can make. Even if we are able to get the first 6 digits correct, it is still an approximation. A good quality temperature sensor might output a reading of 68.5°C, but this does not mean that the temperature is exactly 68.500 000 000... It simply means that the true temperature lies somewhere between 68.45° and 68.55°. Still, up until about 1960, engineering calculations were done on sliderules, which were capable of producing only 3 significant figures, but this was good enough to design oil refineries that would produce products and airplanes that would fly.

In industry, the true value of any process measurement is usually taken to be the value indicated by some standard measurement device. Unfortunately, the majority of measurement standards are not rugged enough to tolerate the plant environment and have to be kept in a laboratory. The sensors that are used in the process operations are then calibrated against the standard.

For this procedure to give satisfactory results, the accuracy of the standard must necessarily be two or three orders better than that of the plant sensor. A fact that helps the situation is that process measurements nearly always have a tolerance that is acceptable for the successful production of products. A temperature measurement that is within ±1 degree of the true value might well be sufficiently accurate for operating a plant.

Errors

It is customary to express the accuracy of a measurement in terms of the error that can be expected. The error is defined as the difference between the observed value of the measurement (OV) and the true or accurate value (TV). Thus, if the error is designated ε, then

$$\varepsilon = OV - TV.$$

An error can be expressed in actual units of measurement or as a fraction or percentage of the true value. An error can be positive or negative, depending on whether the observed value is greater or less than the true value.

Sometimes statements about the accuracy of measurements tend to get sloppy. Such a statement might be "Orifice meters are 3% accurate." This statement could be interpreted incorrectly in either of two ways. First, the statement that the meter is 3% accurate implies that it could be 97% in error, which is unrealistic. Second is the implication that no matter what the reading of the meter might be, it will be in error by 3%. This is not realistic either because most industrial meters can usually be calibrated so that they are accurate at at least two points in their range of measurement. A proper statement of accuracy for the meter is "This meter is accurate to within 3% (of the true value)." The key word is "within."

Errors in Combinations of Quantities

Two quantities, whose true values are X and Y, are known to have possible errors of ε_x and ε_y. If the two quantities are to be added, what will be the error in the sum? If X and Y are the true values, then the measured values will be $X + \varepsilon_x$ and $Y + \varepsilon_y$, bearing in mind that the errors could be positive or negative.

$$(X + \varepsilon_x) + (Y + \varepsilon_y) = (X + Y) + (\varepsilon + \varepsilon_y)$$

The conclusion drawn from this is that *when quantities with known errors are added or subtracted, the **actual** errors are added*. Note that errors are always added, never subtracted, even if the quantities to which the errors belong are subtracted.

If the two quantities X and Y are to be multiplied, then

$$(X + \varepsilon_x) \times (Y + \varepsilon_y) = XY + Y\varepsilon_x + X\varepsilon_y + \varepsilon_x\varepsilon_y\,.$$

Since ε_x and ε_y are hopefully small compared with X and Y, for the purposes of this calculation, their product $\varepsilon_x\varepsilon_y$ can be ignored, Then,

$$XY + Y\varepsilon_x + X\varepsilon_y = XY + XY\left(\frac{\varepsilon_x}{X} + \frac{\varepsilon_y}{Y}\right).$$

The two factors within the brackets are the fractional errors of X and Y. Their sum is the fractional error in the product XY. The sum of the fractional errors multiplied by the product XY, as shown in the expression, will be the actual potential error in the product XY, in whatever units X and Y are measured.

The development for the quotient of X over Y produces the same result. Accordingly, the rule is: *The potential **fractional** error of the product or the quotient of quantities is the sum of the **fractional** errors of the individual quantities.*

Since percentage error is simply another form of fractional error, the words "percentage errors" could be substituted for fractional errors in the rule.

Correction Factor

A correction factor (CF) can be determined if the error is known. It is the quantity that should be added to the observed value in order to correct it to the true value. As a mathematical expression,

$$OV + CF = TV.$$

Rearranging this, OV – TV = – CF.

Recalling that the expression for the error was ε = OV – TV, then the correction factor CF must be equal to the *negative* of the error. Thus, if there is a scale that is showing a weight of 0.5 kg when it should be reading zero, then the correction factor for any reading taken from this scale should be –0.5 kg.

Significant Figures

Simply put, significant figures are numbers that actually mean what they say. It is definitely possible for numbers that have no meaning or fact to them to emerge as a result of a calculation. This is particularly true today when there are pocket calculators which will fill the whole viewing screen with numbers whether they have any meaning or not.

Suppose that the following population data is available for the City of Edmonton; the City of Ft. McMurray, 500 km to the north; and the village of Wandering River, which is halfway in between.

City of Edmonton	941,000
City of Ft. McMurray	43,900
Village of Wandering River	63

The total population of all three places calculates out to be 984,963 but obviously not all of the six figures have any real meaning.

