Appendix **F**

Treatment of Numerical Data

I. Recording Data

When numerical data is recorded, three kinds of information must be conveyed: the magnitude of the number, how well the number is known, and the units used in making the measurement. The magnitude of the number simply specifies "how much"; for example 4, 40, or 400. The units could be 4 °C or pH of 4, 40 milliamps or 40 drops, 400 milliliters or 400 grams. Units must be specified.

Proper use of significant figures relays how well a number is known. If an instrument has a digital readout, all numbers displayed are recorded and are significant. For example, the display on an analytical balance might read 21.3480 g, with six significant figures. The mass displayed for the same object weighed on a top loading balance would be 21.35 g, with four significant figures. The uncertainty is assumed to be ±1 in the right-most digit (e.g., 0.0001 g on an analytical balance).

When an instrument has a scale, the significant figures are all of the digits we know exactly plus one that we estimate. Thus, if a scale is marked off in whole units, we estimate and record in the tenths place. If the scale is marked in tenths, we estimate in the hundredths place. The same is true for volumetric equipment. If a volume of exactly 15 mL is measured in a 50-mL beaker marked off in 10-mL increments, the volume would be recorded as 15 mL, with two significant figures. If a 25-mL buret is used, markings at 0.1 mL intervals permit estimating into the hundredths place, and the volume is recorded as 15.00 mL, with four significant figures. The last digit in both is an estimate and could not be altered after the measurement was taken.

Notice the zeros in 40, 400, and 15.00. Are they significant? Sometimes zeros are used as "placeholders"—to establish the magnitude of the number. At other times, a zero is written because the number in that place is known to be zero. Some rules are useful in writing numbers and interpreting written numbers.

1. Zeros to the right of a nonzero digit and to the right of the decimal point are significant. The number 15.00 has four significant figures.

2. Zeros between significant figures are significant. All of the following numbers have four significant figures.

1005	10.50	150.0	1.005

3. Leading zeros are not significant. These are placeholders to the left of the first nonzero digit. The following numbers all have three significant figures.

0.0400	0.000650	0.00605

4. A whole number that ends in zero, such as 40 or 400, is ambiguous. It is not clear whether the zeros are placeholders or known quantities. It is generally assumed that they are placeholders and are not significant. To remove the ambiguity, scientific notation should be used. In scientific notation, significant figure information is in the coefficient and decimal point information is in the exponent.

If 400 means	*Write*	*Number of Significant Figures*
a number between 300 and 500	4×10^2	one
a number between 390 and 410	4.0×10^2	two
a number between 399 and 401	4.00×10^2	three

Some numbers are *exact*; they have no uncertainty and have an infinite number of significant figures. Numbers can be exact by count or by definition. The 40 drops mentioned previously are exact by count. Numbers used in converting units within the metric system are exact by definition. There are exactly 1000 mL in a liter. There are exactly 1000 g in a kilogram.

II. Calculations Using Significant Figures

When a measured quantity is used in a calculation, the resulting number must still reflect the precision of the measured quantity. A mathematical operation can neither improve a measured number nor make it less precise. There are two rules for using significant figures in calculations. One is for addition and subtraction; the other is for multiplication and division.

THE ADDITION-SUBTRACTION RULE

The right-most digit to be retained is in the last place common to all numbers. The following examples illustrate this rule.

(a)		(b)		(c)	
	34.060		439.5		6.845
	4.2		27		$-$ 5.92
	$+$ 81.57		$+$ 6.01		0.925
	119.830		472.51		rounds to 0.92
	rounds to 119.8		rounds to 473		

In (a), the last place common to all numbers is the tenths place. The rounding rule used here is: If the left-most number to be discarded is less than 5, round down—the extra numbers are dropped. In (b), the last place common to all numbers is the ones place. The rounding

rule here is: If the left-most number to be discarded is greater than 5, or 5 followed by nonzero digits, round up—increase the last digit to be retained by one. In (c), the last place common to both numbers is the hundredths place. The rounding rule here is: If the number to be discarded is 5 or 5 followed only by zeros, round even (your instructor may not be using this rule in lecture). The last number to be retained is unchanged if it is even; it is increased by one if it is odd (i.e., the last number to be retained is rounded even). Rounding the numbers below to two significant figures illustrates the "round even" rule further.

2.35	rounds to 2.4
2.45	rounds to 2.4
2.4523	rounds to 2.5
2.550	rounds to 2.6

THE MULTIPLICATION-DIVISION RULE

The answer has the same number of significant figures as the quantity with the least number of significant figures. The following examples illustrate this rule.

