Biological Chemistry Laboratory
Biology 3515/Chemistry 3515
Spring 2022

Lecture 7:
More on Significant Figures

and

Curve Fitting

Tuesday, 1 February 2022

©David P. Goldenberg
University of Utah
goldenberg@biology.utah.edu

Computer Labs

- Computer Labs next week and the following week.
  - Start at 2:00 PM!
  - Room 150 Biology Building
- This week: Graphing and curve fitting with SciDAVis.
- Next week: Molecular modeling with PyMOL.
- We will use the computers in the lab, not personal laptops.
- But, you should still install SciDAVis and PyMOL on your own computer. Use the versions available on Canvas.

Rules for Significant Figures

- Multiplication and division:

  The calculated result should contain the number of significant figures of the measured quantity with the smallest number of significant figures.

$$15\,\mathrm{g} \div 121.1\,\mathrm{g/mol} = 0.12\,\mathrm{mol}$$

$$\begin{aligned} 15\,\mathrm{mM} \times 25\,\mu\mathrm{L} &= 0.015\,\mathrm{moles/L} \times 2.5 \times 10^{-5}\,\mathrm{L} \\ &= 3.8 \times 10^{-7}\,\mathrm{moles} \\ &= 0.38\,\mu\mathrm{moles} \end{aligned}$$

Rules for Significant Figures

- For addition and subtraction:
  - The last decimal place of the result is determined by last decimal place of the measured quantity with the smallest number of decimal places.

    $$125\,\mathrm{g} + 0.035\,\mathrm{g} = 125\,\mathrm{g}$$

  - The big message: The number of significant figures in a calculated value should not imply more precision than is present in the values going into the calculation!

The Problem with Significant Figures

- Consider two measurements, made with the same balance:
  - 10.3 g
  - 94.5 g
- Both are represented with 3 significant figures.
- Are they equally precise?
  - On an absolute scale? *e.g.,* $\pm$0.1 g?
  - On a relative scale? *e.g.,* $\pm$%1?
- In order to know, we really need multiple measurements and some statistics!

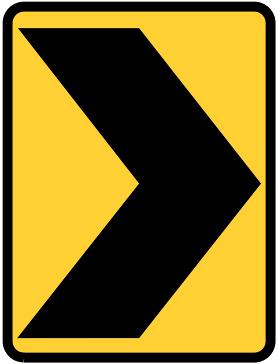
How Many Significant Figures for Our Measurements?

- A common approach: Use the number of digits displayed by the instrument.
  Is that valid?
- To determine precision, we need multiple measurements!
- What do your pipette calibration data suggest?
- Most often, the uncertainty over a range of values is a fixed amount, such as 2 μL, rather than a constant fraction or percentage.
- For the methods that we will be using, volume measurements are the usually the largest sources of error, and, with good technique, we can expect precision corresponding to two or three significant figures.

Significant Figures for this Class

- Rarely have precision greater than three significant figures.
- Can assume that concentrations of prepared solutions have precision of two significant figures.
- Spectrophotometric measurements can have precision of three significant figures.
- Reporting three rather than two significant figures isn't a major sin. (In this class!)
- Reporting more than three significant figures probably is!
- *Do* use extra decimal places for intermediate calculations, to avoid round-off errors.
- Use scientific notation when $|x| < 0.01$ or $|x| > 100$

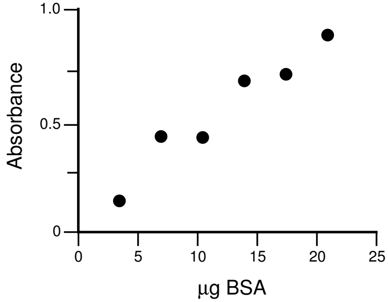
Warning!



