

Error Propagation in Numerical Analysis

Most numerical results in science are the end result of an analysis that involves several mathematical calculations applied to raw data. Raw data are virtually always subject to measurement errors and the question inevitably arises as to how these measurement errors affect the final numerical result.

Measurement errors are usually estimated by replicate measurements that are assumed to obey a normal (binomial) distribution centered around the mean. This mean is usually assumed to represent the exact value of the quantity being measured, and this is a good approximation in the absence of systematic errors in the measurement.

Unfortunately, it would take an inordinately large number of individual measurements to perfectly define the normal distribution and the exact mean. Thus, we are forced to use a limited number of measurements to *estimate* the distribution and the mean (μ). The distribution is characterized by its half-width, called the (estimated) standard deviation (esd, σ). Using these estimates of the correct measurement and the errors involved, combined with the formulae below, allows an estimate of the mean and esd of the distribution of the final numerical result.

Error Propagation Formulae

These formulae assume that measurements u and v have esd's σ_u and σ_v . The result of an individual calculation is called x and we are always trying to estimate the esd of x , σ_x .

$$x = f(u, v) \qquad \sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u} \right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v} \right)^2$$

$$x = au \pm bv \qquad \sigma_x^2 = a^2 \sigma_u^2 + b^2 \sigma_v^2$$

$$\begin{aligned} x &= \pm auv \\ x &= \pm au/v \end{aligned} \qquad \frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2}$$

$$x = au^{\pm b} \qquad \frac{\sigma_x}{x} = b \frac{\sigma_u}{u}$$

$$x = a \exp(\pm bu) \qquad \frac{\sigma_x}{x} = b \sigma_u$$

$$x = a \ln(\pm bu) \qquad \sigma_x = a \frac{\sigma_u}{u}$$

Determination of Metal/Protein Ratio

[Metal] determination

[Metal] is often determined by ICP-AES (Inductively Coupled Plasma-Atomic Emission Spectrometry) or ICP-MS (ICP-Mass Spectrometry) in units of ppm metal. Ppm is parts-per-million based on mass...

$$\text{ppm metal } (M_p) = \frac{\text{g metal}}{10^6 \text{ g sample}} = \frac{\mu\text{g metal}}{\text{g sample}}$$

...the sample is mostly water with a density of 1 g/mL...

$$\text{ppm metal } (M_p) = \frac{\mu\text{g metal}}{\text{g sample}} = \frac{\mu\text{g metal}}{\text{mL sample}}$$

...conversion to molarity (M) requires the atomic mass of the metal, M_{at} , in units of g/mol or $\mu\text{g}/\mu\text{mol}$...

$$\mu\text{g metal} = M_{at} \times \mu\text{mol metal}$$

$$\text{ppm metal } (M_p) = \frac{\mu\text{g metal}}{\text{mL sample}} = \frac{M_{at} \times \mu\text{mol metal}}{\text{mL sample}}$$

$$\text{ppm metal } (M_p) = \frac{M_{at} \times \text{mmol metal}}{\text{L sample}} = M_{at} \times \text{mM metal}$$

$$\text{mM metal } (M_m) = \frac{\text{ppm metal } (M_p)}{\text{Atomic mass } (M_{at})}$$

[Protein] determination

[Protein] is usually determined by BCA or Coomassie (or similar) assay in units of mg/mL. Conversion to molarity (M) requires the protein molecular weight, MW, in units of g/mol (or mg/mmol)...

$$P_g = \frac{\text{mg}}{\text{mL}} = \frac{\text{mg}}{\text{mmol}} \times \frac{\text{mmol}}{\text{mL}} = \text{MW} \times \frac{\text{mmol}}{\text{mL}} = \text{MW} \times \frac{\text{mol}}{\text{L}}$$

$$M_{\text{protein}} = (\text{mg protein}/\text{mL})/\text{MW}$$

or

$$P_m = \text{mM protein} = 1000 \times M = P_g(\text{mg protein}/\text{mL})/(\text{MW}/1000)$$

Error Analysis

First, let's define some terminology...