The figure given for the City of Edmonton really specifies that the population is somewhere between 940 500 and 941 500. In other words, there will be a tolerance in the sum of ±500 persons. In view of this, the population figure for Wandering River has no meaning at all in the sum, while the hundreds digit for Ft. McMurray is relevant, in calculating the sum, only to the extent of bumping the thousands digit from 3 to 4.

When due regard is given to the real significance of the figures, the total population of the three communities should be recorded as 985 000.

The loss of meaning, or significance, always occurs in the trailing digits. The question then becomes: How many of the leading digits are significant? In the case of a sum or difference, *the last digit in the sum or difference that is significant, will be the same as the last significant digit in the greatest term in the summation.*

Until the 1960s, significant figures in products or quotients were not really a problem, since the calculations were done on sliderules, which could generate only three figure answers. Calculators that have come along since that time can create a false impression. Consider the operation of multiplying 102.7 by 3.14. If this is plugged into a calculator, the answer comes out 322.478, but how many of these figures are really significant?

If the two numbers are measurements of some kind (the decimal places suggest that they are), then the number 102.7 says that the measurement it represents lies between 102.65 and 102.75. Similarly the measurement that has a value of 3.14 lies somewhere between 3.135 and 3.145.

Multiplying the two lower values gives $102.65 \times 3.135 = 321.808$.

Multiplying the two higher values gives $102.75 \times 1.345 = 323.149$.

Thus, all that is known for sure about the product of the two numbers is that it lies between 321.808 and 323.149. Consequently, in the number cranked out by the calculator, 322.478, the numbers following the decimal point are meaningless. The only significant digits are 322.

The rule for the significant figures in a product or quotient is that *the number of significant figures in the product or quotient can be no greater than the number of significant figures in the term that has the least number of significant figures.* In the case of the example, the number 3.14 has the fewest significant figures (three), therefore no more than the first three figures in the calculated product will be significant.

Special attention should be given to the matter of trailing zeros after the decimal point. The fact that the relationship between inches and centimetres is written 1 in. = 2.54 cm does not imply that all of the figures after the 4 are zeros. However, it turns out that the next figure after the 4 is in fact a zero, so that the relationship should be recorded as 1 in. = 2.540 cm. This shows that the relationship is known to four significant figures, not three, and furthermore that the fourth significant figure is a zero.

Conversion of Units

Twice during the last few years, somewhat similar incidents were reported by the news media. An Air Canada jet liner ran out of fuel in mid air. Fortunately, the captain was one of the very few who had the necessary skill to land the plane on zero power. The problem was attributed to a mix up between pounds and kilograms of fuel.

Later, NASA crashed a space probe that was designed to land on Mars. The cost of the probe was some hundreds of millions of dollars, not counting the salaries of the highly trained technical people who had to track and guide it through its months long journey from Earth. The cause was said to be a mix up between feet and metres in the rate of descent.

Failure to convert correctly from one system of units to another can have serious consequences, enough so that some attention to a conversion procedure is justifiable. The process of converting measurements in one system of units into another system is sometimes called *scaling*. There are likely numerous methods of performing conversions. The one that will be described here, however, is straightforward and virtually foolproof. The explanation will be clearest if an example is worked.

Your neighbor has come to you with a problem. He has bought some fertilizer for his lawn The directions on the bag say to apply it at a rate of 2 kg per 100 m^2. He is in difficulty because he knows the area of his lawn is 2 560 ft^2. In addition, the only scale he has measures in pounds, not kilograms. How much fertilizer should he put on his lawn?

The first step is to convert 2 kg into pounds. Two statements are required, the first being any known *correct* relationship between pounds and kilograms. You happen to know that 1 lb = 0.453 6 kg, so you use this as the first statement. The second statement is the relationship you need to know, using an x, or a ? for the unknown quantity. The setup looks like this.

1st statement	1 lb = 0.453 6 kg
2nd statement	? = 2 kg.

When these statements are written down as shown, it is *vital* to ensure that the same units in both statements are on the same side of the equals sign. In this case, pounds are on the left side of the equals sign, kilograms on the right, in both statements. It would make no difference if this were reversed (kilograms on the left, pounds on the right), as long as consistency between the two statements is maintained.

With the two statements in place, mentally draw two diagonals across the four numbers.

$$\begin{array}{ccc} 1 \text{ lb} & = & 0.453\ 6 \text{ kg} \\ ? & = & 2 \text{ kg} \end{array}$$

The unknown quantity, (?), will always be equal to the *product* of the two numbers that lie on the path that does *not* go through the (?), divided by the single number that is on the path that includes the (?). Therefore,

$$? = \frac{2 \times 1}{0.453\ 6} = 4.41 \text{ lb}.$$

Use the same procedure to convert 100 m^2 into ft^2. You recall that 1 ft = 0.304 8 m. Therefore, 1 ft^2 = 0.304 8^2 m^2 = 0.092 9 m^2.

