STATISTICAL ANALYSIS OF SPECTROPHOTOMETRIC ASSAYS IN THE PRESENCE OF INTERFERENCE

Edward Gbur

Patti Landers

Roy Sharp

Recommended Citation

STATISTICAL ANALYSIS OF SPECTROPHOTOMETRIC ASSAYS
IN THE PRESENCE OF INTERFERENCE

Edward Gbur, Patti Landers and Roy Sharp

University of Arkansas

Abstract

Indirect measurement of the amount of a specified component in a sample of a chemical compound can be accomplished by spectrophotometry. The underlying principle is Beer's Law, which states that, in a pure system, the amount of light absorbed by a chemical bond is linearly related to its concentration. In some mixtures it may not be possible to find a wavelength at which only the bond of interest absorbs light. Hence, the absorbance is composed of contributions from the bond of interest and one or more other (nuisance) bonds. Chemists refer to this situation as interference. In this paper we study the effect of ignoring interference and develop a statistical version of a deterministic approach used by chemists to account for interference. The results are discussed in the context of the measurement of two starch polymers in rice.

Key words: Calibration; Inverse prediction; Regression.

1. Introduction

Starch is the major component of a rice kernel. It has been estimated that approximately 90% of the dry weight of milled rice is starch; the remaining 10% being composed primarily of protein, fat and lipids. Rice starch exists as two polymers, amylose and amylopectin. Amylopectin is the major component, contributing between 55 and 90% of the overall dry weight. Amylose content is considered to be the single most important predictor of the cooking and processing behavior of rice varieties. Varieties which have very little amylose are referred to as "waxy" types because of their cohesiveness or stickiness when cooked.

The most widely used method for amylose determination has been a colorimetric assay in which iodine binds with amylose to produce a blue color. This color is measured spectrophotometrically at one wavelength (Halick and Keneaster, 1956; Juliano, 1971). Unfortunately, the amylopectin present in the rice flour also color reacts with the iodine and interferes with the direct measurement of the color produced by the amylose-iodine complex. Since there is no wavelength at which only one of the starch polymers absorbs light, the contributions of the two polymers cannot be separated by measuring absorbance at only one wavelength. Even though the interference problem is well-known, rice researchers have continued to ignore it and act as if the absorbance value is due only to amylose.

Under the assumption that the two polymers contribute additively to the total absorbance at each wavelength, it is still possible to determine the concentrations of the two components by measuring absorbance at two wavelengths for each sample and simultaneously estimating the concentration of each component (Williard, Merritt and Bean, 1965). This method is well

known in analytical chemistry but has been applied only recently to
research on starch in potatoes (Hovenkamp-Hermelink et al., 1988).

In this paper we describe the results of an experiment to quantify the
effect of ignoring the amylopectin interference when measuring the amylose
concentration. The simultaneous approach is described and stochastic
estimators of the concentrations are obtained along with their standard
errors. Concentration estimates using these estimators are compared to the
estimates derived when the interference is ignored.

2. Beer's Law and Spectrophotometry

Beer's Law forms the basis for the estimation of the concentration of a
substance in a sample using spectrophotometry. It utilizes the principle
that chemical bonds absorb light at specific wavelengths and not at others.
Roughly speaking, Beer's Law states that, in a pure system, the amount of
light absorbed by a chemical bond is linearly related to the concentration
of the substance. A spectrophotometer is a device which focuses a beam of
light on the sample and measures the amount of light reflected by the
sample. Absorbance is the negative of the common logarithm of the
proportion of light reflected by the sample (Williard, Merritt and Bean,
1965).

To construct a statistical model, let

y_λ = the absorbance at wavelength λ,

x = the concentration of the substance.

From Beer's Law, a calibration curve or standard curve is given by the
simple linear regression model

$$y_\lambda = \beta_0 + \beta_1 x + e ,   \tag{1}$$

where β_0 and β_1 are unknown parameters, depending on λ, and e is random
error. Data to fit the model (1) are obtained by measuring the absorbance
y_λ for a series of pure samples of the substance of interest having known
concentrations x. From the fitted model, the concentration for an unknown
sample is estimated by

$$\hat{x} = (y_\lambda - \hat{\beta}_0)/\hat{\beta}_1 ,   \tag{2}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least squares estimators of β_0 and β_1, respectively.
The simple estimator (2) can be found in any standard regression textbook,
along with a method of obtaining a confidence interval for x, under the
heading of inverse estimation or prediction. The inverse estimator (2) is
biased and has infinite mean squared error. Williams (1969) has shown that
no unbiased estimator of x has finite variance. Oman (1988) and Hunter and
Lamboy (1981) contain additional discussion.