Call	M_p	[metal] in ppm	with esd	$\sigma(M_p)$
	M_m	[metal] in mM	with esd	$\sigma(M_m)$
	M_{at}	metal atomic weight	with esd	$\sigma(M_{at})$
	P_g	[protein] in mg/mL	with esd	$\sigma(P_g)$
	P_m	[protein] in mM	with esd	$\sigma(P_m)$
	MW	molec. wt. protein in g/mol	with esd	$\sigma(MW)$
	R	[metal]/[protein] ratio	with esd	$\sigma(R)$

The metal concentration was calculated as...

$$M_m = M_p / M_{at}$$

...and the error in M_m can be calculated from...

$$\frac{\sigma(M_m)^2}{M_m^2} = \frac{\sigma(M_p)^2}{M_p^2} + \frac{\sigma(M_{at})^2}{M_{at}^2}$$

...but, relatively speaking, the error in the atomic weight is insignificant, $\sigma(M_{at})^2 \approx 0$, thus...

$$\frac{\sigma(M_m)}{M_m} = \frac{\sigma(M_p)}{M_p} \quad \text{or} \quad \sigma(M_m) = \frac{\sigma(M_p)}{M_{at}}$$

The protein concentration was calculated as...

$$P_m = P_g / (MW/1000)$$

...and the error in P_m can be calculated from...

$$\frac{\sigma(P_m)^2}{P_m^2} = \frac{\sigma(P_g)^2}{P_g^2} + \frac{\sigma(MW)^2}{MW^2}$$

...if MW is from a known sequence or from ESI-MS, then $\sigma(MW) \approx 0$. However, if MW is from gel filtration or PAGE, then $\sigma(MW)$ is very significant. The metal/protein ratio, R, is calculated as...

$$R = \text{mM metal} / \text{mM protein} = M_m / P_m$$

...and the error in R can be calculated from...

$$\frac{\sigma(R)^2}{R^2} = \frac{\sigma(M_m)^2}{M_m^2} + \frac{\sigma(P_m)^2}{P_m^2}$$

Example Calculation

Metal Determination

When determining metal content by ICP-AES or ICP-MS, one usually submits replicate (at least 3) samples, along with replicate blanks. The blanks are preferably as similar to the medium containing the metalloprotein as possible; often this can be the ultrafiltrate from the final concentration step that prepared the metalloprotein samples.

In the following example, three samples each of: (a) buffer blank; (b) [Co]hTFIIB-NTD; and (c) [Zn]hTFIIB-NTD are used. [Metal] is measured by ICP-AES, which provides concentrations of a number of metals simultaneously. Here is a partial list of the results of such a determination:

Metal	Buff1	Buff2	Buff3	CoNTD1	CoNTD2	CoNTD3	ZnNTD1	ZnNTD2	ZnNTD3
Cd	0	0	0	0	0.01280	0.02346	0.06505	0.01209	0.00853
Co	0.01751	0	0.05650	5.7392	6.2491	6.3245	0.2945	0.09914	0.05431
Fe	0	0	0	0	0	0	0	0	0
Mn	0.21149	0.25337	0.25655	0.13863	0.15616	0.1968	0.24779	0.30117	0.25974
Ni	0	0.00932	0.0958	0.11548	0.1248	0.00932	0.04712	0.05800	0.00518
Zn	0.69187	0.65564	0.69067	0.83272	0.87967	0.83558	4.0127	4.1436	4.1062

Comparing [metal] for [Co]- and [Zn]hTFIIB-NTD samples against the buffer backgrounds is convincing that only the expected metal is present significantly above background. We need to put this qualitative observation on firm statistical ground. The first step is to calculate the mean (μ) and estimated standard deviation (σ) for the appropriate metals in all three samples (buffer blank, [Co]-, and [Zn]hTFIIB-NTD). For any set of observations y_i ($i = 1-n$), the mean and esd are given by:

$$\mu = (\sum y_i)/n \qquad \sigma^2 = [\sum (y_i - \mu)^2]/(n - 1)$$

For example, using the Co determinations for the three [Co]hTFIIB-NTD samples,

	y_i	$y_i - \mu$	$(y_i - \mu)^2$
CoNTD1	5.7392	-0.3651	0.13327
CoNTD2	6.2491	0.1448	0.02098
CoNTD3	6.3245	0.2202	0.04850
Σ	18.3128		Σ 0.20275
μ	6.1043		σ 0.31839

So, for the [Co]hTFIIB-NTD samples, [Co] = 6.1 ± 0.3 ppm. For further calculations, it is often useful to keep additional (non-significant) figures to avoid roundoff error: [Co] = 6.104 ± 0.318 ppm