(a) $645 \cdot 0.13 = 83.85$ rounds to 84

(b) $645 \div 12.90 = 50$ correctly reported as 50.0 or 5.00×10^1

In (a), the quantity with the least number of significant figures is 0.13 with two. The answer must have two significant figures. In (b), the quantity with the least number of significant figures is 645 with three. To show three significant figures, the answer of 50, which appears on the calculator, must be rewritten.

If a calculation requires a combination of processes, the appropriate rule is applied to each process following the order of operations. The following examples illustrate this statement.

(a) $(12.39 - 5.8147) \div 9.317 = 0.706$

(b) $(12.39 \cdot 0.02317) + 5.8147 = 6.1018$

In (a), the answer to the subtraction is good to the hundredths place, resulting in three significant figures. The result of the division and the final answer must therefore have three significant figures. In (b), the answer to the multiplication has four significant figures, good to the fourth place past the decimal point. Since the number being added is also good to the fourth place, the digit in the fourth place in the answer is also significant.

In addition to these two rules, significant figures apply in logarithms. When taking the logarithm of a number, such as finding pH value from the concentration of H^+, the significant information is displayed after the decimal point. This calculation—a common base 10 log—is defined as "a quantity representing the power 10 must be raised to equal a desired value." For example,

$\log 1500 = 3.1\underline{7}60$, therefore $10^{3.1\underline{7}60} = 1500$

In the above example, you will notice that (while all digits are kept during the calculation) the 7 is underlined as the last significant digit. Why is this the case, if the number 1500 only contains 2 significant figures and 3.17 is 3? A logarithm is actually composed of two parts, the **characteristic** (number before the decimal place) and the **mantissa** (the numbers following the decimal place). Since the characteristic only tells us the order of magnitude, it is

not considered a significant digit. Therefore, only the mantissa contains significant figure information. In other words, if you take the log of a number with N significant figures, the log should have N decimal places.

$$[H^+] = 0.056 \text{ M}$$

$$pH = -\log[H^+] = -\log [0.056] = 1.2\underline{5}1812 = 1.25$$

In this example, the original value, 0.056, has 2 significant figures (N = 2). Therefore, two decimal places are considered significant, and the value reported is 1.25 (N = 2 decimal places). Note: you can not take the log of a unit, and therefore the resulting value does not have a unit.

In all mixed operation calculations, all numbers should be carried through the calculation and the answer correctly rounded at the end. Rounding at each step can produce significant rounding errors.

III. Reporting Results

The determination of a numerical value is rarely done with one experiment; two, three or more trials are performed. The result of N trials is generally reported as an *average*.

$$\text{average} = \frac{\text{sum of results of N trials}}{N}$$

If results of a molecular weight determination are 45.2, 31.8, and 49.7 g/mol, the average is:

$$\text{average} = \frac{45.\underline{2} + 31.\underline{8} + 49.\underline{7}}{3} = 42.\underline{2}3333 = 42.2 \text{ g/mol}$$

To use Microsoft Excel to average a set of data, use the "AVERAGE" function.

Note: when averaging values, the calculated average cannot exceed the precision of the instrument used to obtain the data.

To express the *precision* of the measurements, how close the determinations are to each other, we calculate a deviation and an average deviation. The deviation is the difference between an individual measurement and the average.

$$\text{deviation, } \delta = \text{individual measurement} - \text{average}$$

For the numbers given above, the deviations are:

$$\delta = 45.\underline{2} - 42.\underline{2}3 = 2.\underline{9}6667 = 3.0$$
$$\delta = 31.\underline{8} - 42.\underline{2}3 = -10.\underline{4}3333 = -10.4$$
$$\delta = 49.\underline{7} - 42.\underline{2}3 = 7.\underline{4}6667 = 7.5$$

The *average deviation* is calculated using absolute values, symbolized by enclosing the quantity between vertical lines.

$$\text{average deviation} = \frac{|\delta_1| + |\delta_2| + |\delta_3| + \cdots + |\delta_N|}{N}$$

For the numbers above:

$$\text{average deviation} = \frac{|2.\underline{9}6667| + |10.\underline{4}3333| + |7.\underline{4}6667|}{3} = \frac{20.\underline{8}6667}{3} = 6.\underline{9}5556$$

The average deviation indicates uncertainty in the units place, so the answer should be reported as 42 ± 7 g/mol. Note that all intermediate digits were kept during the calculation.