1st statement	1 ft^2 = 0.092 9 m^2
2nd statement	? = 100 m^2

$$\text{Therefore, } ? = \frac{100 \times 1}{0.092\ 9} = 1\ 076\ \text{ft}^2.$$

The specified application rate of 2 kg per 100 m^2 is therefore equivalent to 4.41 lb per 1 076 ft^2. For 2 560 ft^2,

$$4.41\ \text{lb} = 1\ 076\ \text{ft}^2$$

$$?\ = 2\ 560\ \text{ft}^2$$

$$? = \frac{2\ 560 \times 4.41}{1\ 076} = 10.5\ \text{lb}.$$

When this method is used, it doesn't matter which units are on which side of the equals sign. In fact, it doesn't matter if the statement of the known relationship is the first statement or the second one. What does matter is that in both statements, consistency in the placement of the units with respect to the equals sign is maintained. If this is done then a correct conversion will result.

Converting Formulas to New Units

It sometimes happens that a formula is needed to calculate some required quantity. Unfortunately, the available formula uses a system of units that is not convenient. The problem is to convert the available formula to the units that one wishes to use.

Since a mistake in converting the formula's units will inevitably lead to calculating the wrong value for the required quantity, the procedure that is used should be logical and as error proof as possible. First of all, consider these two statements, both correct, which follow.

$$1\ \text{lb} = 0.453\ 6\ \text{kg} \qquad \text{Therefore, lb} \times 0.453\ 6 = \text{kg}$$

It is the second of these two statements that is needed to convert a formula correctly. Specifically, conversion factors are needed that will convert the desired units into the units required by the formula. Correct conversion of the formula depends, therefore, on coming up with the right conversion factors.

The recommended procedure is to set up the conversion statements, similar to the two above, with the *desired* units on the *left side* of the equals sign, and the units *required* by the formula on the *right side* of the equals sign. The whole process will be illustrated best by working an example.

If the inside diameter of a pipe, the velocity of the flowing stream, and the density of the stream are known, then the mass flow rate (W) of the stream in the pipe will be

$$W\ \text{kg/s} = \pi/4\ v\ d^2\ \rho = 0.785\,4\ v\ d^2\ \rho.$$

In this expression, v is the velocity in per second, d is the inside diameter of the pipe in inches, and ρ is the density of the stream in kilograms per cubic metre. What is required is the equivalent expression with the velocity in feet per second, the inside diameter in inches, and the density in pounds per cubic foot. The starting point is to set up the conversion statements, one at a time, with the desired units on the left and the required units on the right.

For the velocity v, 1 fps = 0.304 8 m/s. Therefore fps × 0.3048 = m/s.

For the diameter d, $1\text{ in.} = \frac{1}{12} \times 0.304\,8\text{m} = 0.025\,4\text{m}$.

Therefore $\text{in}^2 \times 0.025\,4^2 = \text{m}^2$.

For the density ρ, $1\,\text{lb/ft}^3 = 0.453\,6\ \text{kg/ft}^3 = 0.453\,6 \times \frac{1}{0.304\,8^3}\ \text{kg/m}^3$

$$= 16.02\ \text{kg/m}^3.$$

Therefore, $\text{lb/ft}^3 \times 16.02 = \text{kg/m}^3$.

If v_1, d_1, and ρ_1 are the velocity in fps, the diameter in in., and the density in lb/ft^3, then the formula converts to

$$W = 0.785\,4\ (0.304\,8\ v_1)\ (0.025\,4^2\ d_1^2)\ (16.02\ \rho_1)$$

$$= 0.002\,474\ v_1\ d_1^2\ \rho_1.$$

Obviously, this result needs to be tested with some actual numbers. Suppose that in a 6 in. I.D. pipe, a fluid with a density of 50 lb/ft^3 is flowing with a velocity of 5 fps.

$$5\text{ fps} = 5 \times 0.304\,8\text{ m/s} = 1.52\,4\text{ m/s}$$

$$6\text{ in.} = 6 \times \frac{1}{12} \times 0.304\,8\text{ m} = 0.152\,4\text{ m}$$

$$50 \text{ lb/ft}^3 = 50 \times 0.453\ 6 \times \frac{1}{0.304\ 8^3} \text{kg/m}^3 = 800.9 \text{ kg/m}^3$$

Applying the original formula using the metric units,

$$W = 0.785\ 4 \times 1.524 \times 0.152\ 4^2 \times 800.9 = 22.26 \text{ kg/s}.$$

If the formula has been correctly converted, it should produce the same number, using the fps units.

$$W = 0.002\ 474 \times 5 \times 6^2 \times 50 = 22.26 \text{ kg/s}.$$

While some satisfaction can be derived from this, it points out that the job of converting the formula is not yet finished. The new converted formula can accept the fps units and give the correct answer, but it still produces the answer in kg/s, whereas the result is required in pounds per hour. The final step, therefore, is to modify the numerical factor 0.002 474 so that lb/hr are the units of the result.

$$1 \text{ kg/s} = \frac{1}{0.453\ 6} \text{ lb/s} = \frac{1}{0.453\ 6} \times 3\ 600 \text{ lb/hr} = 7\ 937 \text{ lb/hr}$$

Therefore the numerical factor for the formula should be

$$0.002\ 474 \times 7\ 937 = 19.63.$$

The final result is

W (lb/hr) = 19.63 $v\, d^2\, \rho$, with v in ft/s, d in in., and ρ in lb/ft^3.

