Statistics

Chapter 4

Section 4-4
Comparison of Means with Student's *t*

t Test: Comparison of Means

Are the *means* of two sets of measurements "statistically different"?

If you make two sets of measurements of the same quantity, $\bar{x}_1 \neq \bar{x}_2$ generally due to random variations in measurements.

The ***t* test** determines if there is a statistical difference between $\bar{x}_1$ and $\bar{x}_2$.

If $t_{calculated} > t_{table}$, then reject the null hypothesis.

- There is <5% chance that the two data sets came from populations with the same population mean.
- The difference is considered significant.

Comparisons of Means with Student's *t*

Case 1

Compare $\bar{x}$ to a known value:

- Measure quantity several times.
- Obtain $\bar{x}$ and s.

Does $\bar{x}$ compare to accepted answer, μ?

Case 2

Compare $\bar{x}_1$ to $\bar{x}_2$ with replicate samples:

- Measure quantity multiple times by two different methods.
- Obtain $\bar{x}_1 \pm s_1$ and $\bar{x}_2 \pm s_2$ (for each method).

Does $\bar{x}_1$ agree with $\bar{x}_2$ within experimental uncertainty?

Case 3 (paired test)

Compare two methods where samples are not duplicated:

- Measure sample A once by method 1 and once by method 2.
- Measure sample B once by method 1 and once by method 2.

Do the two methods agree within experimental uncertainty?

Case 1: Comparing Measured Result with "Known" Value

A coal sample is certified to contain 3.19 wt% sulfur. A new analytical method measures values of 3.29, 3.22, 3.30, and 3.23 wt% sulfur, giving a mean $\bar{x} = 3.26_0$ and a standard deviation $s = 0.041$.

Does the answer using the new method agree with the known answer?

$$\textbf{Confidence Interval} = \bar{x} \pm \frac{ts}{\sqrt{n}} = 3.26_0 \pm \frac{(3.182)(0.004\,1)}{\sqrt{4}} = \mathbf{3.260 \pm 0.006\,5}$$

- Confidence interval = 3.19_5 to 3.32_5 wt%
- Known value, **3.19 wt%**, outside 95% confidence interval
- Method gives "**different**" result from known result

  (Result is so close scientist might want to complete a few more trials to confirm)

Case 2a: Comparing Replicate Measurements When Standard Deviations Are *Not* Significantly Different

Recall HCO_3^- in horse blood is measured after each race.:

	Original instrument	Substitute instrument
Mean ($\bar{x}$, mM)	36.14	36.20
Standard deviation (s, mM)	0.28	0.47
Number of measurements (n)	10	4

$$F_{\text{calculated}} = \frac{s_1^2}{s_2^2} = \frac{(0.47)^2}{(0.28)^2} = 2.8_2$$

$F_{calculated} = 2.8_2 < F_{table} = 5.08$, so s_1 and s_2 are ***not*** *significantly different.*

Do the means of the two methods differ?

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

$$s_{\text{pooled}} = \sqrt{\frac{\sum_1 (x_i - \bar{x}_1)^2 + \sum_2 (x_j - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{s_1^2 (n_1 - 1) + s_2^2 (n_2 - 1)}{n_1 + n_2 - 2}}$$

Case 2a: Comparing Replicate Measurements When Standard Deviations Are *Not* Significantly Different

Do the means of the two methods differ?

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{n_1+n_2-2}} = \sqrt{\frac{0.28^1(10-1)+0.47^2(4-1)}{10+4-2}} = 0.33_8$$

$$t = \frac{|\bar{x}_1-\bar{x}_2|}{s_{pooled}}\sqrt{\frac{n_1 n_2}{n_1+n_2}} = \frac{|36.14-35.20|}{0.33_8}\sqrt{\frac{10\times 4}{10+4}} = 0.30_0$$

	Original instrument	Substitute instrument
($\bar{x}$, mM)	36.14	36.20
(s, mM)	0.28	0.47
(n)	10	4

95% critical value of t in Table 4-4 for $(n_1 + n_2 - 2) = 12$ degrees of freedom lies between 2.228 and 2.131

If $t_{calculated} > t_{table}$, then reject the null hypothesis. $t_{calculated}(0.300) < t_{table}(2.131)$

- There is <5% chance that the two data sets came from populations with the same population mean.
- The difference in means is *not* considered significant.

Lord Rayleigh and the Discovery of Argon

Dry air is composed of ~1/5 oxygen and ~4/5 nitrogen

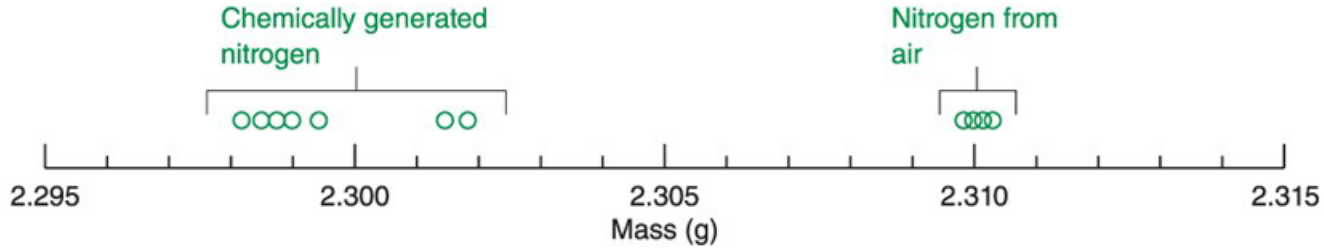
Measured nitrogen with two experiments:

(at constant temperature, pressure, and volume)

- Mass of N_2 after removing O_2 *from air*
- Mass of N_2 generated *from chemical decomposition*

Do the means of the two methods differ?

Figure 4-7



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Table 4-5

From air (g)	From Chemical decomposition (g)
2.310 17	2.301 43
2.309 86	2.298 90
2.310 10	2.298 16
2.310 01	2.301 82
2.310 24	2.298 69
2.310 10	2.299 40
2.310 28	2.298 49
—	2.298 89
Average $2.310\ 10_9$	$2.299\ 47_2$
Standard deviation $0.000\ 14_3$	$0.001\ 37_9$

Case 2b: Comparing Replicate Measurements When Standard Deviations *Are* Significantly Different

Do the means of the two methods differ?

If the standard deviations of the two methods differ (*F*-test), the *t* test equations become:

$$t_{calculated} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\left(u_1^2\right)+\left(u_2^2\right)}}$$

Standard deviation of the mean

$$u_i = s_i/\sqrt{n_i}$$

$$\text{Degrees of freedom} = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}} = \frac{\left(u_1^2+u_2^2\right)^2}{\frac{u_1^4}{n_1-1} + \frac{u_2^4}{n_2-1}}$$

Round degrees of freedom to the nearest integer.

Compare $t_{calculated}$ to t_{table} at 95% confidence using appropriate degrees of freedom.

Example: Is Rayleigh's N_2 from Air Denser Than N_2 from Chemicals? (1 of 4)

The average mass of nitrogen from air is $\bar{x}_1 = 2.310\,10_9$ g, with a standard deviation of $s_1 = 0.000\,14_3$ (for $n_1 = 7$ measurements).

The average mass from chemical decomposition is $\bar{x}_2 = 2.299\,47_2$ g, with a standard deviation of $s_2 = 0.001\,37_9$ (for $n_2 = 8$ measurements).

Are the two masses significantly different?

