

The Weakest Link

A chemical analysis generally consists of three steps: obtaining a representative sample, preparing it for analysis, and taking the appropriate measurements. Each of these steps is subject to random errors that are characterized by a standard deviation, σ , or variance, σ^2 ; thus, σ_{samp}^2 is the variance for obtaining a sample, σ_{prep}^2 is the variance for preparing the sample, and σ_{meas}^2 is the variance for the measurements. Further subdivisions also are possible. In a spectrophotometric analysis, for example, we can divide σ_{meas}^2 into the variance due to the spectrometer's source, detector, and optics, σ_{spect}^2 , and the variance due to the positioning of the sample cell within the spectrometer, σ_{pos}^2 .

The step with the largest variance is the method's weakest link. Improving the weakest link's variance improves the method's overall precision, which is a good thing. Improving the variance of other steps without improving the weakest link has little to no effect. Knowing the value for each variance, therefore, is essential if you wish to improve an analytical method.

Determining values for these variance isn't trivial because results for the analysis of replicate samples gives the total variance for the analysis, σ_{total}^2 , which includes contributions from each step; thus

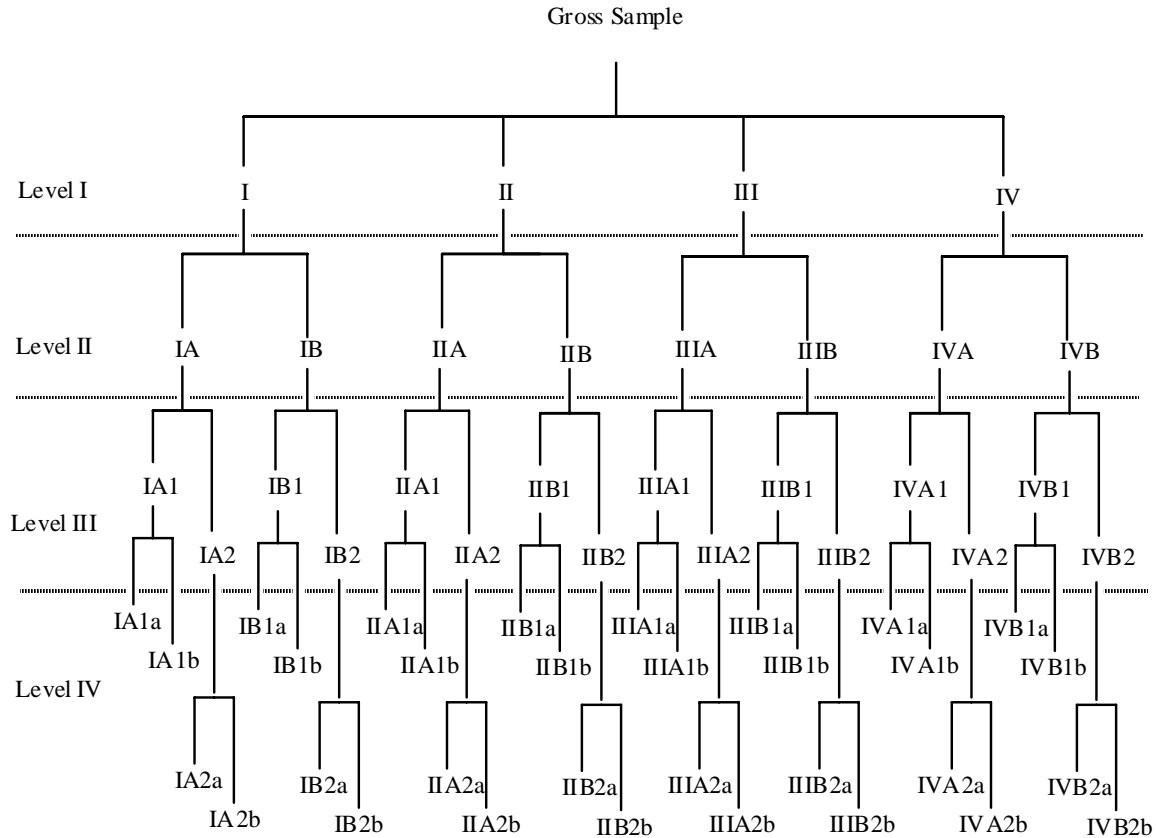
$$\sigma_{\text{total}}^2 = \sigma_{\text{samp}}^2 + \sigma_{\text{prep}}^2 + \sigma_{\text{meas}}^2$$

$$\sigma_{\text{total}}^2 = \sigma_{\text{samp}}^2 + \sigma_{\text{prep}}^2 + \sigma_{\text{pos}}^2 + \sigma_{\text{spect}}^2$$

A simple determination of σ_{total}^2 , therefore, does not provide enough information to partition the overall variance into its component parts.

