Chapter 2

Data Handling

- Sample size, accuracy, precision and number of analyses needed.

2.1 Accuracy and Precision

Accuracy: the degree of agreement between the measured value and the true value

Precision: the degree of agreement between replicate measurements of the same quantity

A systematic error does not affect the precision, but it does affect the accuracy.

The higher the degree of precision, the greater the chance of obtaining the true value.

2.2 Significant Figures

Ex:

0.92067 → five

0.092067 → five

9.3660×10^5 → five

936600 → four

7.270 → four

- Multiplication and Division

  Key number: the one with the least number of significant figures.

  Ex: 2.3 , 2.4

2.2-3

- Addition and Subtraction

  The answer is the same number of units as the number containing the least significant unit.

  The overall accuracy can be improved if desired ether by making the key number larger, or by making the measurement to an additional figure if possible.

- Logarithms :

  The mantissa determines the number of significant figures.

  Ex :

  $[H^+]=2.0\times10^{-3}$

  $pH=-\log 2.0\times10^{-3}=-(-3+0.30)=2.70$

  antilogarithm of 0.072 $\Rightarrow$ 1.18

  logarithm of 12.1 $\Rightarrow$ 1.083

2.3 Rounding Off

9.47 → 9.5

8.65 → 8.6

9.43 → 9.4

8.75 → 8.8

8.55 → 8.6

2.4 Determinate Errors (systematic errors)

Errors that are determinable and that presumably can be either avoided or corrected.

The variation is unidirectional.

Ex:

1) Instrumental errors
2) Operative errors
3) Errors of the method

How to reduce: run a reagent blank

2.5 Indeterminate errors (accidental or random errors)

They cannot be predicted or estimated.

They will follow a random distribution.
(Ex: Gaussian curve)

2.6 Ways of Expressing Accuracy

- Absolute Errors:
  (the measured value - the true value)
  mean error
- Relative Errors:

$$\frac{\text{absolute error or mean error}}{\text{true value}} \times 100\%$$

Ex : 2.6

2.7 Standard Deviation: an indication of the precision of the analysis

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

$$\bar{x} = \sum \left(\frac{x_i}{N} \right)$$

μ : mean value when $N \to \infty$

$\bar{x}$: mean value when $N \not\to \infty$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}}$$

$$s = \sqrt{\frac{\sum x_i^2 - \left(\sum x_i\right)^2 / N}{N - 1}}$$

s : estimated standard deviation

N – 1 : degrees of freedom

The arithmetical mean derived from N measurements can be shown to be $\sqrt{N}$ times more reliable than a single measurement.

Standard deviation of the mean

$$= s(\text{mean}) = \frac{s}{\sqrt{N}} = \text{standard error}$$

coefficient of variation : $\frac{s}{\bar{x}} \times 100\%$

or relative standard deviation :

The precision of a measurement can be improved by increasing the number of observations.

Variance : s^2

2.8 Propagation of Error

- Addition and Subtraction:

  For addition and subtraction, absolute uncertainties are additive.

  The absolute variance of the answer is the sum of the individual variances.

  $s_a^2 = s_b^2 + s_c^2 + s_d^2$ for $a = b + c - d$

  $s_a = (s_b^2 + s_c^2 + s_d^2)^{½}$

  Ex: 2.10

- Multiplication and Divison

The relative uncertainties are addtve

for a = bc/d

$$(s_a^2)_{rel}=(s_b^2)_{rel}+(s_c^2)_{rel}+(s_d^2)_{rel}$$

$$\Rightarrow (s_a)_{rel}=[(s_b^2)_{rel}+(s_c^2)_{rel}+(s_d^2)_{rel})]^{1/2}$$

Ex: 2.11, 2.12*

2.9 Significant Figures and Propagation of Error

The number of significant figures in an answer is determined by the uncertainty due to propagation of error.

Ex: 2.14

2.10 Control Charts

Fig 2.3

- Quality control chart:

  a time plot of a measured quantity

2.11 The Confidence Limit

$$\bar{x} \pm \frac{ts}{\sqrt{N}}$$

confidence level

t value: Table 2.1

Ex: 2.15

Confidence levels of 90-95% are generally accepted as reasonable.

The more measurements you make, the more confident you will be that the true value lies within a given range.

2.12 Tests of Significance

- The F test: F test=variance-ratio test
  1. for two method
  2. provides a simple method for comparing the precision of two sets of measurements.

The F test is used to determine if two variances are statistically different.

The sets do not necessarily have to be obtained from the same sample as long as the samples are sufficiently alike that the sources of indeterminate error can be assumed to be the same.

$$F = \frac{s_1^2}{s_2^2} \qquad s_1^2 > s_2^2$$

Table 2.2

Ex: 2.16

2.12-4

- The student t test: Table 2.1

The t test is used to determine if two sets of measurements are statistically different.

1. t test when an accepted value is known.

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{N}} \Rightarrow \pm t = (\bar{x} - \mu)\frac{\sqrt{N}}{s}$$

As the precision improves; it is easier to distinguish nonradom differences.