Table 4-5

From air (g)	From Chemical decomposition (g)
2.310 17	2.301 43
2.309 86	2.298 90
2.310 10	2.298 16
2.310 01	2.301 82
2.310 24	2.298 69
2.310 10	2.299 40
2.310 28	2.298 49
—	2.298 89
Average $2.310\,10_9$	$2.299\,47_2$
Standard deviation $0.000\,14_3$	$0.001\,37_9$

Example: Is Rayleigh's N_2 from Air Denser Than N_2 from Chemicals? (2 of 4)

Solution: The *F*-test told us that the standard deviations are significantly different, so we use Equations 4-9b and 4-10b:

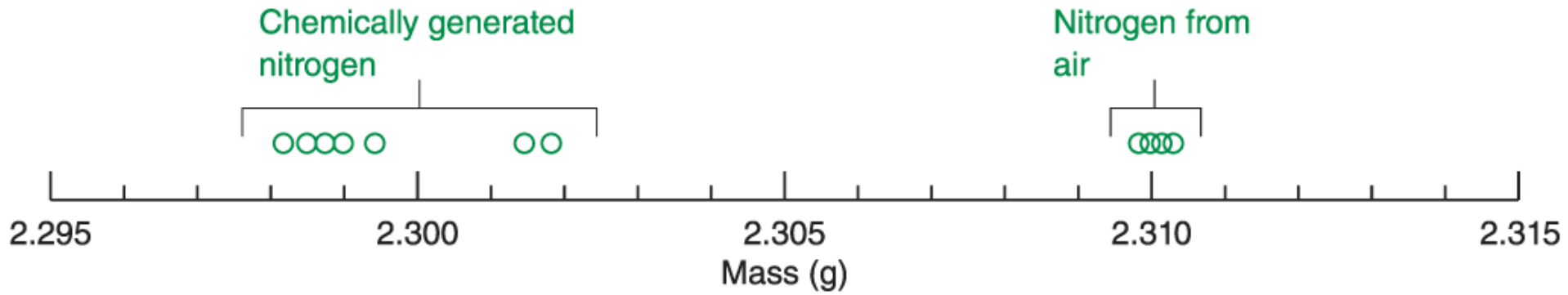
$$t_{\text{calculated}} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\sqrt{\left(s_1^2/n_1\right)+\left(s_2^2/n_2\right)}} = \frac{\left|2.310\ 10_9 - 2.299\ 47_2\right|}{\sqrt{0.000\ 14_3{}^2/7 + 0.001\ 37_9{}^2/8}} = 21.7$$

$$\text{Degrees of freedom} = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1}+\frac{\left(s_2^2/n_2\right)^2}{n_2-1}} = \frac{\left(0.000\ 14_3{}^2/7 + 0.001\ 37_9{}^2/8\right)^2}{\frac{\left(0.000\ 14_3{}^2/7\right)^2}{7-1}+\frac{\left(0.001\ 37_9{}^2/8\right)^2}{8-1}} = 7.17$$

Example: Is Rayleigh's N_2 from Air Denser Than N_2 from Chemicals? (3 of 4)

Solution: Equation 4-10b gives us 7.17 degrees of freedom, which we round to 7. For 7 degrees of freedom, the critical value of t in Table 4-4 for 95% confidence is 2.365. The observed value $t_{calculated}$ = 21.7 far exceeds t_{table}. The obvious difference between the two data sets in Figure 4-7 is highly significant.

Figure 4-7



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Example: Is Rayleigh's N_2 from Air Denser Than N_2 from Chemicals? (4 of 4)

Test Yourself: If the difference between the two mean values were half as great as Rayleigh found, but the standard deviations were unchanged, would the difference still be significant?

Case 3: Paired *t* Test for Comparing Individual Differences

Do the methods give the same answer?

- Use two methods to make single measurements on several different samples..
- No measurement is duplicated.

$$s_d = \sqrt{\frac{\sum(d_1 - \bar{d})^2}{n-1}}$$

$$= \sqrt{\frac{(0.01 - \bar{d})^2 + (0.37 - \bar{d})^2 + (-0.14 - \bar{d})^2 + \cdots}{8-1}} = 0.40_1$$

$$t_{calculated} = \frac{|\bar{d}|}{s_d}\sqrt{n} = \frac{|0.11_4|}{0.40_1}\sqrt{8} = 0.80_3$$

Figure 4-8

	A	B	C	D
1	Comparison of methods for measuring nitrate			
2				
3		Nitrate (ppm) in plant extract		
4		Spectrophotometry	Experimental	
5	Sample	with Cd reduction	biosensor	Difference (d_i)
6	1	1.22	1.23	0.01
7	2	1.21	1.58	0.37
8	3	4.18	4.04	–0.14
9	4	3.96	4.92	0.96
10	5	1.18	0.96	–0.22
11	6	3.65	3.37	–0.28
12	7	4.36	4.48	0.12
13	8	1.61	1.70	0.09
14			Mean difference =	0.114
15		Standard deviation of differences =		0.401
16			Number of samples =	8
17	D6 = C6-B6		$t_{calculated}$ =	0.803
18	D14 = AVERAGE(D6:D13)		t_{table} =	2.365
19	D15 = STDEV.S(D6:D13)			
20	D16 = COUNT(D6:D13)			
21	D17 = ABS(D14)*SQRT(A13)/D15 (ABS = absolute value)			
22	D18 = T.INV.2T(0.05,A13-1)			

One-Tailed and Two-Tailed Significance Tests

Two tailed tests: *t* test calculations assume:

- *Certified value lies in the **outer 5%** of the area under the curve*

One tailed tests: compare mean with regulatory limit

- *5% region lies only on one side of the certified mean*

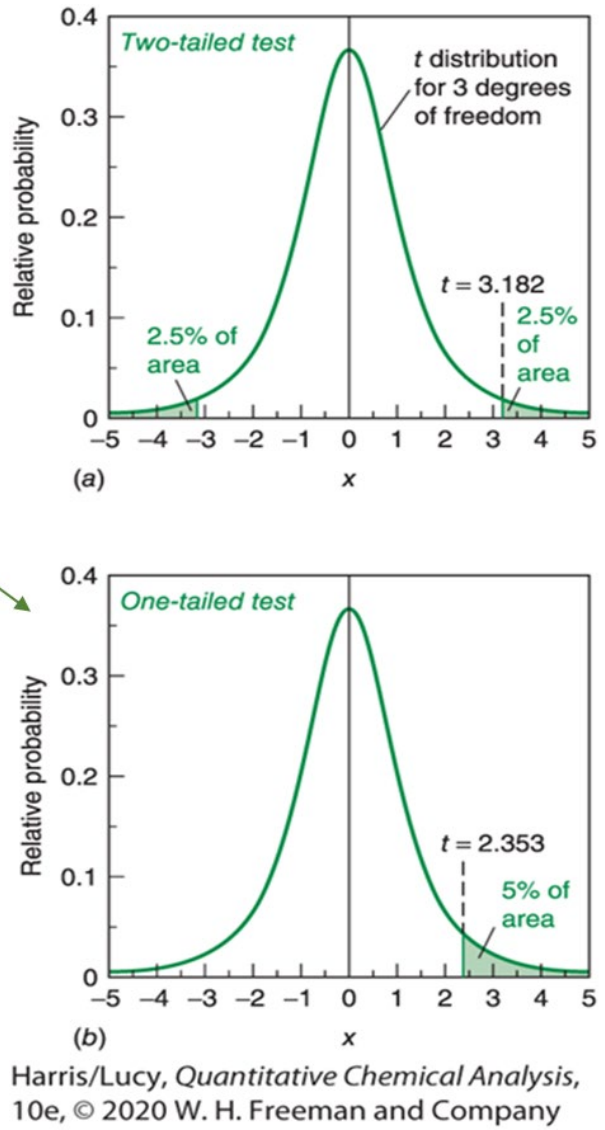
Consider drinking water: We are concerned only if the probability of arsenic (As) in water *exceeds **the limit***.