Analogous calculations for the other samples yields

	Buffer	[Co]hTFIIB-NTD	[Zn]hTFIIB-NTD
[Co]	0.025 ± 0.029 ppm	6.104 ± 0.318 ppm	0.149 ± 0.128 ppm
[Zn]	0.679 ± 0.021 ppm	0.849 ± 0.026 ppm	4.088 ± 0.067 ppm

The purpose behind measuring the [Co] and [Zn] concentrations of the blank (buffer) was to allow correction for the amount of these metals in the buffer. For example the corrected concentration of [Zn] in the [Zn]hTFIIB-NTD sample is calculated as

$$[\text{Zn}]_{\text{Z,corr}} = [\text{Zn}]_{\text{Z}} - [\text{Zn}]_{\text{B}} = 4.088 - 0.679 \text{ ppm} = 3.409 \text{ ppm}$$

$$\sigma^2([\text{Zn}]_{\text{Z,corr}}) = \sigma^2([\text{Zn}]_{\text{Z}}) + \sigma^2([\text{Zn}]_{\text{B}}) = (0.067 \text{ ppm})^2 + (0.021 \text{ ppm})^2 = 0.00493 \text{ ppm}^2$$

$$\text{or } \sigma([\text{Zn}]_{\text{Z,corr}}) = 0.070 \text{ ppm}$$

Corrected [M] for hTFIIB-NTD samples

	[Co]hTFIIB-NTD	[Zn]hTFIIB-NTD
[Co]	6.079 ± 0.319 ppm	0.124 ± 0.131 ppm
[Zn]	0.170 ± 0.033 ppm	3.409 ± 0.070 ppm

Now we need to convert these concentrations to units of μM . We need the atomic weights (M_{at}) for Co (58.93 g/mol) and Zn (65.39 g/mol). Since we consider the errors in these numbers to be insignificant, we can just divide the [M] in ppm by M_{at} and the $\sigma([\text{M}], \text{ppm})$ by M_{at} to get the numbers in terms of μM .

	[Co]hTFIIB-NTD	[Zn]hTFIIB-NTD
[Co]	103.16 ± 5.41 μM	2.104 ± 2.223 μM
[Zn]	2.600 ± 0.505 μM	52.13 ± 1.07 μM

[Protein] determination

Determination of protein concentration in the absence of a definitive ϵ_{280} depends on a calibration curve (usually using known concentrations of bovine serum albumin, BSA). Again, performing replicate measurements of A for each concentration of BSA is needed to get a feeling of the precision of the calibration. For example, here are triplicate measurements of a 4-point calibration

[P], mg/mL	A1	A2	A3
0.125	0.187	0.242	0.158
0.250	0.325	0.247	0.257
0.500	0.550	0.562	0.584
1.000	0.991	0.948	0.908

The standard treatment is to fit the A vs [P] calibration curves using linear least squares and then use the slope and intercept of this fit to calculate [P] from measured A values for each unknown. This is an appropriate analysis, as long as we allow for errors in the calibration measurements. We need a way to find esd's for the slope and intercept. This is statistically possible through a variance-covariance matrix associated with the least squares fit, and most good statistical programs (Mathematica, Maple, Igor) will calculate esd's of slope and intercept automatically when you perform the least squares fit.

