



Gaussian Distribution

- If the errors for a set of measurements are random, then the measurements can be represented by a Gaussian distribution
- Two quantities are used to describe a Gaussian curve
 - The *Mean*—the arithmetic average of all measurements
 - The *Standard Deviation*—a measure of the spread in the measurements



Gaussian Distribution

- Mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

n = total number of measurements

- Standard deviation

$$s = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$$

the factor n-1 in the denominator is called the degrees of freedom



Gaussian Distribution

Measurement	Value	Measurement	Value
1	0.3410	9	0.3430
2	0.3350	10	0.3420
3	0.3470	11	0.3560
4	0.3590	12	0.3500
5	0.3530	13	0.3630
6	0.3460	14	0.3530
7	0.3470	15	0.3480
8	0.3460		

$$\bar{x} = .3486$$

$$s = .0073_1$$



Gaussian Distribution

- For small, finite number of measurements, the mean and standard deviation are only approximations of the “true” mean (μ) and standard deviation of the sample (σ)
- For a large number of measurements:

$$\lim_{n \rightarrow \infty} \bar{x} = \mu$$

$$\lim_{n \rightarrow \infty} s = \sigma$$



Gaussian Distribution

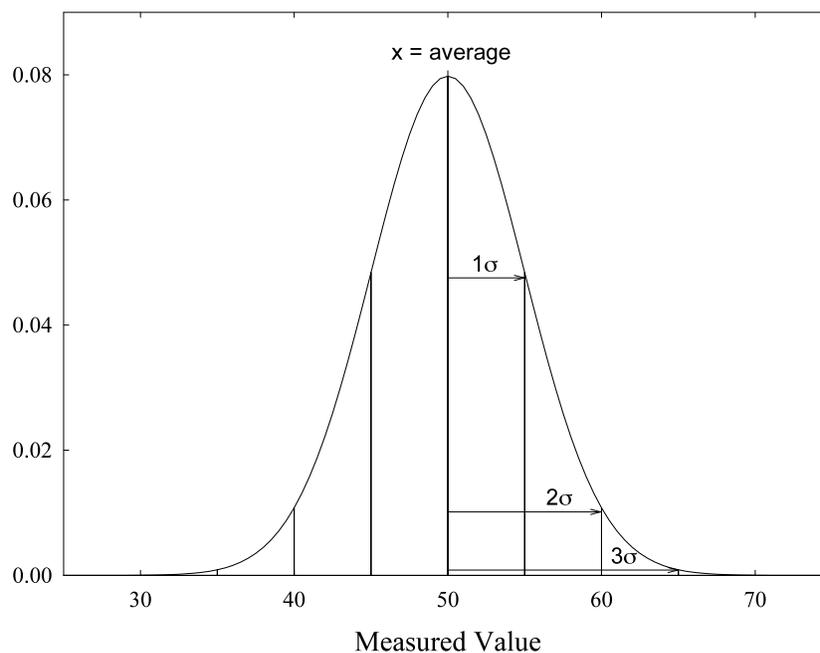
- The mathematical expression for a Gaussian distribution is:

$$y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

The factor $1/\sigma(2\pi)^{1/2}$ is a normalization constant and assures that the area under the curve for the Gaussian function equal unity