EPA maximum permissible level = 10 µg As/L

Water samples → 10.06, 12, 19, and 10.04 µg As/L; $\bar{x} = 10.10_{25} \pm 0.06_{75}$ µg/L

$$t_{calculated} = \frac{|\bar{x} - \text{regulatory limit}|}{s}\sqrt{n} = \frac{|10.10_{25} - 10|}{0.06_{75}}\sqrt{4} = 3.0_4$$

Figure 4-9



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Section 4-5
t Tests with a Spreadsheet

t Tests with a Spreadsheet

Spreadsheet for comparing mean values of Rayleigh's nitrogen measurements

Figure 4-

	A	B	C	D	E	F	G
1	Analysis of Rayleigh's Data				t-Test: Two-Sample Assuming Equal Variances		
2						*Variable 1*	*Variable 2*
3		Mass of gas (g) collected from			Mean	2.310109	2.299473
4		air	chemical		Variance	2.03E-08	1.9E-06
5		2.31017	2.30143		Observations	7	8
6		2.30986	2.29890		Pooled Variance	1.03E-06	
7		2.31010	2.29816		Hypothesized Mean Diff	0	
8		2.31001	2.30182		df	13	
9		2.31024	2.29869		t Stat	20.21372	
10		2.31010	2.29940		P(T<=t) one-tail	1.66E-11	
11		2.31028	2.29849		t Critical one-tail	1.770932	
12			2.29889		P(T<=t) two-tail	3.32E-11	
13	Average	2.31011	2.29947		t Critical two-tail	2.160368	
14	Std Dev	0.00014	0.00138				
15					t-Test: Two-Sample Assuming Unequal Variances		
16	B13 = AVERAGE(B5:B12)					*Variable 1*	*Variable 2*
17	B14 = STDEV.S(B5:B12)				Mean	2.310109	2.299473
18					Variance	2.03E-08	1.9E-06
19					Observations	7	8
20					Hypothesized Mean Diff	0	
21					df	7	
22					t Stat	21.68022	
23					P(T<=t) one-tail	5.6E-08	
24					t Critical one-tail	1.894578	
25					P(T<=t) two-tail	1.12E-07	
26					t Critical two-tail	2.364623	

Section 4-6
Grubbs Test for an Outlier

Grubbs Test: Check for Outliers

Should a data point that looks like an anomaly be discarded?

If you make several replicate measurements, results should fall within a Gaussian distribution about the mean. But when ***n*** is small, it can be difficult to determine if an outlying data point falls within the normal distribution.

The **Grubbs test** is a statistical test to decide whether to discard a datum that appears discrepant (an ***"outlier"***).

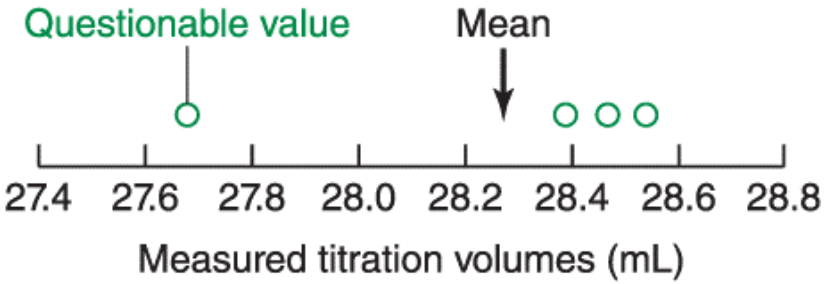
If $G_{calculated} > G_{table}$, then reject the null hypothesis.

- There is <5% chance that the suspicious data point is a member of the same population as the other measurements.
- The difference is considered significant.

Grubbs Test for an Outlier (1 of 3)

When a single data point lies far from the other data in a set of measurements:

- First, check your notebook.
- Are there any recorded observations about the anomalous data point (for example, a note that solution was lost during transfer)?
- Any data point based on *recorded* faulty procedure should be discarded, no matter how well it fits the rest of the data (*"blunder"*).

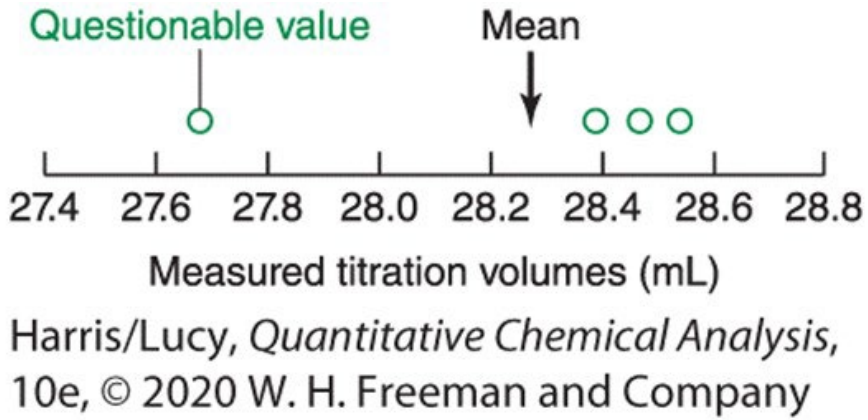


Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Grubbs Test for an Outlier (2 of 3)

In the absence of a recorded blunder, use the **Grubbs test**.

$$G_{\text{calculated}} = \frac{|\,\text{questionable value} - \bar{x}\,|}{s}$$



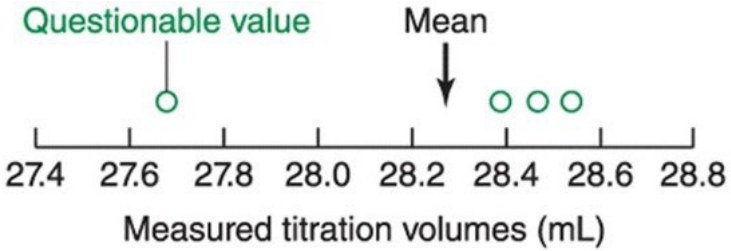
Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

- If $G_{\text{calculated}}$ is greater than G in Table 4-6, the questionable point should be discarded.
- Only one outlier may be rejected using the Grubbs test.

Grubbs Test for an Outlier (3 of 3)

In the absence of a recorded blunder, use the **Grubbs test**.

$$G_{\text{calculated}} = \frac{|\,\text{questionable value} - \bar{x}\,|}{s}$$



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Volumes for replicate titrations (mL): **28.54, 28.39, 28.47, 27.68**

$\bar{x} = 28.27 \pm 0.40$ mL; RSD = 1.4% ← (Larger than expected precision for this titration)

$$G_{\text{calculated}} = \frac{|27.68 - 28.27|}{0.40} = 1.482 \qquad G_{\text{table}} = 1.463 \text{ for 4 observations}$$

$G_{\text{calculated}} > G_{\text{table}}$, so reject the null hypothesis.

- The questionable data point is an "outlier" and should be discarded.

Table 4-6 Critical values of *G* for rejection of outlier

Number of observations	*G* (95% confidence)	Number of observations	*G* (95% confidence)
3	1.153	10	2.176
4	1.463	11	2.234
5	1.672	12	2.285
6	1.822	15	2.409
7	1.938	20	2.557
8	2.032	30	2.745
9	2.110	50	2.956

$G_{calculated} = |questionable\ value - mean|/s$. If $G_{calculated} > G_{table}$, the value in question can be rejected with 95% confidence. Values in this table are for a one-tailed test, as recommended by ASTM.