The applicability of model (1) and the inverse estimator (2) rests on
the assumption that the only chemical bond in the sample absorbing light at
wavelength λ is found in the substance of interest. If another substance
in the mixture also absorbs at that wavelength, then the observed
absorbance overestimates the expected value of y_λ, $E(y_\lambda)$. This situation
is referred to as interference. Since the slope of the fitted model will

be positive, the estimator $\hat{x}$ in (2) tends to overestimate the true concentration x when there is interference.

3. Amylose Measurement Ignoring Interference

An experiment was conducted to study the effect of ignoring the amylopectin interference when estimating amylose concentration. To this end, the model (1) was fitted using pure amylose samples. Absorbance values y_λ at $\lambda = 620$ nm were obtained from a spectrophotometer for 68 samples made from commerically available pure amylose. Seventeen distinct amylose concentrations, x, ranging from 2 to 21 μg/ml with 3 to 7 replications per concentration were used. Weighted least squares with weights proportional to the reciprocal of amylose concentration was used to fit the model (1). The fitted model was

$$\hat{y}_\lambda = 0.00051 + 0.023x , \tag{3}$$

where the estimated standard errors are $s\hat{e}(\hat{\beta}_0) = 0.00152$ and $s\hat{e}(\hat{\beta}_1) = 0.0002$, respectively. From (3), the estimated concentration is given by

$$\hat{x} = (y_\lambda - 0.00051)/0.023 . \tag{4}$$

To test the fitted model, 88 amylose-amylopectin mixtures of known concentrations were made. Amylose concentrations ranged from 0 to 17.5 μg/ml in increments of 2.5 μg/ml. For each of the eight amylose concentrations, a mean amylopectin concentration was selected to reflect what was thought to be a reasonable amylose-amylopectin ratio. Eleven equally spaced amylopectin concentrations in 0.5 μg/ml increments, centered about the mean, were used. There were three replicates for each mixture combination. Absorbance readings were obtained for each sample and the amylose concentration was estimated from (4). Summary statistics are presented in Table 1.

The results in Table 1 clearly show that there is an upward bias in the estimates of the amylose concentration. However, there are two sources for this bias which must be taken into consideration. The first is the bias of the estimator $\hat{x}$ for x and the second is the bias due to the amylopectin interference.

Using a second order Taylor series expansion of $\hat{x} = g(y_\lambda, \hat{\beta}_0, \hat{\beta}_1)$ about x, it can be shown that

$$\hat{x} - x \simeq (y_\lambda - E[y_\lambda])/\beta_1 - (\hat{\beta}_0 - \beta_0)/\beta_1 - x(\hat{\beta}_1 - \beta_1)/\beta_1$$

$$- (y_\lambda - E[y_\lambda])(\hat{\beta}_1 - \beta_1)/\beta_1^2 + (\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)/\beta_1^2$$

$$+ x(\hat{\beta}_1 - \beta_1)^2/\beta_1^2 .$$

Taking expectations yields

$$Bias(\hat{x}) \simeq [x \, var(\hat{\beta}_1) + cov(\hat{\beta}_0, \hat{\beta}_1)]/\beta_1^2$$

$$\simeq (x - \bar{x})var(\hat{\beta}_1)/\beta_1^2 . \tag{5}$$

For the fitted model (3), an estimate of the bias (5) is approximately

$$\hat{Bias}(\hat{x}) \simeq 0.0001x - 0.0004 \ .$$

The approximate bias of $\hat{x}$ is negligible compared to the biases from Table 1, which, on average, range from 7.4 to 10.3. Thus, the effect of the amylopectin interference is the primary cause of the bias and cannot be ignored.

4. Simultaneous Spectrophotometric Determination

The assumption of additivity of the absorbances of several components in a mixture forms the basis for simultaneously determining the component concentrations. More formally, assume that the total absorbance at wavelength λ is the sum of the absorbances which the several components would have individually if the components were in separate solutions under similar conditions and had the same concentrations as in the mixture. To set notation for a two component system, let

$y_{\lambda j}$ = the absorbance at wavelength λ_j , $j = 1, 2$,
x_i = the concentration of the ith component, $i = 1, 2$,
ϵ_{ij} = the molar absorptivity of the ith component at wavelength λ_j .

Then additivity implies that

$$y_{\lambda 1} = \epsilon_{11}x_1 + \epsilon_{21}x_2$$

$$y_{\lambda 2} = \epsilon_{12}x_1 + \epsilon_{22}x_2 \ . \tag{6}$$

Chemists typically estimate the ϵ_{ij} from separate experiments, assume that they are "known", and solve the deterministic system (6) simultaneously for the concentrations x_1 and x_2 . This yields

$$x_1 = (\epsilon_{22}y_{\lambda 1} - \epsilon_{21}y_{\lambda 2})/\Delta$$

$$x_2 = (\epsilon_{11}y_{\lambda 2} - \epsilon_{12}y_{\lambda 1})/\Delta \ , \tag{7}$$

where $\Delta = \epsilon_{11}\epsilon_{22} - \epsilon_{12}\epsilon_{21}$ is the determinant of the system of equations.