Example 1: Reynolds Number

In the SI metric system, the basic relationship for computing the Reynolds number Re is

$$Re = \frac{v d \rho}{\mu}.$$

In this relationship,

v = the velocity of the stream in metres per second (m/s)
d = the internal diameter of the pipe in metres (m)
ρ = the density of the fluid in kilograms per cubic metre (kg/m^3)
μ = the absolute viscosity of the fluid in Pascal seconds (Pa.s)

When the SI metric relationship is used, the constant multiplying the expression is 1.00.

Starting with the basic metric relationship, the assignment is to develop a formula for the Reynolds number having the form

$$Re = C\frac{Q\,sg}{d\,\mu}.$$

where

Q = the volume flow rate of the liquid in barrels per hour (bph)
sg = the specific gravity of the liquid
d = the internal diameter of the pipe in inches (in.)
μ = the absolute viscosity of the liquid in centipoise (cps)
C = the constant that is required to accommodate the new units (to be determined).

The first step is to convert the metric formula so that it is in terms of volume flow rate and specific gravity instead of velocity and density. Volume flow rate is the product of the stream velocity and the cross sectional area of the pipe. In appropriate units,

$$Q = \frac{\pi}{4}d^2\,v,\ \text{from which}\ d\,v = \frac{4Q}{\pi d}.$$

The specific gravity $sg = \dfrac{\rho}{\rho_s}$

where ρ_s is the density of water at 15°C.

Thus, $\rho = sg \times 999.1\ kg/m^3$. Inserting these values in the original formula,

$$Re = \frac{4Q}{\pi d} \times 999.1 sg \times \frac{1}{\mu} = 1\,272\frac{Q\,sg}{d\,\mu},\ \text{all in metric units.}$$

The next step is to convert the individual terms, bearing in mind that the procedure calls for keeping the desired units on the left side of the equals sign.

$$1\ bbl = 159.0\ dm^3 = 0.159\ m^3$$

$$1\ bph = 0.159 \times \frac{1}{3\,600}m^3/s = 0.000\,044\,17\ m^3/s$$

$$\therefore\ bph \times 0.000\,044\,17 = m^3/s$$

$$1\ in. = \frac{1}{12}ft = \frac{1}{12} \times 0.304\,8\ m = 0.025\,4\ m$$

$$\therefore \text{ in.} \times 0.025\,4 = \text{m}$$

$$1 \text{ cps} = 1 \text{ mPa.s} = 0.001 \text{ Pa.s}$$

$$\therefore \text{ cps} \times 0.001 = \text{Pa.s}$$

Substituting these values into the metric relationship for Re,

$$\text{Re} = \frac{1\,272 \times 0.000\,044\,17\,Q \times sg}{0.025\,4\,d \times 0.001\,\mu} = 2\,212\frac{Q\,sg}{d\mu}\text{, in the desired units.}$$

This result should be tested with some actual data. In the test application, a liquid stream with a specific gravity of 0.80 and a viscosity of 0.70 cps is flowing through a pipe with an internal diameter of 6 in. at a rate of 500 bph. The Reynolds number computes to be,

$$\text{Re} = \frac{2\,212 \times 500 \times 0.80}{6.0 \times 0.70} = 211\,000\,.$$

Verifying this value for Re,

Flow rate $Q = 500 \times 0.000\,044\,17 = 0.022\,09$ m^3/s

Density $\rho = 0.80 \times 999.1 = 799.3$ kg/m^3

Viscosity $\mu = 0.70$ cps $= 0.70$ mPa.s $= 0.000\,70$ Pa.s

Diameter $d = 6 \times 0.0254 = 0.152\,4$ m

$$\text{Velocity} = \frac{Q}{\text{Area}} = \frac{0.022\,09}{\frac{\pi}{4} \times 0.152\,4^2} = 1.211 \text{ m/s}$$

Inserting these values in the original metric formula,

$$\text{Re} = \frac{1.211 \times 0.152\,4 \times 799.3}{0.000\,70} = 211\,000\,.$$

Example 2: Pressure of a Column of Liquid

One situation that comes up frequently concerns the pressure built up by a head of liquid. When a textbook formula is found, it usually does not employ the most convenient units. The basic relationship in the SI system is

$$p = \rho\, g\, h.$$

The pressure is in Pascals when ρ is in kg/m^3, g is in m/s^2, and h is in m.

The formula desired is one that gives the pressure (p) in psi, with the density expressed as specific gravity (sg), and the head of liquid (h) in feet.

By definition,

$$\text{specific gravity} = \frac{\text{Density of the liquid } (\text{kg}/\text{m}^3)}{\text{Density of water at } 15^{\text{O}}\text{ C } (\text{kg}/\text{m}^3)} = \frac{\rho}{999.1}.$$

$$\therefore \rho = 999.1 \times \text{sg}$$

Also, in the SI system, g = 9.812 m/s^2, and p (kPa) = p (Pa) ÷ 1000.