Section 4-7
The Method of Least Squares

The Method of Least Squares

For most chemical analyses, the response obtained by the given lab procedure must be compared to known quantities (called ***standards***).

In this way the response from an *unknown* quantity can be interpreted.

- Prepare a **calibration curve** from known standards.
- Work in a region where the calibration curve is linear (usually).

Method of least squares: used to draw the "best" straight line through experimental data points that contain some scatter

- Some points will lie above and some below the line.
- Equation $y = mx + b$ can be used to quantify the unknown from

Finding the Equation of the Line

Assume:

- Uncertainty in y values is much greater than uncertainty in x values ($s_y >> s_x$).
- Uncertainties of all y values are similar.

Draw a line to minimize vertical deviations between points and line.

Slope

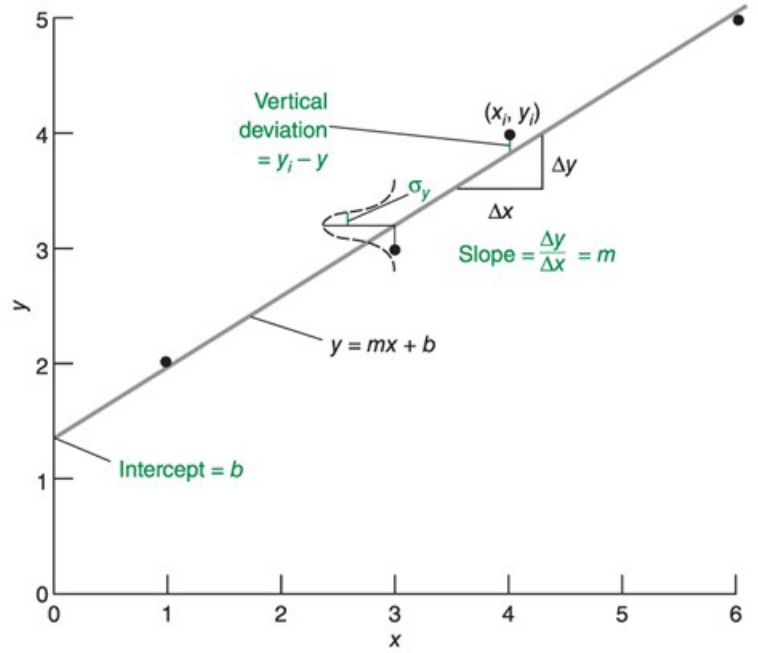
$$y = mx + b$$

y-intercept

- Vertical deviation $d_i = y_i - y = y_i - (mx_i + b)$

Deviations can be positive or negative. To minimize magnitude, irrespective of sign, square the deviation: $d_i^2 = (y_i - mx_i - b)^2$ ← **method of least squares**

Figure 4-11

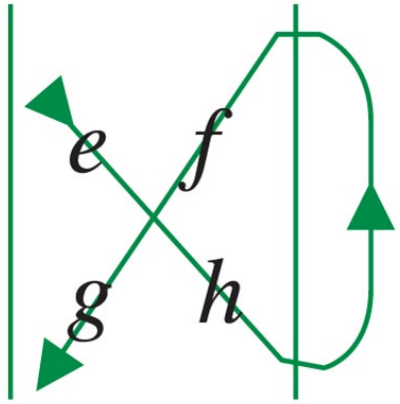


Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Determinants

Mathematically finding values of *m* and *b* that minimize the sum of the squares involves some calculus.

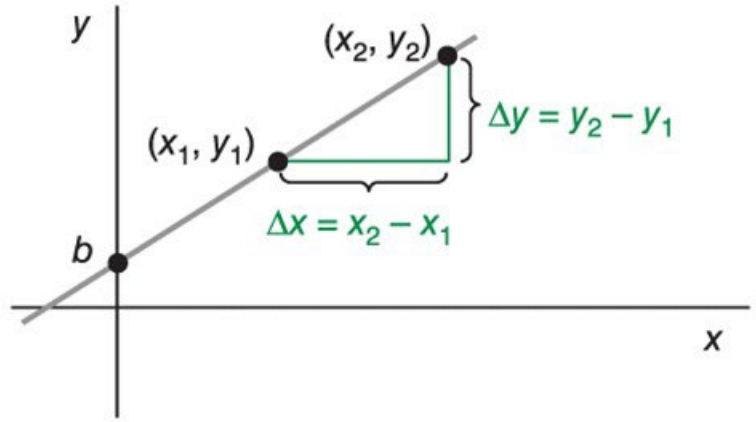
We express the final solution for *m* and *b* as **determinants**.



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

$$\begin{vmatrix} e & f \\ g & h \end{vmatrix} = eh - fg$$

$$\begin{vmatrix} 6 & 5 \\ 4 & 3 \end{vmatrix} = (6 \times 3) - (5 \times 4) = 2$$



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Determinants to Solve Method of Least Squares

$$\text{Least-squares "best" line} \begin{cases} \text{Slope}: & m = \begin{vmatrix} \sum(x_i y_i) & \sum x_i \\ \sum y_i & n \end{vmatrix} \div D \\ \text{Intercept}: & b = \begin{vmatrix} \sum(x_i^2) & \sum(x_i y_i) \\ \sum x_i & \sum y_i \end{vmatrix} \div D \end{cases}$$

where D is

$$D = \begin{vmatrix} \sum(x_i^2) & \sum x_i \\ \sum x_i & n \end{vmatrix}$$

and n is the number of points.

Table 4-7 Calculations for least-squares analysis

x_i	y_i	$x_i y_i$	x_i^2	$d_i (= y_i - mx_i - b)$	d_i^2
1	2	2	1	0.038 46	0.001 479 3
3	3	9	9	−0.192 31	0.036 982
4	4	16	16	0.192 31	0.036 982
6	5	30	36	−0.038 46	0.001 479 3
$\overline{\sum x_i = 14}$	$\overline{\sum y_i = 14}$	$\overline{\sum (x_i y_i) = 57}$	$\overline{\sum (x_i^2) = 62}$		$\overline{\sum (d_i^2) = 0.076\,923}$

$$m = \begin{vmatrix} 57 & 14 \\ 14 & 4 \end{vmatrix} \div \begin{vmatrix} 62 & 14 \\ 14 & 4 \end{vmatrix} = \frac{(57 \times 4) - (14 \times 14)}{(62 \times 4) - (14 \times 14)} = \frac{32}{52} = 0.615\ 38$$

$$b = \begin{vmatrix} 62 & 57 \\ 14 & 14 \end{vmatrix} \div \begin{vmatrix} 62 & 14 \\ 14 & 4 \end{vmatrix} = \frac{(62 \times 4) - (57 \times 14)}{(62 \times 4) - (14 \times 14)} = \frac{70}{52} = 1.346\ 15$$

$$y = 0.\ 615\ 38\ x + 1.\ 346\ 15$$

Example: Finding Slope and Intercept with a Spreadsheet (1 of 3)

Excel has functions called SLOPE and INTERCEPT, whose use is illustrated here:

	A	B	C	D	E	F
1	x	y			Formulas:	
2	1	2		slope =		
3	3	3		0.61538	D3 = SLOPE(B2:B5,A2:A5)	
4	4	4		intercept =		
5	6	5		1.34615	D5 = INTERCEPT(B2:B5,A2:A5)	

Example: Finding Slope and Intercept with a Spreadsheet (2 of 3)

The slope in cell D3 is computed with the formula "=SLOPE(B2:B5,A2:A5)", where B2:B5 is the range containing the *y* values and A2:A5 is the range containing *x* values.