Direction Change

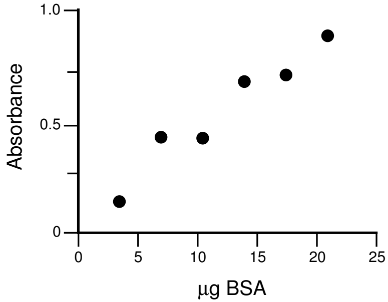
Curve Fitting

The Curve-Fitting Problem



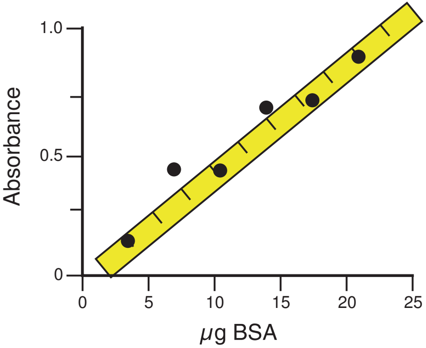
- How do we find the equation of the line (or other function) that best “fits” the experimental data?
- What assumptions do we make when fitting data to a function?
- How we determine how well the function (model) fits the data?

The Curve-Fitting Problem



- Two general kinds of situations:
  - We have a theoretical model that predicts a particular function, and we want to both test the model and estimate parameters that define the model.
  - We don't have a particular model in mind, but there is an empirical relationship that fits the data, and we want to estimate the parameters.
- Which category does the Bradford calibration fall into?

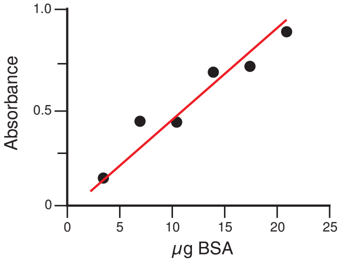
The Ruler Method



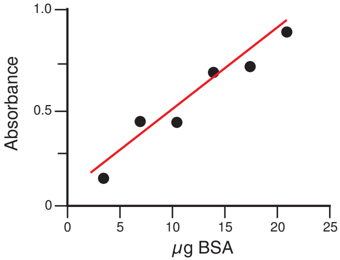
- Advantages?
  - It's easy and inexpensive!
- Disadvantages?
  - It's subjective. Different individuals will get different results.
    The more scatter, the bigger the problem!
  - Pretty much limited to the straight-line function.
  - Doesn't provide a measure of how well the data fit the function.

Clicker Question #1

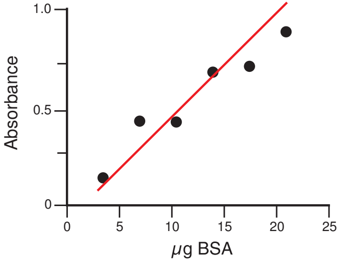
Which is the “best” fit?



A)



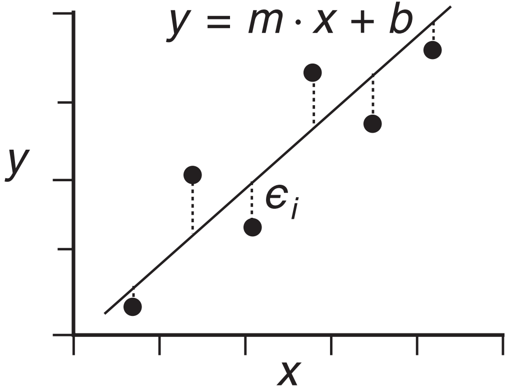
B)



C)

All answers count (for now)!

The Method of Least Squares



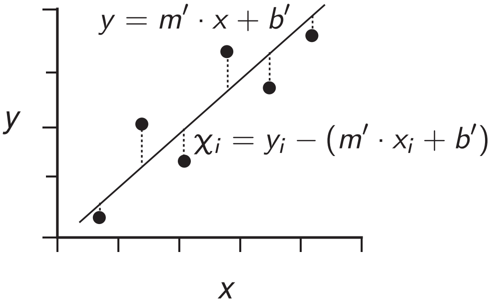
- Key Assumption: The experimental values of y are determined by a linear (or other) function of x *and* random error in the measurements.

$$y_i = m \cdot x_i + b + \epsilon_i$$

m and b are the “true” values of the parameters, and ϵ_i is the error in each measurement.