Substituting these facts into the basic formula gives,

$$\text{p (kPa)} = \frac{(999.1 \times \text{sg}) \times 9.812 \times \text{h}}{1000}$$

$$= 9.803 \times \text{sg} \times \text{h (m)}.$$

The formula is required to accept h in feet. Following the procedure:

$$1 \text{ ft} = 0.304\,8 \text{ m}. \quad \therefore \text{ ft} \times 0.304\,8 = \text{m}$$

$$9.803 \times \text{sg} \times \text{h (m)} = 9.803 \times \text{sg} \times 0.304\,8 \times \text{h (ft)} = 2.988 \text{ sg h (ft)}.$$

For the resulting pressure to be in psi, not kPa:

$$1 \text{ psi} = 6.895 \text{ kPa}$$

$$? = 2.988 \text{ sg h kPa}$$

$$? = \frac{1 \times 2.988 \text{ sg h}}{6.895} = 0.433 \text{ sg h psi}\cdot$$

Finally, p (psi) = 0.433 sg h (ft).

The Most Representative Value (MRV)

Sometimes a number of measurements of the same entity are made, possibly at different times. This leads to the question: What one value could be chosen as being the most representative value for the entity of concern?

The daily price of gasoline, in cents per litre, over a 15 day period, was noted to be:

62.5	64.5	65.5	66.5	61.5
61.5	62.5	66.5	67.5	67.5
68.5	68.5	66.5	65.5	63.5

Which value would be most representative of the price of gasoline over the 15 day period? There are actually four possibilities.

Average Value

The average value, or what the statisticians call the *arithmetic mean*, is probably the most familiar. To calculate the average, the 15 readings are summed, and the sum is divided by the number of readings, i.e. 15. For this example, the average turns out to be 65.2¢/*l*.

Although the average value is often used as the most representative value, there is obviously a problem in this case because the value 65.2 does not appear anywhere in the list of readings. This raises some doubt as to its appropriateness as the representative value.

Weighted Average

When a number of readings are taken, it may happen that some readings will be more relevant than others. At the beginning of the year, one might be attempting to estimate how much of one's income will go into paying for natural gas. The cost of gas for each of the last five years is available.

A simple average may not be as accurate as a weighted average. It would be more practical to assume that the cost of gas in more recent years should be given greater influence than the cost in earlier years. If G5, G4, G3, G2, and G1 were the yearly natural gas costs for each of the last five years, C1 being the cost for the year just ended, then the weights might logically be assigned as follows.

$$G5 \times 1 \qquad G4 \times 2 \qquad G3 \times 3 \qquad G2 \times 4 \qquad G1 \times 5$$

In this weighting, the cost of gas for the year just ended would receive five times the emphasis as the cost five years ago. Giving one of the readings a weight of five is equivalent to putting that reading into the sum five times rather than once, and this fact must be taken into account in the denominator of the expression. Consequently, the mathematical expression for the weighted average will be:

$$\text{Weighted Average} = \frac{G5 \times 1 + G4 \times 2 + G3 \times 3 + G2 \times 4 + G1 \times 5}{1 + 2 + 3 + 4 + 5}$$

For the gasoline prices, it may be considered that the latest five readings should receive twice the weight of the first five, and the second group of five readings should have 1.5 times the weight of the first five readings.

The sum of the first five readings is 320.5.

The sum of the second five readings is 325.5.

The sum of the last five readings is 332.5.

The weighted average equals

$$\frac{320.5 \times 1 + 325.5 \times 1.5 + 332.5 \times 2}{5 \times 1 + 5 \times 1.5 + 5 \times 2} = \frac{1\,473.75}{22.5} = 65.5\,.$$

The fact that the digit after the decimal place turned out to be a 5, the same as for the individual readings, is just a coincidence. If any of the 15 readings had had a different value, then the third digit in the weighted average would have been something other than a 5.

Mode

The *mode* is the reading that appears most often in the set of readings. In the case of the gasoline prices, the reading 66.5 appears the most times (3), and is accordingly the mode. In this situation, a case could be made for promoting 66.5¢/*l* as the most representative value.

Median

The *median* is the central value in the set of readings. To find the median, first set down the readings in descending order of magnitude. Then cross off the highest and lowest readings, then the next highest and next lowest, and so on, until there is only one reading left. That reading will be the median. Its significance is that there will be as many readings that are greater than the median, as there are less than the median.

For the set of 15 gasoline prices, the median turns out to be 65.5¢/*l*.

One data assembling situation in which the median is often used is in the recording of salaries. If a person is earning the median salary, then half of the people reporting are earning less than he or she is, and the other half are earning more. The compilers of the data will sometimes list the salaries reported in descending order of magnitude, and then divide the list into 10 equal groups. When this is done the groups are called *deciles*. The desirable situation, of course, is for one to be in the upper decile. If the same list is divided into 4 groups instead of 10, then the groups are called *quartiles*.

In general, there are no hard and fast rules that decide which of the four possibilities should be chosen as the most representative value. The particular situation itself governs.