	A	B	C	D	E	F
1	x	y			Formulas:	
2	1	2		slope =		
3	3	3		0.61538	D3 = SLOPE(B2:B5,A2:A5)	
4	4	4		intercept =		
5	6	5		1.34615	D5 = INTERCEPT(B2:B5,A2:A5)	

Example: Finding Slope and Intercept with a Spreadsheet (3 of 3)

Test Yourself: Change cell A3 from 3 to 3.5 and find the new slope and intercept.

How Reliable Are Least Squares Parameters?

To estimate uncertainties in *m* and *b*, an uncertainty analysis must be performed.

Estimate σ_y, the population standard deviation for all *y*, by calculating s_y.

$$s_y = \sqrt{\frac{\sum(d_i^2)}{n-2}} \qquad u_m^2 = \frac{s_y^2 n}{D} \qquad u_b^2 = \frac{s_y^2 \sum(x_i^2)}{D}$$

The first digit of the uncertainty is the last significant figure. We often retain extra, insignificant, digits to prevent round-off errors in further calculations.

Slope: $\begin{matrix} 0.615\,38 \\ \pm 0.054\,39 \end{matrix} = 0.62 \pm 0.05 \text{ or } 0.61_5 \pm 0.05_4$

Intercept: $\begin{matrix} 1.346\,15 \\ \pm 0.214\,15 \end{matrix} = 1.3 \pm 0.2 \text{ or } 1.3_5 \pm 0.2_1$

Example: Finding s_y, u_m, and u_b with a Spreadsheet (1 of 3)

Excel function LINEST returns the slope and intercept and their standard uncertainties in a table (a *matrix*). As an example, enter *x* and *y* values from Table 4-7 in columns A and B. Then highlight the 3-row × 2-column region E3:F5 with your mouse. This block of cells is selected for the output of LINEST. On the Formulas ribbon, go to Insert Function. In the window that appears, in "Or select a category" select Statistical and double click LINEST. The new window asks for four inputs to the function. For *y* values, enter B2:B5. Then enter A2:A5 for *x* values. The next two entries are both "TRUE". The first TRUE tells Excel that we want to compute the *y*-intercept of the line and not force the intercept to be 0. The second TRUE tells Excel to print out the uncertainties as well as the slope and intercept. The formula you have just entered is "=LINEST(B2:B5,A2:A5,TRUE,TRUE)". Now press CONTROL+SHIFT+ENTER on a PC or CONTROL+SHIFT+RETURN on a Mac.

Example: Finding s_y, u_m, and u_b with a Spreadsheet (2 of 3)

Excel prints out a matrix in cells E3:F5. Write labels around the block to indicate what is in each cell. The slope and intercept are on the top line. The second line contains u_m and u_b. Cell F5 contains s_y, and cell E5 contains a quantity called R^2, which is defined in Equation 5-3 and is a measure of the goodness of fit of the data to the line. The closer R^2 is to unity, the better the fit.

	A	B	C	D	E	F	G
1	x	y			Output from LINEST		
2	1	2			Slope	Intercept	
3	3	3		Parameter	0.61538	1.34615	
4	4	4		u_m	0.05439	0.21414	u_b
5	6	5		R^2	0.98462	0.19612	s_y
6	Highlight cells E3:F5						
7	Type "= LINEST(B2:B5,A2:A5,TRUE,TRUE)"						
8	Press CTRL+SHIFT+ENTER (on PC)						
9	Press CTRL+SHIFT+RETURN (on Mac)						

Example: Finding s_y, u_m, and u_b with a Spreadsheet (3 of 3)

Test Yourself: Change cell A3 from 3 to 3.5 and apply LINEST. What is the value of s_y from LINEST?

Section 4-8
Calibration Curves

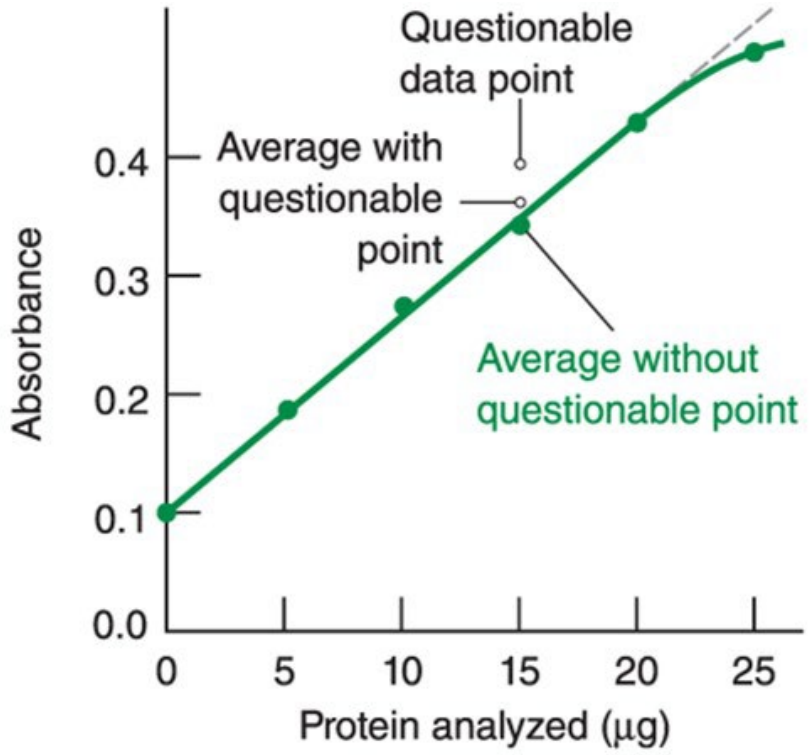
Calibration Curves

A **calibration curve** shows the response of an analytical method to known quantities of analyte.

- **Standard solutions** contain known concentrations of analyte.
- **Blank solutions** contain all reagents and solvents used in the analysis, but contain *no* deliberately added analyte.

A spectrophotometer measures the absorbance of light (*y*-axis), which is proportional to the quantity of protein analyzed (*x*-axis).

Figure 4-12



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

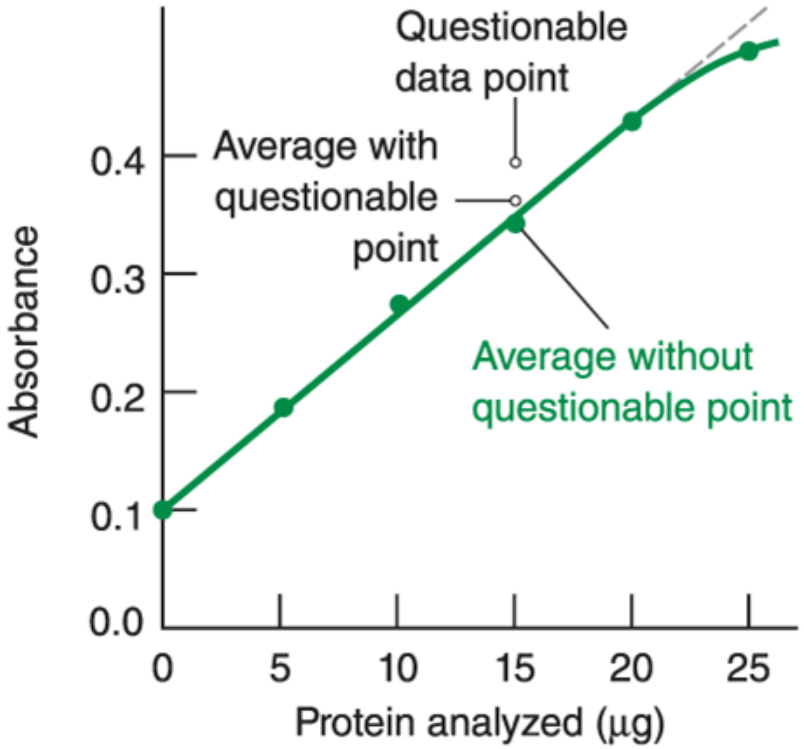
Table 4-8 Spectrophotometer data used to construct calibration curve