Chose an Arbitrary Line Given by the Equation:

$$y = m' \cdot x + b'$$

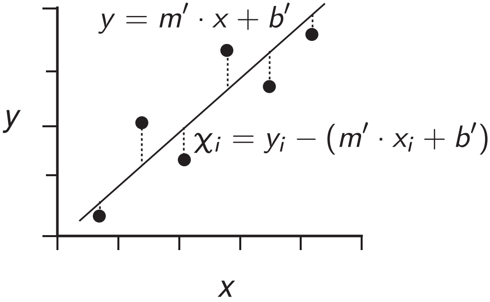


- m' and b' are not the "true" values of m and b, but we want to find values that are the best estimates of the true values.
- The observed values of y are then described by the equation:

$$y_i = m' \cdot x_i + b' + \chi_i$$

- χ_i is referred to as the "residual" for each point, and is distinct from the error, ϵ_i. (We never really know the values of ϵ_i.)

χ^2: The Sum of the Squares of the Residuals

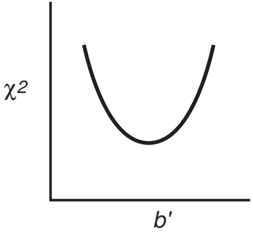


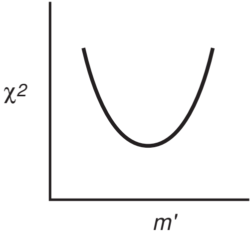
$$\chi^2 = \sum \chi_i^2$$
$$= \sum (y_i - (m' \cdot x_i + b'))^2$$

- Adjust m' and b' to minimize the value of χ^2 for the particular values of x_i and y_i in the experimental data set.
- Why are the residuals squared?

Minimization of χ^2

- χ^2 is a function of both b' and m'





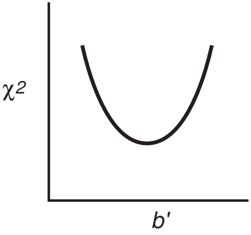
- How do we find values of b' and m' that minimize χ^2?
- Take derivatives of χ^2 with respect to b' and m' and set them to zero.

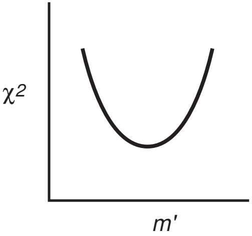
$$\frac{d}{db'} \sum (y_i - (m' \cdot x_i + b'))^2 = 0$$

$$\frac{d}{dm'} \sum (y_i - (m' \cdot x_i + b'))^2 = 0$$

Two equations, with unknowns b' and m'.

Minimization of χ^2





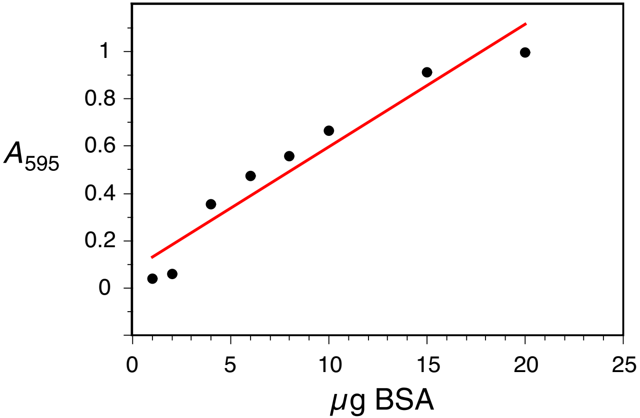
- For a linear function, there is a (relatively) simple solution to the two equations:

$$b' = \frac{\sum x_i^2 \sum y_i^2 - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$m' = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

N is the number of experimental x, y pairs.