Predicting Future Values

Sometimes the purpose of determining the most representative value is to estimate what that value will be at some future time. Preparing a budget is a prime example of this task. For example, suppose that a car owner has the following data for what he has spent on gasoline over the last five years.

Table 11-1. History Data for Predicting Future Values

Years Ago	$ Spent for Gasoline
5	850
4	680
3	810
2	1070
1	1060

Based on these numbers, how much should be budgeted for gasoline this year?

From the appearance of the numbers, the best and easiest way out of the dilemma would be to determine the weighted average, which turns out to be $950. However, to get some idea of any trend that may be developing, it is necessary to plot the values on a time base, as in Figure 11-1. In this graph, the negative signs indicate years in the past.

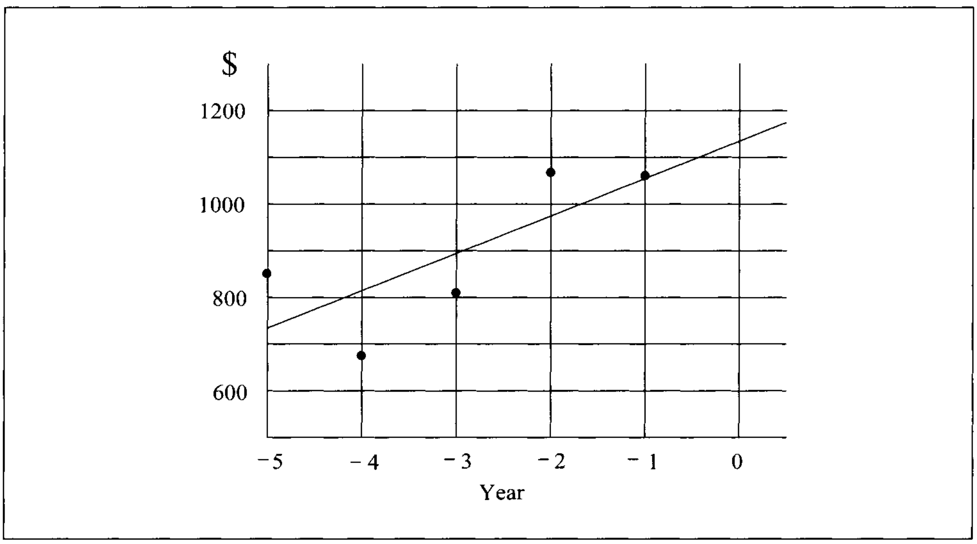

Figure 11-1. Estimating the Cost of Gasoline for the Coming Year.

Unfortunately, the way the points are dispersed makes it difficult to eyeball in the best straight line, but help is on the way by virtue of a mathematical procedure that involves the use of calculus, but which we need not go into here. The equation of the best line will have the form y = a + bx. The procedure assumes that as far as the various points are concerned, the x values are correct (no error), but the y values deviate from their logical values represented by the best line. Expressions can then be devised for the constants a and b based on minimizing the sum of the squares of all of the y deviations.

These expressions turn out to be:

$$a = \frac{\sum x \sum xy - \sum y \sum x^2}{(\sum x)^2 - n\sum x^2}, \text{ and } b = \frac{\sum x \sum y - n\sum xy}{(\sum x)^2 - n\sum x^2}$$

where n is the number of points, and Σ is the summation sign.

This theory should be applied to the data on gasoline costs. From the five data points obtained:

Table 11-2. Predicting Future Values Example

	x	y	xy	x^2
	–5	850	–4 250	25
	–4	680	–2 720	16
	–3	810	–2 430	9
	–2	1 070	–2 140	4
	–1	1 060	–1 060	1
Σ	**–15**	**4 470**	**–12 600**	**55**

$$\therefore a = \frac{[-15 \times -12\ 600] - [4\ 470 \times 55]}{[-15^2] - [5 \times 55]} = 1\ 137$$

$$b = \frac{[-15 \times 4\ 470] - [5 \times -12\ 600]}{[-15^2] - [5 \times 55]} = 81$$

The equation for the best line is consequently y = 1 137 + 81 x.

The purpose of obtaining the equation for the best line was to estimate the value of y, the cost of gasoline, at x = 0, which corresponds to the current year. When x = 0, y is equal to $1 137, although judging from the original data from which the line was determined, four significant figures are hardly justified.

In summary, using the data on the cost of gasoline over the previous five years was useful for demonstrating the procedure for establishing the best straight line through a collection of points. However, as a means of estimating how much gasoline is going to cost in the current year, it is obviously a less reliable method than say, determining the weighted average. The best straight line method using least squares is better applied to collections of points that the laws of nature say *should* lie on a straight line, but do not, due to inaccuracies in the data.

If the best line procedure is used, one should take care to distinguish between Σx^2 and $(\Sigma x)^2$.

How Much Confidence in the Most Representative Value?

It often happens that the most representative value (MRV) for a particular entity has to be determined from a number of measurement readings, with the added complication that not all of the readings are the same. Differences in the readings can occur for a number of reasons.