Amount of protein (μg)	Absorbance of independent standards			Range	Corrected absorbance		
0	0.099	0.099	0.100	0.001	-0.000_3	-0.000_3	0.000_7
5.0	0.185	0.187	0.188	0.003	0.085_7	0.087_7	0.088_7
10.0	0.282	0.272	0.272	0.010	0.182_7	0.172_7	0.172_7
15.0	0.345	0.347	(0.392)	0.047	0.245_7	0.247_7	—
20.0	0.425	0.425	0.430	0.005	0.325_7	0.325_7	0.330_7
25.0	0.483	0.488	0.496	0.013	0.383_7	0.388_7	0.396_7

Constructing a Calibration Curve (1 of 2)

1. Prepare known samples of analyte covering the range (0 to 150%) of concentrations expected for unknowns.
    - Tabulate amount of analyte in each standard and response.
2. Subtract the average absorbance of the blank solutions from each measured absorbance (corrected absorbance).
    - Blanks measure the response of the procedure when no analyte is present.
3. Make a graph of corrected absorbance vs. quantity of analyte.
    - Inspect the graph for linearity, outliers, and consistent *y*-uncertainty.

Figure 4-12



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

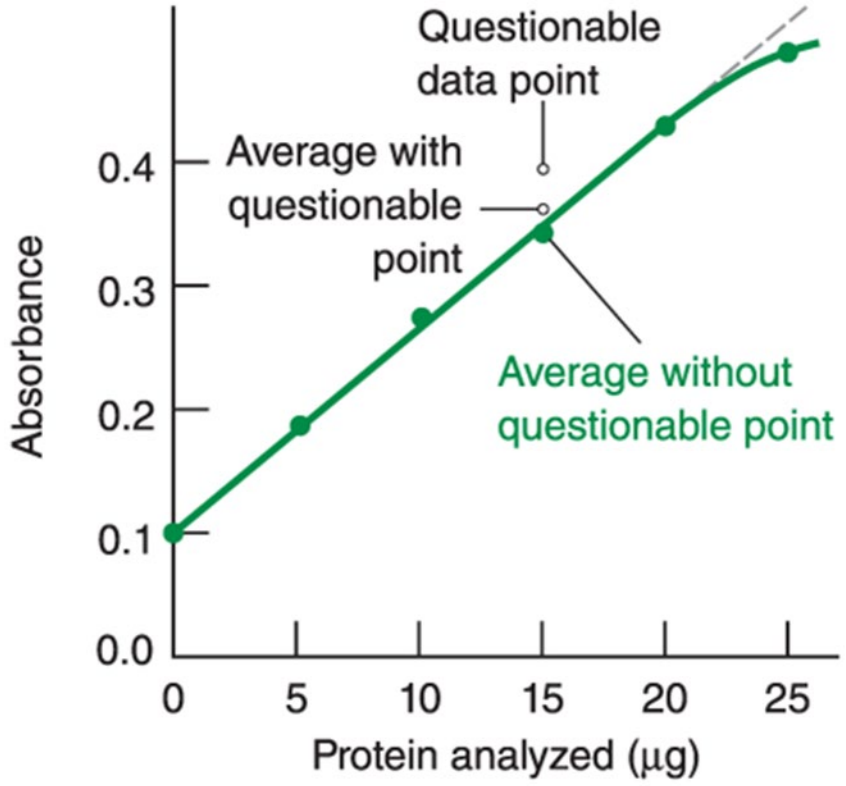
Constructing a Calibration Curve (2 of 2)

4. Use the least-squares procedure to find the best straight line through the linear portion of the data.

$$\underbrace{\text{Corrected absorbance}}_{y} = (0.016\,3_0)\underbrace{(\mu\text{g protein})}_{x} + 0.004_7$$

5. If you analyze an unknown at a future time, run a blank *at that time*.
    - Subtract the new blank signal from the unknown to correct.

Figure 4-12



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Example: Using a Linear Calibration Curve (1 of 3)

An unknown protein sample gave an absorbance of 0.406, and a blank had an absorbance of 0.104. How many micrograms of protein are in the unknown?

$$\underbrace{\text{Corrected absorbance}}_{y} = (0.016\,3_0)\underbrace{(\mu\text{g protein})}_{x} + 0.004_7$$

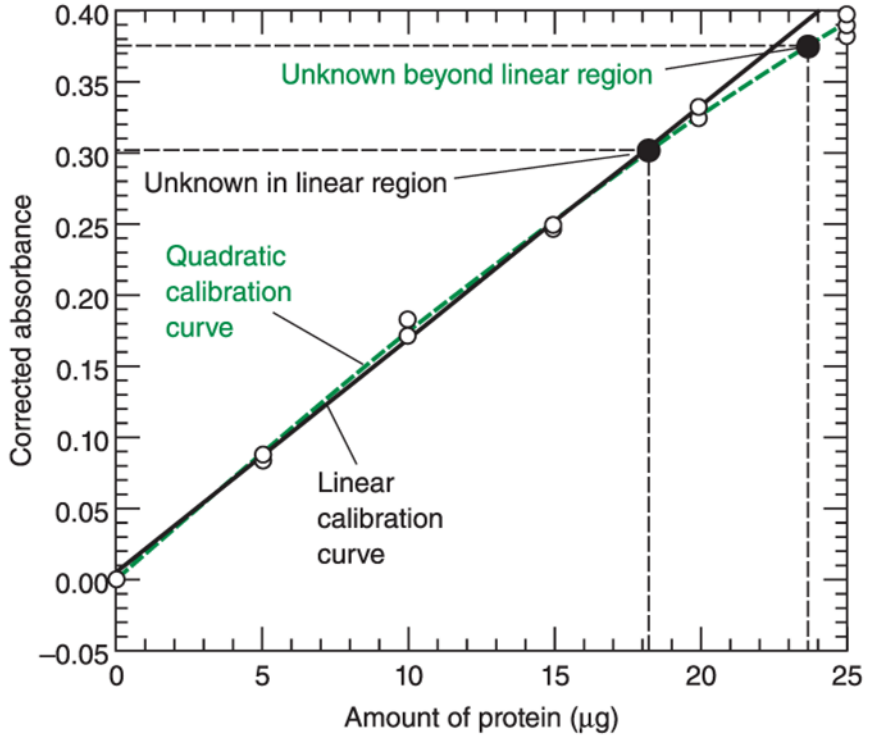
Example: Using a Linear Calibration Curve (2 of 3)

Solution: The corrected absorbance is 0.406 − 0.104 = 0.302, which lies on the linear portion of the calibration curve in Figure 4-13. Rearranging Equation 4-25 gives:

$$\mu g \text{ of protein} = \frac{\text{corrected absorbance} - 0.004_7}{0.016\,3_0}$$

$$= \frac{0.302 - 0.004_7}{0.016\,3_0} = 18.2_4\ \mu g$$

Figure 4-13



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020
W. H. Freeman and Company

Example: Using a Linear Calibration Curve (3 of 3)