In contrast, we shall begin with a system of equations as in (6) in which the absorbance y is random and the left hand side represents the mean absorbance; that is,

$$E[y_{\lambda 1}] = \epsilon_{11}x_1 + \epsilon_{21}x_2$$

$$E[y_{\lambda 2}] = \epsilon_{12}x_1 + \epsilon_{22}x_2 \ .$$

One approach would be to treat the resulting system as a bivariate regression problem. One disadvantage of such an approach is that mixtures of the substances must be made, whereas in the chemist's approach they are not. In addition, in some applications the range of reasonable (x_1, x_2) pairs needed for the regression approach may not be known. The regression approach will not be pursued here. Instead, following the chemist's approach, the ϵ_{ij} will be estimated from independent experiments and mean absorbances will be estimated directly from the pure samples of the material of interest. Then the concentration estimators are given by

$$\hat{x}_1 = (\hat{\epsilon}_{22}\hat{E}[y_{\lambda 1}] - \hat{\epsilon}_{21}\hat{E}[y_{\lambda 2}])/\hat{\Delta}$$

$$\hat{x}_2 = (\hat{\epsilon}_{11}\hat{E}[y_{\lambda 2}] - \hat{\epsilon}_{12}\hat{E}[y_{\lambda 1}])/\hat{\Delta} \ , \tag{8}$$

where $\hat{\Delta} = \hat{\epsilon}_{11}\hat{\epsilon}_{22} - \hat{\epsilon}_{12}\hat{\epsilon}_{21}$ and the $\hat{E}[y_{\lambda i}]$ are the sample means of the absorbances for the replicate samples of the unknown mixture. The estimated standard errors of the concentration estimators $\hat{x}_1$ and $\hat{x}_2$ in (8) can be obtained by a straightforward application of the Multivariate Delta Theorem (e.g., Bishop, Fienberg, and Holland, 1976).
Define

$$\eta = [\ x_1, \ x_2 \ ]' \ ,$$

$$\theta = [\ \epsilon_{11}, \ \epsilon_{12}, \ \epsilon_{21}, \ \epsilon_{22}, \ E(y_{\lambda 1}), \ E(y_{\lambda 2}) \ ]' \ .$$

Then $\eta = g(\theta) = [\ g_1(\theta), \ g_2(\theta) \ ]'$ is given by (7) with $E(y_{\lambda j})$ replacing $y_{\lambda j}$. Let $\hat{\theta}$ be a consistent estimator of θ and let $\hat{\eta} = g(\hat{\theta})$. Then the mean of $\hat{\eta}$ is approximately η and the covariance matrix of $\hat{\eta}$ is approximately

$$cov(\hat{\eta}) = G \ cov(\hat{\theta}) \ G' \ , \tag{9}$$

where G is the 2×6 matrix of partial derivatives of $g_i(\hat{\theta})$ with respect to $\hat{\theta}_j$ and $cov(\hat{\theta})$ is block diagonal since the ϵ_{ij} are estimated in pairs from separate experiments.

5. Amylose Measurement Accounting for Interference

The simultaneous determination method described above was applied to the rice starch problem. Since amylose and amylopectin are polymers, it was not possible to determine the molar absorptivities ϵ_{ij}. Instead, absorptivity values a_{ij} were estimated and expressed as unit absorptivity (μg/ml). Wavelengths of 620 and 560 nm were chosen corresponding to the absorption spectra maxima for amylose and amylopectin, respectivity. The absorptivity values for amylose were estimated from an experiment using 30 pure amylose samples. For each sample, an absorptivity value was calculated at each of the two wavelengths. The means of these sets of values over the 30 samples were used as the absorptivity estimates $\hat{a}_{1j}$. The corresponding experiment for amylopectin used 28 samples. The absorptivity estimates, their estimated standard errors, and estimated correlations from the two experiments are presented in Table 2. The values in Table 2 yield the first four entries of $\hat{\theta}$ and the first two blocks of $cov(\hat{\theta})$.