One approach to determining the individual variances is to use a nested experimental design (see figure on the following page), which consists of several levels, where the number of levels equal to the number of parameters you wish to evaluate. In this experiment you will determine the contribution of sampling, sample preparation, sample cell positioning, and a spectrometer to the total variance of a spectrophotometric analysis, requiring four levels. The first level consists of four random samples from the gross sample (identified using the Roman numerals I, II, III, and IV). Duplicate samples from each Level I sample are prepared for analysis; these eight samples represents the second level and are identified using the uppercase letters A and B. For the third level, each Level II sample is divided in half before measuring its absorbance; these 16 samples are identified using the numbers 1 and 2. Finally, each Level III sample is placed in the spectrometer and its absorbance measured twice without repositioning the sample, providing the 32 Level IV samples. Note that each Level I sample is divided into two

Level II samples, that each Level II sample is divided into two Level III samples, and that each Level III sample is, in turn, divided into two Level IV samples. Look closely at Level IV and note that the **difference** between any two samples coming



from the **same** Level III sample is due to indeterminate errors in the spectrometer. The variance for these differences, s_{IV}^2 , provides an estimate for σ_{spect}^2 . Because the variance is determined using differences, the normal equation for a standard deviation is not used. Instead, the variance is given by

$$s_{IV}^2 = s_{\text{spect}}^2 = \frac{\sum (d_{IV})_i^2}{8n}$$

where d_{IV} is the difference between related Level IV samples (e.g. IA1a and IA1b), and n is the number of Level I samples (4 in this case). The variance s_{IV}^2 has $4n$ degrees of freedom.

The variance for Level III, s_{III}^2 , includes contributions from the spectrometer and the positioning of the sample cell; thus

$$s_{\text{III}}^2 = s_{\text{pos}}^2 + \frac{s_{\text{spect}}^2}{2} = \frac{\sum (d_{\text{III}})_i^2}{4n}$$

where d_{III} is the difference between related Level III samples (e.g. IA1 and IA2). The factor of 2 in the term for the spectrometer's variance accounts for the fact that each Level III sample includes results from two Level IV samples. The variance s_{III}^2 has $2n$ degrees of freedom.

The variance for the Level II samples, s_{II}^2 , includes contributions from the spectrometer, the positioning of the sample cell, and sample preparation; thus

$$s_{\text{II}}^2 = s_{\text{prep}}^2 + \frac{s_{\text{pos}}^2}{2} + \frac{s_{\text{spect}}^2}{4} = \frac{\sum (d_{\text{II}})_i^2}{2n}$$

where d_{II} is the difference between related Level II samples (e.g. IA and IB). The variance s_{II}^2 has n degrees of freedom.

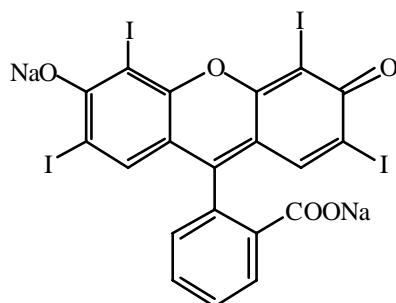
Finally, the variance for Level I, s_{I}^2 , is determined using the standard equation for the variance, and includes contributions from sampling, sample preparation, the positioning of the sample cell, and the spectrometer

$$s_{\text{I}}^2 = s_{\text{samp}}^2 + \frac{s_{\text{prep}}^2}{2} + \frac{s_{\text{pos}}^2}{4} + \frac{s_{\text{spect}}^2}{8} = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

where X_i is the result for each Level I sample and \bar{X} is the average result for all Level I samples. The variance s_{I}^2 has $n-1$ degrees of freedom.

Directions

Erythrosin B, whose structure is shown here, is a dye that has several analytical uses, including a biological stain for bacteria in soils, a plasma stain for nerve cells (when used in conjunction with methylene blue), a phosphorescent probe for studying the diffusion of



membrane proteins, and in the quantitative determination of phospholipids. It also is an acid-base indicator whose color is orange in strongly acidic solutions and red in aqueous solutions of pH greater than 3. And, of course, erythrosin B, which also is known as Acid Red 51 and FD&C red dye no. 3, has been used as a food coloring in maraschino cherries.

The sample to be analyzed is a mixture of erythrosin B (the analyte) and NaCl (the matrix). The erythrosin B adheres to the surface of the salt crystals, imparting a pink color to the NaCl (in fact, this is a common way to dispense indicators that are unstable in solution and, therefore, cannot be dispensed in solution form). You will be provided with approximately 50 g of this mixture, which represents your gross sample.