Test Yourself: What mass of protein gives a corrected absorbance of 0.250?

$$\underbrace{\text{Corrected absorbance}}_{y} = (0.016\,3_0)\underbrace{(\mu\text{g protein})}_{x} + 0.004_7$$

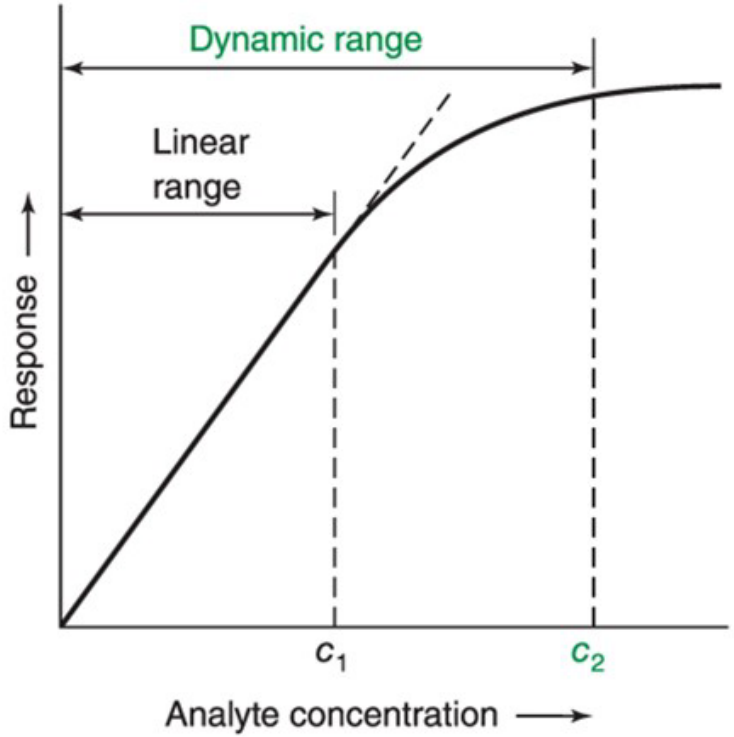
Linear Response

The **linear range** of an analytical method is the analyte concentration range over which response is proportional to concentration.

Dynamic range is the concentration range over which there is a measurable response to analyte, even if the response is not linear.

- Calibration procedures with a *linear response* are preferred.
- Corrected analytical signal $\propto$ quantity of analyte.
- It is possible to obtain valid results beyond the linear region by fitting with a nonlinear equation.

Figure 4-14



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Box 4-2 Using a Nonlinear Calibration Curve

- Consider an unknown whose corrected absorbance of 0.375 lies beyond the linear range.
- Fit all the data points with a quadratic equation:

$$y = -1.1_7 \times 10^{-4} x^2 + 0.0185\,5_8 x - 0.000\,7$$

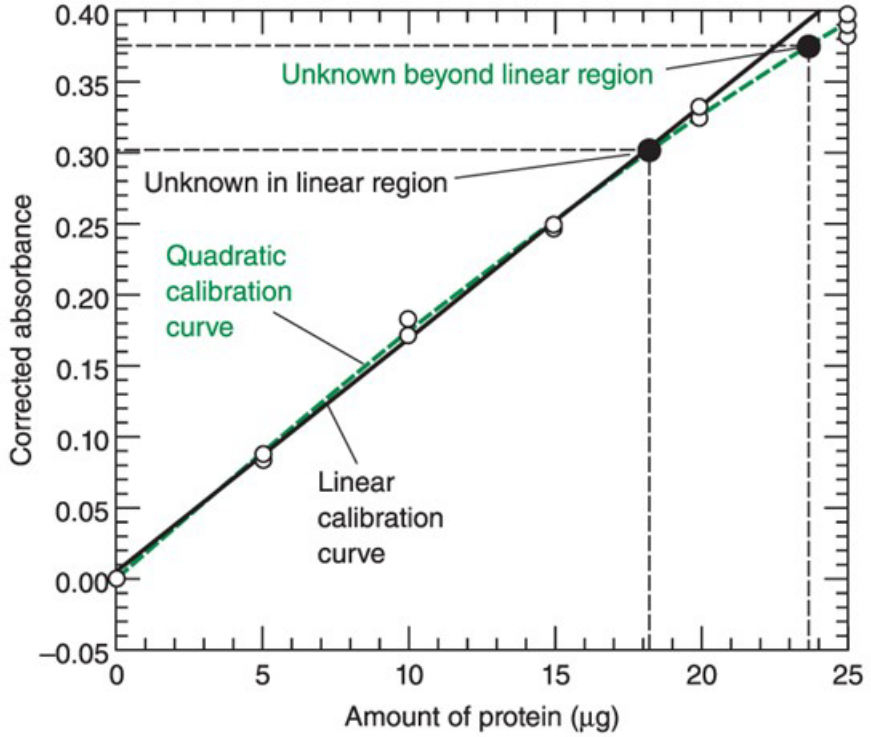
- Insert $y = 0.375$ into the equation and rearrange to the form

$$ax^2 + bx + c = 0$$

- Solve for x.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = 135\ \mu g \text{ or } 23._8\ \mu g$$

Figure 4-13

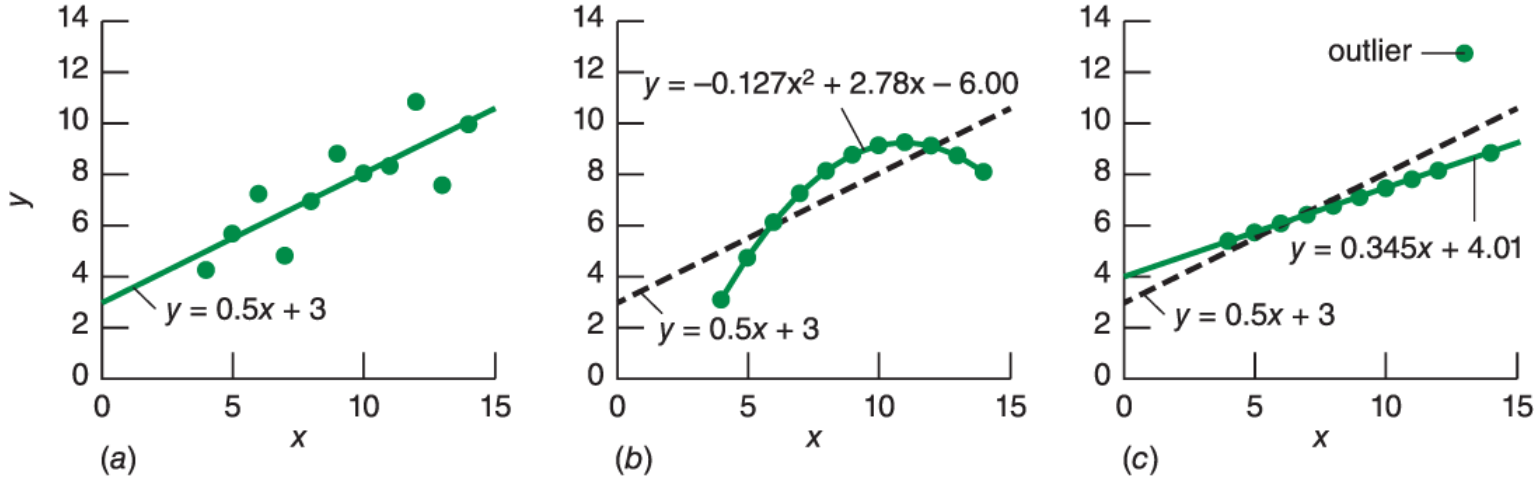


Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020
W. H. Freeman and Company