IV. Identifying and Handling Outliers

You might have noticed that one of the values in the previous example is significantly different from the other two. Are we justified in ignoring the value of 31.8 g/mol in calculating the average? To answer that question, we calculate a *standard deviation*, σ. Standard deviation measures how closely the data clusters around the average.

$$\sigma = \sqrt{\frac{\delta_1^2 + \delta_2^2 + \delta_3^2 + \cdots + \delta_N^2}{N-1}}$$

For the numbers above, the standard deviation is:

$$\sigma = \sqrt{\frac{(2.96667)^2 + (-10.43333)^2 + (7.46667)^2}{2}} = \sqrt{\frac{8.80113 + 108.85437 + 55.75116}{2}}$$

$$= \sqrt{\frac{173.40667}{2}} = \sqrt{86.7033} = 9.31146$$

Note that all digits were kept during the intermediate calculations. Note the final value is reported to two significant figures because it cannot exceed the precision of the instrument used to obtain the data. To use Microsoft Excel to find the standard deviation of a set of data, use either the "STDEV.S" function or the "STDEV" function.

In this course, if the deviation for an individual measurement is **greater than** the standard deviation, that measurement can be discarded. Trial 2, with a deviation of 10.4, can be ignored in calculating the average. Recalculation of the average and average deviation produces the following:

$$\text{average} = \frac{45.2 + 49.7}{2} = 47.45$$
$$\delta = 45.2 - 47.45 = -2.25$$
$$\delta = 49.7 - 47.45 = 2.25$$
$$\text{average deviation} = \frac{2.25 + 2.25}{2} = 2.25$$

The number would be reported as 47 ± 2 g/mol.

The relative deviation or relative error expresses the deviation or error in relation to the average or true value. For the molecular weight example discussed above, the relative deviation, expressed as a percent, is:

$$\text{relative deviation} = \frac{\text{average deviation}}{\text{average}} \cdot 100 = \frac{2.25}{47.45} \cdot 100 = 4.74684\%$$

If the true value is known to be 47.9 g/mol, the percent error can be calculated:

$$\text{percent error} = \frac{\text{experimental value} - \text{true value}}{\text{true value}} \cdot 100 = \frac{47.45 - 47.9}{47.9} \cdot 100 = -0.93946\%$$

The negative sign indicates that the experimental value is underestimating the true value. Percent error is sometimes reported as the absolute value, 1% in this case. However, as the sign tells us some information about the data, we often use it in the lab.

V. Calculating Slope Using a Graph

Table F.1 Variation of concentration of reactant A with time.

time, seconds	[A] (M)	*In* [A]
0	0.1000	−2.3026
2.00×10^2	0.0735	−2.610
4.00×10^2	0.0488	−3.020
6.00×10^2	0.0340	−3.381
8.00×10^2	0.0230	−3.772
1.000×10^3	0.0166	−4.098
1.200×10^3	0.0116	−4.457

Enclosing A in square brackets stands for the molar concentration of A. The subscripts t and 0 indicate "at time t" and "at time zero."

The equation of a straight line,

$$y = mx + b$$

where m is the slope of the line and b is the y-intercept, the point at which the line crosses the y-axis. This equation is often used to find slope for a number of physical processes. The example graph below is for the equation,

$$ln[A]_t = -k_t + ln[A]_0$$

If $ln[A]_t$ is graphed on the y-axis and t is graphed on the x-axis, a straight line should be produced with a slope of −k and a y intercept of $ln\,[A]_0$. This graph has been constructed in Figure F.1, using information from Table F.1. The x-axis is the time, in seconds, with scale divisions of 10 seconds. The y-axis is $ln[A]$, and the numbers plotted on the y-axis are between −2.30 and −4.50 so zero does not appear on the scale for the y-axis. Negative numbers are still located so that they are increasingly positive going up. Scale divisions are 0.025.

The value of the constant k is characteristic of the relationship between concentration of A and time. It is easily determined from the slope of the graph.