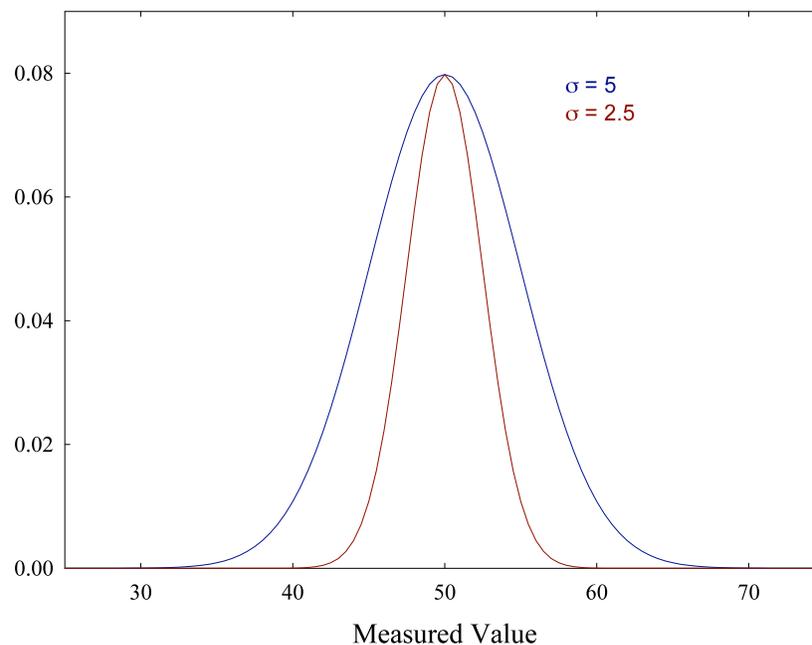


Gaussian Distribution





Gaussian Distribution



Gaussian Distribution

- 68.3% of measurements fall within $\pm 1\sigma$ of the mean for a Gaussian distribution
 - If we report the error bars of a measurement as $\pm 1\sigma$, we are confident that 68% of our results will lie within 1σ of the true value
- 95.5% of measurements fall within $\pm 2\sigma$ of the mean
- 99.7% of measurements fall within $\pm 3\sigma$ of the mean



Confidence Intervals

- Reported confidence intervals can be calculated using the *student's t* values (found in Table 4-1)

$$\mu = x_{\text{avg}} \pm (t \cdot s) / n^{1/2}$$

the value of t depends on the number of degrees of freedom and the confidence level you want to report



Confidence Intervals

Values of student t at 95% confidence level

<u>d.f.s</u>	<u>95% confidence</u>
1	12.706
2	4.303
3	3.182
4	2.776
5	2.571
10	2.228
40	2.021



Confidence Intervals

Example: The following data were determined for analysis of Ca^{2+} in water:

6.34 ppm 5.87 ppm 6.12 ppm 5.71 ppm 6.48 ppm

Report average and uncertainty at 95% confidence level

$$\bar{x} = 6.10_4 \quad s = 0.31_9$$

$$t = 2.776$$

$$t \cdot s/n^{1/2} = (2.776)(0.31_9)/5^{1/2} = 0.39_6$$

$$[\text{Ca}^{2+}] = 6.10_4 \pm 0.39_6 \text{ ppm}$$



Student t's

- Student t's can also be used to compare different measurements to see if they give the same results within experimental errors
- There are three different cases to consider:
 - Comparing measured result with “true” value
 - Comparing results from replicate measurements
 - Comparing individual differences from alternate experimental techniques



Student t's

- Comparing measured result with “true” value
- Use confidence interval expression at desired level of confidence to see if “true” value lies within prescribed range for results

$$\mu = x_{\text{avg}} \pm (t \cdot s) / n^{1/2}$$



Student t's

- Comparing results from replicate measurements

$$t_{\text{calc}} = \frac{|\bar{x}_1 - \bar{x}_2|}{s_{\text{pooled}}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$s_{\text{pooled}} = \sqrt{\frac{\sum_{\text{set 1}} (x_i - \bar{x}_1)^2 + \sum_{\text{set 2}} (x_j - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

If $t_{\text{calc}} > t_{\text{table}}$, the two results are statistically different



Student t's

- Comparing individual differences from alternate experimental techniques
define d_i as the difference between results for same sample by different experimental methods

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

If $t_{\text{calc}} > t_{\text{table}}$, the results are statistically different

$$t_{\text{calc}} = \frac{|\bar{d}|}{s_d} \sqrt{n}$$

Exclusion of Data Points—the Q Test



- Suppose you have a series of measurements for a given sample, and one data point seems to be far outside the average range of the other data
- We use the Q test to see if that point can be excluded from the calculation of the mean and standard deviation
 - Calculated range between extremes of data—highest data value minus lowest data value
 - Calculate gap between suspect data and nearest neighbor

Exclusion of Data Points— the Q Test



- Q is defined as ratio of data gap to data range

$$Q_{\text{calc}} = \frac{\text{gap}}{\text{range}}$$

- If $Q_{\text{calc}} > Q_{\text{table}}$, the point may be excluded

<u>Q (90%)</u>	<u>n</u>
0.765	4
0.642	5
0.560	6
0.507	7
0.468	8
0.437	9

Exclusion of Data Points— the Q Test



Example: The following data were measured for analysis of X in a sample:

5.678 5.589 5.431 4.998 5.486

Decide if the point, 4.995, can be discarded:

$$\text{range} = 5.678 - 4.998 = 0.680$$

$$\text{gap} = 5.431 - 4.998 = 0.433$$

$$Q_{\text{calc}} = 0.433/0.680 = 0.637$$

$$Q_{\text{table}} = 0.642$$

$$Q_{\text{calc}} < Q_{\text{table}} \Rightarrow \text{value may not be discarded}$$

Exclusion of Data Points— the Q Test



Example: The following data were measured for analysis of X in a sample:

Result including low value:

$$x = 5.436 \pm 0.326 \text{ (95\% confidence level)}$$

Result excluding low value:

$$x = 5.546 \pm 0.175 \text{ (95\% confidence level)}$$

Excluding low point gives much tighter uncertainty, but it is not justified in this case.