- The entity varies from time to time, as in the case of gasoline prices, or the outdoor temperature, but one reading has to be selected for the MRV to make cost or other projections.
- Not all of the readings are taken by the same person, and there is a question of skill involved.
- Not all of the readings are taken using the same type or quality of equipment, and there is a question of the potential error or the reliability.

The result is that when one number is designated to be the MRV out of a group of readings representing the same entity, the question then arises: How much assurance can one have in the accuracy of the MRV?

A possible answer lies in determining how closely the various readings are grouped around the MRV, or conversely, how badly they are scattered. In this procedure, the MRV is, by definition, the arithmetic mean (AM) or average. For the set of gasoline prices (P1 to P15), the AM is 65.2¢/*l*. The deviations of the individual readings are designated D1, D2,...D15, where

$$D1 = P1 - AM, D2 = P2 - AM... D15 = P15 - AM.$$

All deviations (D1 to D15) are considered to be positive even if the AM is greater than an individual price reading. The next step is to calculate the *average deviation*, which is the sum of all the deviations divided by the number of readings. For the gasoline price example, the average deviation

$$\begin{aligned} AD &= \frac{2.7+0.7+0.3+1.3+3.7+3.7+2.7+1.3+2.3+2.3+3.3+3.3+1.3+0.3+1.7}{15} \\ &= 2.1. \end{aligned}$$

The average deviation, in itself, is a fairly good indicator of the confidence one can place in the arithmetic mean as the MRV. If the price of gasoline is stated to be 65.2 ± 2.1¢/*l*, it implies that the price posted on any particular day has a 50% chance of being in the bracket 63.1 to 67.3¢/*l*.

The Standard Deviation

The standard deviation is a number developed by statisticians to indicate the extent of the dispersion of the readings around the arithmetic mean. The standard deviation is not the same as the average deviation, although the average deviation is involved in the calculation of the standard deviation.

Assuming that the average deviation AD has been determined as described already, then the standard deviation SD is equal to

$$\sqrt{\frac{\sum(D-AD)^2}{n-1}}.$$

This is the abbreviated version, in which Σ is the summation operator, D stands for the deviations of the individual readings from the AM, AD is the average deviation, and n is the number of readings. In the long form,

$$SD = \sqrt{\frac{(D1-AD)^2+(D2-AD)^2+(D3-AD)^2+\ldots..+(D15-AD)^2}{n-1}}.$$

For the gasoline price situation, the standard deviation computes to be 1.7.

Statistically, all of the readings that were involved in the calculation should be not more than three times the standard deviation different from the arithmetic mean. If any individual reading has a deviation greater than this from the AM, it is assumed to be invalid and it is scrubbed from the list. The calculation is then redone, from the start, using the remaining readings. This will generate a new value for the standard deviation and will require a second check of the validity of the readings. For the example, all 15 readings were within 3 × 1.7 = 5.1¢/*l* of the AM, 65.2¢/*l*, and are accordingly accepted as valid.

The statistical significance of the standard deviation is that of all of the readings that were taken, or which will subsequently be taken, 68% of the readings will lie within the bracket AM plus or minus the standard deviation, 95% will be in the bracket AM plus or minus twice the standard deviation, and 99% will be in the bracket AM plus or minus three times the standard deviation.

Stated in another way, if the standard deviation can be determined from a set of readings, then it can be expected that 99% of the time, future readings will be within three times the standard deviation of the AM, 95% of the time they will be within twice the standard deviation of the AM, and 68% of the time they will be within the bracket AM plus or minus the standard deviation. This implies a certain level of assurance.

When the MRV for an entity is stated, it is usually important to know the confidence level that should be given to the statement. If, for example, the price of gasoline is said to be 65.2¢/*l* at the 99% confidence level, it will mean that the potential error should be considered to be plus or minus three times the standard deviation.

The validity of this statistical theory depends heavily on the number of readings on which the calculations are based. Generally, the more, the better, although a point of diminishing returns can be reached. Fewer than 10 readings in most cases are inadequate.

Curve Fitting

It sometimes happens that a computer needs to know the relationship between two variables, but the relationship unfortunately does not follow any particular mathematical law. That is, the relationship is not parabolic, hyperbolic, sinusoidal, exponential, logarithmic, or whatever. Physical properties of naturally occurring substances often fall into this category. Enough data are available that the dependent variable can be plotted against the independent variable, but unfortunately, computers are not very good at reading graphs.

The graph in Figure 11-2 shows the variation of the vapor pressure of water with temperature. The problem is, given any temperature between 0 and 100°C, to have a computer come up with the correct value of the vapor pressure.

