# Statistical Analysis of Data Collected

When performing laboratory experiments, a certain variation in results will always occur regardless of the care taken in measuring. Though all systematic errors have been avoided, the student will notice that the values obtained will deviate from each other. Random errors are unavoidable. Variations occur because each measurement is subject to the sensitivity of the measuring instrument and/or the estimation of the experimenter.

Examples: Sample masses determined using an analytical balance to determine the sample mass has a precision of ±0.0001 g. Each of the two buret readings used to determine the volume of titrant has a precision of ±0.02 mL. Finally, the experimenter's estimation of the endpoint is ±0.03 mL minimum. These sources of error can either add or cancel each other, producing the normal distribution (bell) curve shown. The individual trials vary about the average (mean). The challenge is to determine which trials to use and which should be discarded.

## Test for Discarding Outlying Results

### 4D Test

**1** Average all trials *EXCEPT* the outlying value.

**2** Determine the average deviation of the trials retained.

**3** If the deviation of the outlying value is greater than four times the average deviation, the sample potentially can be discarded.

Perform several additional trials. With a larger sample size, one can feel more confident in discarding data if all results except the outlier are in close agreement.

## Standard Deviation

The standard deviation for a small set of data will generally be larger than the average deviation since only a small part of the normal distribution curve has been sampled. The two deviations will approach each other as the number of trials increases. The STANDARD DEVIATION is generally better to use since this small sample size is taken into account by dividing by $n - 1$ rather than $n$ trials.
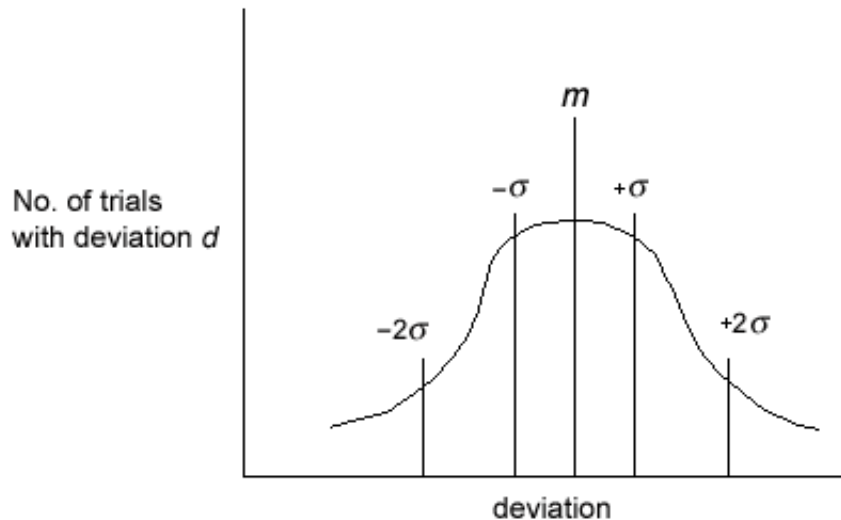
### Normal Distribution Curve

One standard deviation ($\pm 1\sigma$) contains 68.3% of data.

Two standard deviations ($\pm 2\sigma$) contain 95% of data.

Three standard deviations ($\pm 3\sigma$) contain 99.7% of data.

Four standard deviations ($\pm 4\sigma$) contain 99.9+% of data.

Figure

# SAMPLE CALCULATION

## Sample Data

| SAMPLE DATA | | | | |
|---|---|---|---|---|
| | **Trial # 1** | **Trial # 2** | **Trial # 3** | **Trial # 4** |
| **Molarity (mol/L)** | 0.3221 | 0.3231 | 0.3218 | 0.3561* |

*Outlier

## Statistical Analysis of Data

**4D Test**

**1** Compute average "good" values (all except outlier).

$$\frac{0.3221 + 0.3231 + 0.3218}{3} = 0.3223 \ M$$

**2** Compute the average deviation of "good" values. (Note: deviation is the absolute value of the difference from the mean.)

$$
\begin{aligned}
|0.3221 - 0.3223| &= 0.0002 \\
|0.3231 - 0.3223| &= 0.0008 \\
|0.3218 - 0.3223| &= \underline{0.0005} \\
& \quad 0.0015/3 = 0.0005
\end{aligned}
$$

**3** Compute the deviation of outlying value, and compare it to 4D or $4 \times$ average deviation.

$$|0.3561 - 0.3223| = 0.0338 \quad \text{compared to:} \quad (0.0005) = 0.0020$$
$$0.00338 > 0.0020$$

$\therefore$ Based upon the 4D test, results **MAY** be discarded.

Once results that may be discarded have been determined, the standard deviation can be calculated.

## Standard Deviation

$$s = \sqrt{\frac{\sum(x_1 - m)^2}{n - 1}} \qquad \begin{aligned} x_i &= \text{sample value} \\ m &= \text{mean} \\ n &= \text{number of trials} \end{aligned} \tag{1}$$

$$m = \frac{0.3221 + 0.3231 + 0.3218}{3} = 0.3223 \ M$$

$$s = \sqrt{\frac{((0.3221 - 0.3223)^2 + (0.3231 - 0.3223)^2 + (0.3218 - 0.3223)^2)}{(3 - 1)}}$$

$$s = \sqrt{\frac{(4E^{-8} + 6.4E^{-7} + 2.5E^{-7})}{2}}$$

$$= 6.8E^{-4} \ (\text{calculator})$$

Final answer: $0.3223 \pm 0.0007 \ M$. (In your final answer, express your standard deviation to one significant figure and your molarity to four significant figures.)

You can also calculate standard deviation using your calculator:

- You will need to put lists of information (DATA) into your calculator. Do this by pressing STAT 1:Edit.

- Enter your data in one of the columns, for example, L1.

- Now go to STAT > CALC and select 1:1-Var Stats.

- Enter L1 ($2^{\text{nd}}$ 1).

- $Sx$ is the standard deviation, which for this example is $6.806859E^{-4}$(express as 0.0007).

## YOU WILL ALSO CALCULATE % ERROR AND % ACCURACY IN MANY OF THE LABS

The % error is how "far off" your results are.

$$\% = \frac{\text{theoretical} - \text{experimental}}{\text{theoretical}} \text{ x } 100 \tag{2}$$

The % accuracy is how "close" your results are.

$$\% = \frac{\text{experimental}}{\text{theoretical}} \text{ x } 100 \tag{3}$$