Good Practice (1 of 2)

Always make a graph of your data

Figure 4-15



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

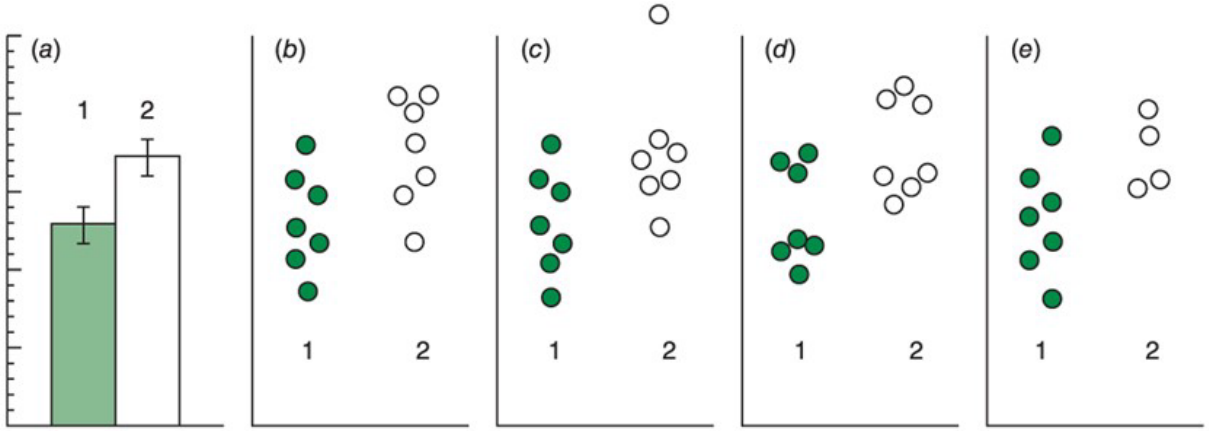
- Helps reject bad data, stimulus to repeat a measurement, or decision that a straight line is not appropriate
- All three data sets were fit to $y = 0.5x + 3$

Good Practice (2 of 2)

- It is not reliable to extrapolate any calibration curve beyond the measured range of standards.
- At *least* six calibration concentrations and two replicate measurements of each unknown are recommended.
- Make each standard solution from a certified material.
- Avoid serial dilution of a single stock solution (serial dilution propagates systematic error).
- Measure calibration solutions in random order, not in consecutive order of increasing concentration.

Box 4-3 Importance of Graphs to Visualize Data

A good graph reveals key characteristics of data and guides statistical analysis.



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

- Heights on the bar graph (a) give the mean values of two data sets. The error bars correspond to ±standard deviation of the mean.
- Data plots (b–e) show different characteristics of the data that are not evident in the bar graph.

Propagation of Uncertainty with a Calibration Curve

- An unknown with a corrected absorbance of $y = 0.302$ had a protein content of $x = 18.2_4$ µg. What is the uncertainty in x?
- Standard uncertainty in x = standard deviation of the mean =

$$u_x = \frac{s_y}{|m|}\sqrt{\frac{1}{k}+\frac{1}{n}+\frac{(y-\bar{y})^2}{m^2\sum(x_i-\bar{x})^2}}$$

k = number of replicate measurements

n = number of data points

$$u_x = \pm 0.2_3 \text{ µg}$$

- Confidence interval for x is $\pm tu_x$, where t is Student's t for $n - 2$ degrees of freedom

$$\pm tu_x = \pm(2.179)(0.2_3) = \pm 0.5_0 \text{ µg}$$

Section 4-9
A Spreadsheet for Least Squares

Figure 4-16: Spreadsheet for Linear Least-Squares Analysis

	A	B	C	D	E	F	G	H	I
1	Least-Squares Spreadsheet								
2									
3	Highlight cells B10:C12	x	y						
4	Type "= LINEST(C4:C7,	1	2						
5	B4:B7,TRUE,TRUE)	3	3						
6	For PC, press	4	4						
7	CTRL+SHIFT+ENTER	6	5						
8	For Mac, press								
9	COMMAND+RETURN	LINEST output:							
10	m	0.6154	1.3462	b					
11	u_m	0.0544	0.2141	u_b					
12	R^2	0.9846	0.1961	s_y					
13									
14	n =	4	B14 = COUNT(B4:B7)						
15	Mean y =	3.5	B15 = AVERAGE(C4:C7)						
16	$\Sigma(x_i - \text{mean } x)^2$ =	13	B16 = DEVSQ(B4:B7)						
17									
18	Measured y =	2.72	Input						
19	k = Number of replicate measurements of y =	1	Input						
20	Derived x =	2.2325	B20 = (B18-C10)/B10						
21	u_x =	0.3735	B21 = (C12/ABS(B10))*SQRT((1/B19)+(1/B14)+((B18-B15)^2)/(B10^2*B16))						

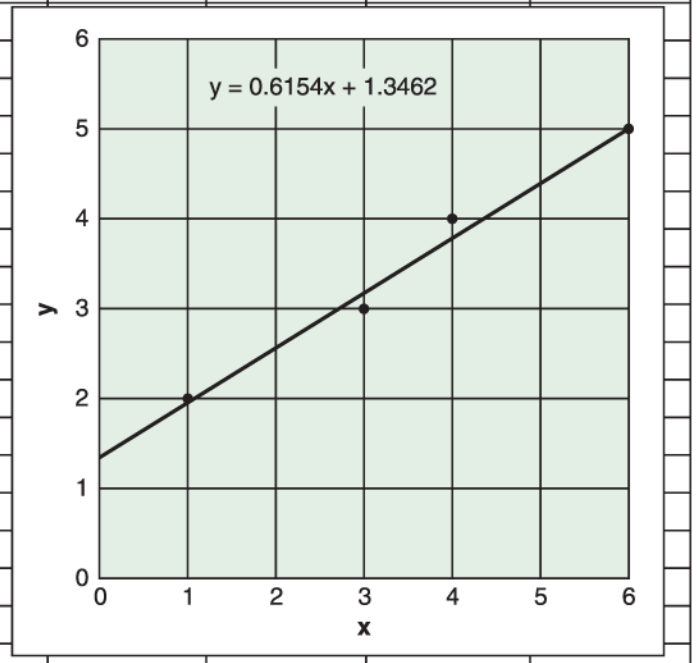


Figure 4-17: Adding Error Bars to a Graph

	A	B	C	D	E	F	G	H
1	Adding error bars to a graph							
2	Protein	Mean	Std dev	95%				
3	(μg)	absorbance	absorbance	confidence				
4	0	0.0993	0.00058	0.0014				
5	5	0.1867	0.00153	0.0038				
6	10	0.2753	0.00577	0.0143				
7	15	0.3613	0.02658	0.0660				
8	20	0.4267	0.00289	0.0072				
9	25	0.4890	0.00656	0.0163				
10								
11	95% confidence: D4 = T.INV.2T(0.05,2)*C4/SQRT(3)							
12								
13	Excel 2016 instructions:							
14	To insert error bars for uncertainties in column D:							
15	Double click in graph to get Chart Tools/Design ribbon							
16	Click Add Chart Element, select Error Bars							
17	Choose More Error Bar Options							
18	In Error Bar Options menu, click bar graph icon							
19	Click Custom and Specify Value							
20	For Positive Error Value, highlight D4:D9							
21	For Negative Error Value, highlight D4:D9							
22	Click OK and error bars appear on the chart							
23	Click on one x error bar and press Delete to remove all x error bars