$$-k = slope = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

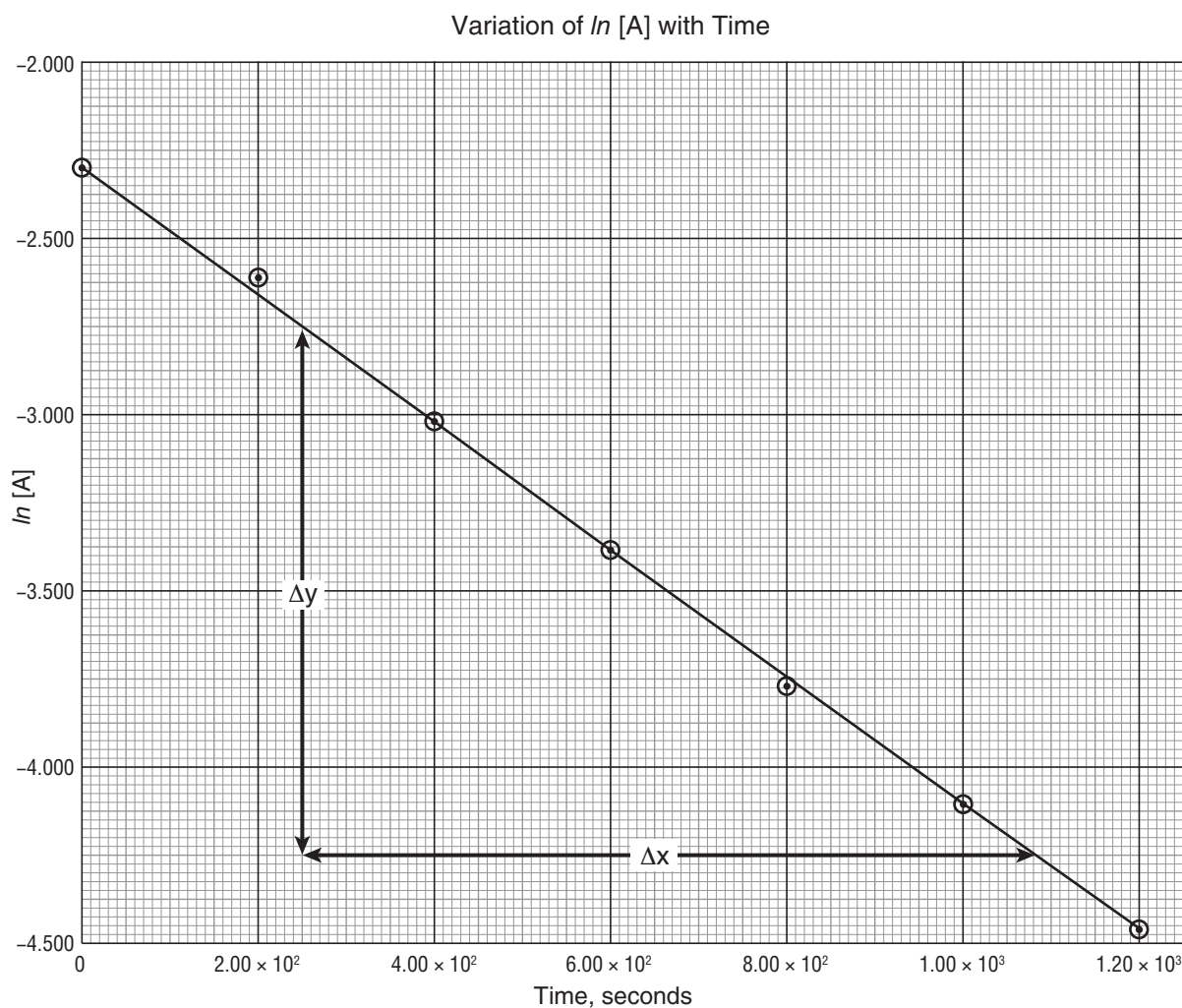Variation of *In* [A] with Time

Figure F.1 Graph of *In* [A] versus time.

To calculate the slope, use two points *on the line*, not two experimental data points. Choose points that are far apart from each other where the line crosses an intersection of grid lines. The slope of the line in Figure F.1 is calculated as follows.

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{-4.250 - (-2.750)}{(1.080 \times 10^3 - 2.50 \times 10^2)\,\text{s}} = \frac{-1.500}{8.30 \times 10^2\,\text{s}} = -1.81 \times 10^{-3}\,\text{s}^{-1}$$

The value of k is 1.81×10^{-3} s^{-1}.

If the two experimental data points that the trend line does not cross through had been chosen to calculate the slope, the value would be

$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{-3.772 - (-2.610)}{(8.00 \times 10^2 - 2.00 \times 10^2)\,\text{s}} = \frac{-1.162}{6.00 \times 10^2\,\text{s}} = -1.94 \times 10^{-3}\,\text{s}^{-1}$$

The percent deviation is

$$\frac{-1.94 \times 10^{-3} - (-1.81 \times 10^{-3})}{-1.81 \times 10^{-3}} \times 100 = 7.2\%$$

Using points from the "best fit" straight line reduces the effect of random errors in individual data points. While this method accurately calculates the slope between two data points, it ignores the rest of the data set, making the value inaccurate.

In this course, you may be asked to show a sample calculation for slope using the data you collected. If that slope is used for further calculations, you should proceed with the slope from Excel. Be sure to indicate you are proceeding with the slope from Excel.