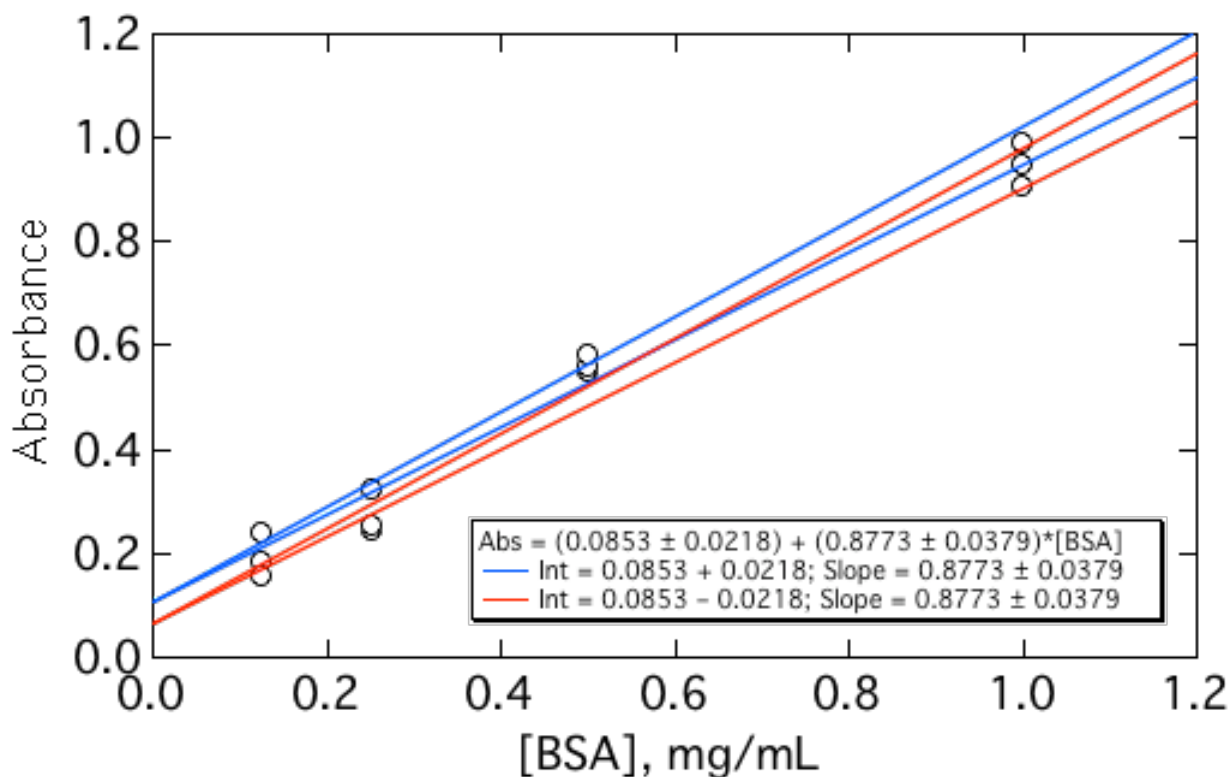
The plot on the next page indicates the linear least squares fit of these data, giving rise to the equation

$$A = I + S \times [\text{BSA}] \quad \text{or} \quad [\text{P}] = (A - I)/S$$

$$I = 0.08533 \quad S = 0.87733$$

$$\sigma(I) = 0.02185 \quad \sigma(S) = 0.03792$$

This equation, the parameter values, and proper error propagation formulae now allow us to use the errors in measurement of absorbance to calculate the errors in [P] determinations.



For the [Co]- and [Zn]hTFIIB-NTD samples and for a buffer blank, the protein assay yielded the following absorbance (A) values

Sample	A1	A2	A3	μ	σ
[Zn]hTFIIB-NTD	0.343	0.330	0.318	0.3303	0.0125
[Co]hTFIIB-NTD	0.717	0.707	0.727	0.7170	0.0100
Buffer	0.025	0.035	0.046	0.0353	0.0105

First we need the absorbance reading corrected for absorbance of the buffer blank. For example, for the [Zn]hTFIIB-NTD sample,

$$A_{Z,\text{corr}} = \mu_Z - \mu_{\text{buff}} = 0.3303 - 0.0353 = 0.2950$$

$$\sigma^2(A_{Z,\text{corr}}) = \sigma^2(A_Z) + \sigma^2(A_{\text{buff}}) = (0.0125)^2 + (0.0105)^2 = 0.000267$$

$$\sigma(A_{Z,\text{corr}}) = 0.0163$$

Performing the same calculations for [Co]hTFIIB-NTD yields

Sample	A_{corr}	$\sigma(A_{\text{corr}})$
[Zn]hTFIIB-NTD	0.2950	0.0163
[Co]hTFIIB-NTD	0.6817	0.0145

The next step is to convert these absorbance readings into [P] (in mg/mL). We use the least squares results from the BSA calibration curve for this. This is straightforward algebra, but we must consider the error propagation at each step of the calculation.