A Linear Least-squares Fit to Bradford Calibration Data



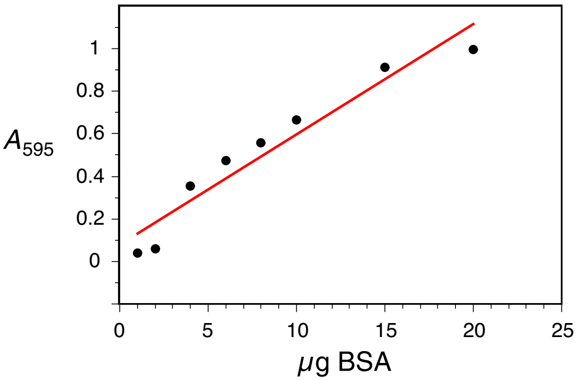
- The estimated parameters for $y = mx + b$:

$$m = 0.052 \pm 0.006$$

$$b = 0.08 \pm 0.06$$

- The uncertainties are analogous to the standard error of the mean.
- What are the units for the parameters?

Clicker Question #2



For the fit function, $y = mx + b$, what are the units of m?

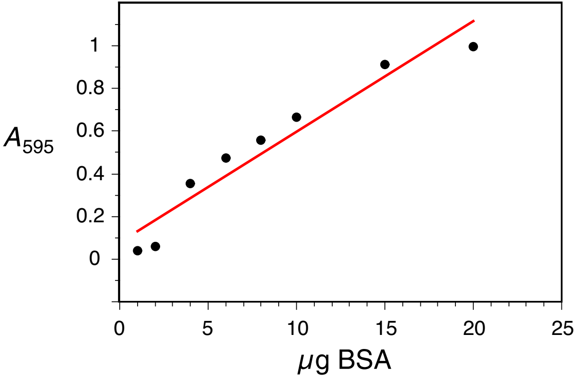
A) None (m is dimensionless)

B) μg^{-1}

C) μg

D) $\mu g/mL$

Clicker Question #3



For the fit function, $y = mx + b$, what are the units of b?

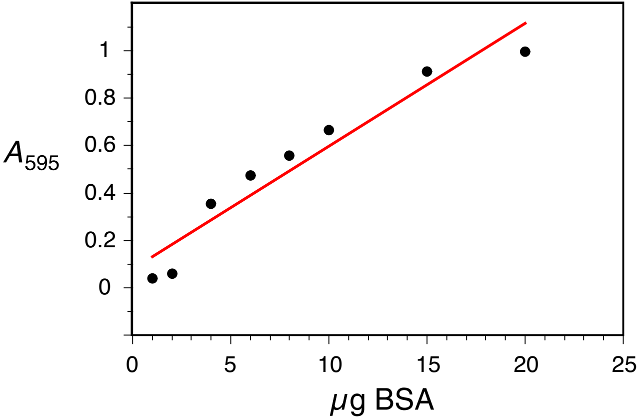
A) None (b is dimensionless)

B) μg^{-1}

C) μg

D) $\mu g/mL$

How Do We Judge the Goodness of Fit?

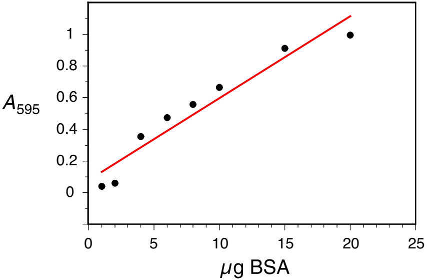


- The final, minimized, value of χ^2 for this fit:

$$\chi^2 = 0.062$$

- Can use this value to compare this fit to fits of other functions to the same data.
- But, actual value does not have a clear meaning.
- Adding more measurements will almost always increase χ^2.

The Coefficient of Determination, R^2



- What fraction of the total variation of y-values is accounted for by the fit function?
- For this fit:

$$R^2 = 0.93$$

Compare χ^2 to the total variation in y-values.

- Define the total sum of squares:

$$SS_{\text{tot}} = \sum (y_i - \bar{y})^2$$

$\bar{y}$ = mean of y-values.

- The ratio:

$$\frac{\chi^2}{SS_{\text{tot}}}$$

represents the fraction of the variation that is *not* accounted for by the fit function.

- The fraction of variation that *is* accounted for the function:

$$R^2 = 1 - \frac{\chi^2}{SS_{\text{tot}}}$$

R^2 should lie between 0 and 1.