This can be done through the use of a curve fitting program. The general equation for the curve fit has the form

$$y = f(z) = I_0 + I_1 z + I_2 z^2 + I_3 z^3 + I_4 z^4$$

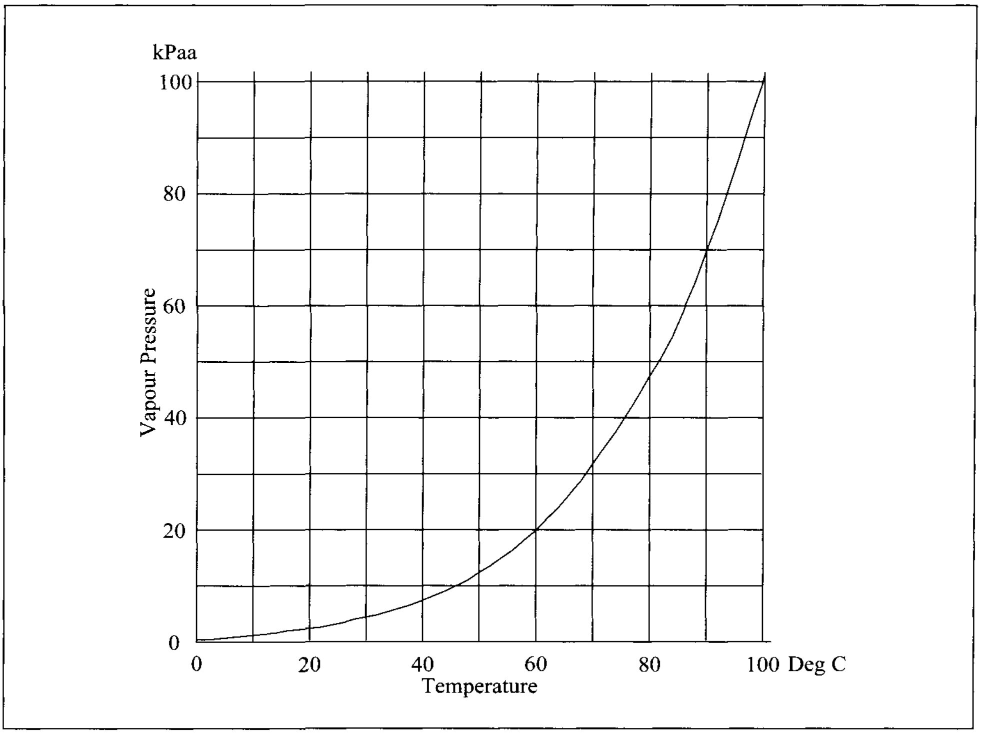

Figure 11-2. Variation of the Vapor Pressure of Water with Temperature.

where the I's are appropriate constants to be determined, and z is an intermediate variable, which will be clarified shortly.

The procedure is as follows.

1. Divide the range of the independent variable (x) into four equal sectors. The division points are identified as x_0 (the starting point), x_1, x_2, x_3, and x_4 (the end point). If each sector is w units wide, in units of the independent variable, then

$$w = \frac{x_4 - x_0}{4}.$$

2. Record the values of the independent variable (y), which correspond to x_0, x_1, x_2, x_3, and x_4. Call these y_0, y_1, y_2, y_3, and y_4 The required values of the constants I_0, I_1, I_2, I_3, and I_4 will then be:

$$I_4 = \frac{y_4 - 4y_3 + 6y_2 - 4y_1 + y_0}{24}$$

$$I_3 = \frac{y_3 - 3y_2 + 3y_1 - y_0}{6} - 6I_4$$

$$I_2 = \frac{6y_2 - 9y_1 + 4y_0 - y_3}{6} - 2I_3 - I_4$$

$$I_1 = y_1 - y_0 - I_2 - I_3 - I_4$$

$$I_0 = y_0.$$

3. The intermediate variable

$$z = \frac{x - x_0}{w}$$

where x is the value of the independent variable at which the value of the dependent variable y is to be calculated. Then,

$$y = I_0 + I_1 z + I_2 z^2 + I_3 z^3 + I_4 z^4.$$

Table 11-3 shows how well the curve fit values compare with the actual vapor pressure values. The curve fit graph has not been plotted in Figure 11-2, since except for a small region around 10°C, it coincides almost exactly with the real vapor pressure curve.

Note that the curve fit values are dead accurate at the 0, 25, 50, 75, and 100% points. This is an inherent characteristic of the curve fit program and will apply any time it is used. If greater accuracy is required, the graph can be divided into two sections and a curve fit equation developed for each section. The combination will then be exactly accurate at nine points along the graph instead of five.

The curve fit program can be applied to graphs that change direction, such as process reaction curves, which have the characteristic S shape. The accuracy of the fit in these cases is usually quite good.

An important word of caution: The curve fit equation should never be extrapolated outside of the specified lower and upper limits (x_0 and x_4). The accuracy falls off rapidly outside of these limits.

Table 11-3. The Curve Fit Program

Temperature (x) Deg C	Vapor Pressure kPa	Curve Fit Value (y)
0	0.60	0.60
10	1.22	1.13
20	2.33	2.30
25	3.16	3.16
30	4.23	4.26
40	7.35	7.38
50	12.29	12.29
60	19.85	19.82
70	31.06	31.04
75	38.43	38.43
80	47.20	47.25
90	69.89	70.00
100	101.03	101.03