Pour your gross sample onto a large watchglass, shape it into a flattened cone, and divide the cone into quarters. Obtain an approximately 1-g sample from each quarter; these are your four Level I samples.

Transfer each Level I sample into a clean glass or agate mortar and grind each sample for several minutes to reduce the particle size and further homogenize the sample. Divide each of these processed samples in half and obtain an approximately 0.25-g sample from each half. Quantitatively transfer each sample to a 50-mL volumetric flask and dilute to volume with distilled water. Be sure to thoroughly mix each solution. These are your eight Level II samples.

Divide each of your Level II samples approximately in half by transferring into separate test tubes (any remaining Level II sample can be discarded). These are your 16 Level III samples.

Adjust the spectrometer to a wavelength of 526 nm. Using distilled water as a reference, measure the absorbance, A , of each Level III sample twice, without removing the sample cell from the spectrometer between measurements. These 32 absorbances are the results for Level IV. Calculate the concentration of erythrosin B in each Level IV sample using Beer's law

$$A = abC$$

where a is erythrosin B's absorptivity, which has a value of $0.0916 \text{ cm}^{-1} \text{ ppm}^{-1}$, b is the pathlength, which is 1.00 cm, and C is the concentration of erythrosin B in ppm. Convert these values to % w/w erythrosin B by accounting for the sample's mass and its dilution. To determine the % w/w erythrosin B for the Level III samples, average the associated results for Level IV. For example, the % w/w erythrosin B for sample IA1 is the average result for samples IA1a and IA1b. The results for the Level II and Level I samples are found in a similar manner.

Hazards

There are no specific health hazards associated with this experiment, although you should exercise appropriate caution when working with any chemicals.

Waste Disposal

All solutions may be discarded down the drain with water.

Report

Summarize your data using four tables. In the first table reports results for the Level IV samples using the following headings: sample ID, mass of sample, absorbance, and % w/w erythrosin B. The remaining tables reports results for the Level III, Level II, and Level I samples and include the following headings: sample ID and % w/w erythrosin B.

Using your data, calculate values for σ_{samp}^2 , σ_{prep}^2 , σ_{pos}^2 , σ_{spect}^2 , and σ_{total}^2 . In addition, briefly answer the following questions

1. Are the differences between σ_{samp}^2 , σ_{prep}^2 , σ_{pos}^2 , and σ_{spect}^2 statistically significant?
2. Based on your results, which step is the weakest link in this analysis?
3. How might you go about improving the overall variance for this analysis?
4. Explain why the difference between the results for samples IA1a and IAab are influenced only by indeterminate errors. Is the same true for the difference between samples IA1 and IA2? How about for samples IA and IB, or samples I and II?
5. The data from your four-level nested design also can be used to establish a confidence interval for the overall analysis. The best experimental estimate of the % w/w erythrosin B is the average result for your eight Level II samples (why is this?). Using the total variance for your analysis, determine the 95% confidence interval for the % w/w erythrosin B.
6. When sampling is the weakest link the sampling plan should be designed to minimize its contribution to the overall variance. One approach is to find a way to increase the gross sample's homogeneity before collecting individual samples. For example, you could grind the gross sample to decrease the average particle size. Another approach is to make a composite sample by collecting several portions of the gross sample and combining them before they are analyzed. If the sampling variance for a particular sample size is σ_{samp}^2 , and σ_{meth}^2 is the variance due to the analysis, then the total variance for the analysis of one sample is

$$\sigma_{\text{total}}^2 = \sigma_{\text{samp}}^2 + \sigma_{\text{meth}}^2$$

If you collect and analyze n separate samples (of the same size), then the total variance is

$$\sigma_{\text{total}}^2 = \frac{\sigma_{\text{samp}}^2}{n} + \frac{\sigma_{\text{meth}}^2}{n}$$

If, however, you form a composite sample by mixing together these n samples and removing k samples (of the same size) for analysis, then the total variance is

$$\sigma_{\text{total}}^2 = \frac{\sigma_{\text{samp}}^2}{nk} + \frac{\sigma_{\text{meth}}^2}{n}$$

Using your results for σ_{samp}^2 and σ_{meth}^2 (which is $\sigma_{\text{total}}^2 - \sigma_{\text{samp}}^2$), determine the total variance for (a) the analysis of four separate samples taken from the gross sample, and (b) the analysis of two portions of a composite sample formed by collecting and mixing four samples from the gross sample. Finally, find a combination of n and k that will give you a total variance that is less than that in (b), but that requires the analysis of only sample.

Acknowledgments

This experiment is adapted from an experiment developed by Frank Settle and Michael Pleva at Washington and Lee University. Should you wish to consult their paper, the reference is *Anal. Chem.* **1999**, *71*, 538A-540A.