2. Paired t test:

For the two sets of data

$$\pm t = \frac{\overline{x_1} - \overline{x_2}}{s_p}\sqrt{\frac{N_1 N_2}{N_1 + N_2}}$$

s_p : pooled standard deviation

$$s_p = \sqrt{\frac{\sum (x_{i1} - \overline{x_1})^2 + \sum (x_{i2} - \overline{x_2})^2 + \cdots + \sum (x_{ik} - \overline{x_k})^2}{N - K}}$$

In applying the t test between two methods, it is assumed that both methods have essentially the same standard deviation.

Ex: 2.18

3. t test with multiple sample

$$t = \frac{\bar{D}}{s_d}\sqrt{N} \qquad s_d = \sqrt{\frac{\sum (D_i - \bar{D})^2}{N - 1}}$$

D_i : the individual difference between the two methods for each sample with regard to sign

$\bar{D}$: the mean of all the individual differences

Ex: 2.19

The smaller the calculated t value, the more confident you are that there is no significant difference between the two methods.

Type I error:
employ too low a confidence level

Type II error:
employ too high a confidence level

2.13 Rejection of A Result: The Q Test

- Q test: Table 2.3

$$Q = \frac{a}{w}$$

  Ex: 2.20
- Median:

2.14 Statistics for Small Data Sets

- Median versus the mean:

$$\text{median}\,(M) \leftrightarrow \text{mean}\,\bar{x}$$

The efficiency of M (E_M):
the ratio of the variances of sampling distributions of these two estimates of the true mean value.

Table 2.4

2.14-2

- Range versus the standard deviation:
  efficiency of the range: E_R Table 2.4
  $S_r = RK_p$
  S: standard deviation
  K_R: range deviation factor
  As N increases, the efficiency of the range decreases.
- Confidence limits using the range
  confidence limit $= x \pm Rt_r$

2.15 Linear Least Squares

$y = mx + b$

method of least squares:

the sum of the squares of the deviations of the points from the line is minimum.

$$S = \sum (y_i - y_\ell)^2 = \sum [y_i - (mx_i + b)]^2$$

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad b = \bar{y} - m\bar{x}$$

$$m = \frac{\sum x_i y_i - \left[\left(\sum x_i \sum y_i \right) \big/ n \right]}{\sum x_i^2 - \left[\left(\sum x_i \right)^2 \big/ n \right]}$$

Ex : 2.21

$$S_y = \sqrt{\frac{\sum \left[y_i - (mx_i + b)\right]^2}{N-2}}$$

$$= \sqrt{\frac{\left[\sum y_i^2 - \left(\sum y_i\right)^2 \big/ N\right] - m^2 \left[\sum x_i^2 - \left(\sum x_i\right)^2 \big/ N\right]}{N-2}}$$

$$S_m = \sqrt{\frac{S_y^2}{\sum\left(\bar{x} - x_i\right)^2}} = \sqrt{\frac{S_y^2}{\sum x_i^2 - \left(\sum x_i\right)^2 / N}}$$

$$S_b = S_y \sqrt{\frac{\sum x_i^2}{N\sum x_i^2 - \left(\sum x_i\right)^2}} = S_y \sqrt{\frac{1}{N - \left(\sum x_i\right)^2 / \sum x_i^2}}$$

2.16 The Correlation Coefficient

A measure of the correlation between two variables.

- Pearson correlation coefficient:

$$r = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{nS_xS_y} \qquad -1 \le r \le 1$$

$$= \frac{\sum x_i y_i - n\overline{xy}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\sum y_i^2 - n\bar{y}^2\right)}}$$

$$= \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n\sum x_i^2 - \left(\sum x_i\right)^2][n\sum y_i^2 - \left(\sum y_i\right)^2]}}$$

2.16-2

In general:

0.90<r<0.95 → a fair curve

0.95<r<0.99 → a fair curve

r<0.99 → excellent linearity

2.17 Detection Limits

The limit of detection is the lowest concentration level that can be determined to be statistically different from an analyte blank.

detection limit:

the concentration that gives a signal three times the standard deviation of the background signal.

2.17-2

For quantitative measurements, concentrations should be at least 10 times the detection limit.

2.18 Statistics of Sampling

- The precision of a result:

  The accuracy and precision of an analysis is limited by the sampling rather the measurement step.

$$S_o^2 = S_s^2 + S_a^2$$

- The true value:

$$\mu = \bar{x} \pm \frac{ts}{\sqrt{N}}$$

- Minimum sample size:

  Ingamell's sampling constant, K_s

  w: the weight of sample analyted

  R: the percent relative standard deviation of the sample composition

  K_s: the weight of sample for 1% sampling uncertainty at a 68% onfidence level

  Ex: 2.25

- Minimum number of sample:

$$n = \frac{t^2 S_s^2}{R^2 \bar{x}^2} \qquad R = \frac{S_x}{\bar{x}} \qquad n = \frac{t^2 S_s^2}{S_x^2}$$

S_s^2 : sampling variance

S_s : adsolute standard deviation

R : acceptable relative standard deviation of the average of the analytical results, $\bar{x}$

for Gaussian distribution

for poisson distribution:

$$S_s^2 \approx \bar{x} \text{ or } \mu$$

$$n = \frac{t^2}{R^2 \bar{x}} \cdot \frac{S_s^2}{\bar{x}} = \frac{t^2}{R^2 \bar{x}}$$