The first calculation involves subtracting the intercept (I) from the Abs reading (we will continue to carry through the [Zn]hTFIIB-NTD analysis).

$$A-I = A_{Z,corr} - I = 0.2950 - 0.0853 = 0.2097$$

$$\sigma^2(A-I) = \sigma^2(A_{Z,corr}) + \sigma^2(I) = (0.0163)^2 + (0.0218)^2 = 0.000743$$

$$\sigma(A-I) = 0.0273$$

Then this result needs to be divided by the slope (S) of the least squares line to give the protein concentration

$$[P]_Z = (A-I)/S = 0.2097 / 0.8773 = 0.2390 \text{ mg/mL}$$

$$\frac{\sigma([P]_Z)^2}{[P]_Z^2} = \frac{\sigma(A-I)^2}{(A-I)^2} + \frac{\sigma(S)^2}{S^2}$$

$$\frac{\sigma([P]_Z)^2}{(0.2390)^2} = \frac{(0.0273)^2}{(0.2097)^2} + \frac{(0.0379)^2}{(0.8773)^2}$$

$$\sigma([P]_Z)^2 = 0.001075$$

$$\sigma([P]_Z) = 0.0328 \text{ mg/mL}$$

Using the same procedure to get the protein concentration of the [Co]hTFIIB-NTD sample yields

Sample	[P], mg/mL	$\sigma([P])$, mg/mL
[Zn]hTFIIB-NTD	0.2390	0.0328
[Co]hTFIIB-NTD	0.6798	0.0419

To convert these concentrations to units of μM protein, we need the MW of the proteins. Apo-hTFIIB-NTD has a calculated MW of 6362.06 g/mol. Adding the atomic weights of the metals and subtracting the atomic weights of three protons (that must be removed from the protonated cysteine sidechains to accommodate the metals) yields 6.42445 and 6.41799 kDa for [Zn]- and [Co]hTFIIB-NTD, respectively.

Sample	[P], μM	$\sigma([P])$, μM
[Zn]hTFIIB-NTD	37.20	5.10
[Co]hTFIIB-NTD	105.92	6.53

Metal/Protein Ratio

Now we can combine the results of concentration determinations for metals and protein. Since everything is in the same units (μM), the ratios are trivial, but we need to consider the errors...

	[Co]hTFIIB-NTD	[Zn]hTFIIB-NTD
[Co]	103.16 \pm 5.41 μM	2.104 \pm 2.223 μM
[Zn]	2.600 \pm 0.505 μM	52.13 \pm 1.07 μM
[P]	105.92 \pm 14.12 μM	37.20 \pm 5.10 μM

Using the Co/protein ratio for [Co]hTFIIB-NTD as an example...

$$R = M_m/P_m = 103.16 \mu\text{M} / 105.92 \mu\text{M} = 0.9739$$

$$\frac{\sigma(R)^2}{R^2} = \frac{\sigma(M_m)^2}{M_m^2} + \frac{\sigma(P_m)^2}{P_m^2}$$

$$\frac{\sigma(R)^2}{(0.9739)^2} = \frac{(5.41)^2}{(103.16)^2} + \frac{(6.53)^2}{(105.92)^2}$$

$$\sigma(R) = 0.0788$$

Now that we have finished this calculation, we should trim the result to the correct number of significant figures. The correct report of the Co/protein ratio for this [Co]hTFIIB-NTD sample is:

$$[\text{Co}]/[\text{protein}] = 0.97 \pm 0.08$$

Thus, we have demonstrated that the [Co]hTFIIB-NTD sample contains one Co per protein monomer "within experimental error". Carrying through this analysis for the other metal determinations on this and the [Zn]hTFIIB-NTD sample yields the following summary...

	[Co]hTFIIB-NTD	[Zn]hTFIIB-NTD
[Co]/[prot]	0.97 ± 0.08	0.057 ± 0.060
[Zn]/[prot]	0.025 ± 0.005	1.40 ± 0.19

Alternatively, these can be reported in the following notation

	[Co]hTFIIB-NTD	[Zn]hTFIIB-NTD
[Co]/[prot]	0.97(8)	0.057(60)
[Zn]/[prot]	0.025(5)	1.40(19)

Conclusions

- [Co]hTFIIB-NTD contains one cobalt per protein within experimental error
- [Co]hTFIIB-NTD contains a small amount of zinc (at the level of 2-3% of protein)
- [Zn]hTFIIB-NTD contains negligible amounts of cobalt
- [Zn]hTFIIB-NTD contains ca. 20-60 % more zinc than protein. This implies that either:
 - the buffer analyzed as a blank did not reflect the amount of zinc in the buffer containing the protein
 - additional zinc beyond that expected to occupy the binding site is adventitiously bound to the protein