Substituting the estimates in Table 2 into (8), the estimated concentrations for an unknown sample are obtained from

$$\hat{x}_1 = (0.0072\hat{E}[y_{\lambda 1}] - 0.0056\hat{E}[y_{\lambda 2}])/0.000072$$

$$\hat{x}_2 = (0.0233\hat{E}[y_{\lambda 2}] - 0.0171\hat{E}[y_{\lambda 1}])/0.000072 \ . \tag{10}$$

The estimated matrix of partial derivatives is given by

$$\hat{G} = \begin{bmatrix} -100.0\hat{x}_1 & 77.8\hat{x}_1 & 100.0\hat{x}_2 & 77.8\hat{x}_2 & 100.0 & -77.8 \\ 326.6\hat{x}_2 & 237.5\hat{x}_2 & -237.5\hat{x}_1 & 323.6\hat{x}_1 & 237.5 & -323.6 \end{bmatrix} ,$$

where, for example, the (1,1) entry of $\hat{G}$ is obtained by substituting estimates into

$$\frac{\partial x_1}{\partial \epsilon_{11}} = -\frac{(\epsilon_{22}E[y_{\lambda 1}] - \epsilon_{21}E[y_{\lambda 2}])\epsilon_{22}}{\Delta^2}$$

$$= - x_1\epsilon_{22}/\Delta .$$

The estimated covariance matrix of $\hat{\theta}$ is block diagonal with three 2×2 blocks given by

$$\begin{bmatrix} 1.0 \times 10^{-6} & 7.1 \times 10^{-13} \\ 7.1 \times 10^{-13} & 8.1 \times 10^{-7} \end{bmatrix} , \quad \begin{bmatrix} 4.0 \times 10^{-8} & 1.5 \times 10^{-15} \\ 1.5 \times 10^{-15} & 4.0 \times 10^{-7} \end{bmatrix}$$

and

$$\begin{bmatrix} \hat{var}(\hat{E}[y_{\lambda 1}]) & \hat{cov}(\hat{E}[y_{\lambda 1}], \hat{E}[y_{\lambda 2}]) \\ \hat{cov}(\hat{E}[y_{\lambda 1}], \hat{E}[y_{\lambda 2}]) & \hat{var}(\hat{E}[y_{\lambda 2}]) \end{bmatrix} .$$

The simultaneous model was tested using the 88 amylose-amylopectin mixtures described in Section 3. The sample means of the absorbances at 620 and 560 nm from the three replications for each mixture combination were used in (10) to obtained estimated concentrations. In turn, these estimates were used to calculate $\hat{G}$. The third block of $\hat{cov}(\hat{\theta})$ was estimated using the sample variances and covariance from the three replicates of each mixture combination. Finally, $\hat{cov}(\hat{\eta})$ was calculated from (9).

To illustrate the calculations, from the three replications for mixture #48, we have

$$\hat{E}[y_{620}] = 0.444, \qquad \hat{se}(\hat{E}[y_{620}]) = 0.010,$$

$$\hat{E}[y_{560}] = 0.432, \qquad \hat{se}(\hat{E}[y_{560}]) = 0.005,$$

and $\quad \hat{\rho} = 0.803 .$

The actual and estimated concentrations for both starchs are given in Table 3. The estimated correlation of these concentration estimates was -0.89.

For each of the 8 amylose concentrations, summary statistics were calculated from the set of 11 mixture estimates at that concentration and are presented in Table 4. Comparing the results in Table 4 with the corresponding estimates in Table 1 shows that accounting for the amylopectin interference has greatly reduced the bias in the estimated amylose concentration. Although four of the eight mean estimated concentrations are within two estimated standard errors of the true concentration, there is still a tendency to overestimate the true amylose concentration, especially at the lower concentrations. Possible reasons for this include the failure of Beer's Law itself to hold, the failure to

satisfy the underlying assumptions of Beer's Law, and difficulties in conducting the colorimetric assay to obtain absorbance readings.

The estimated variances in Table 4 are relatively large since they are based on only the three replications for each mixture combination. If the concentrations are to be estimated for an unknown mixture, more replications would be required to adequately estimate the entries of the third block of $\text{cov}(\hat{\theta})$.

6. Conclusion

Using the rice starch example, we have demonstrated empirically that interference in spectrophotometric assays causes a serious upward bias in the concentration estimates for the substance of interest. The magnitude of the bias depends on the concentration of the nuisance bonds in the mixture. The simultaneous estimation procedure used by chemists was shown to greatly reduce the bias. Using the stochastic version of the simultaneous procedure, standard error estimates for the estimated concentrations can be calculated when replications are available.

The above conclusions relate to a specific problem. There are, however, broader implications. For decades agricultural scientists have utilized statistical design and analysis concepts for "field type" experiments in which uncontrolled variation is large and clearly visible. More recently, some agricultural experimentation has moved to laboratory settings in which data is collected using highly sophisticated equipment. There is often an appearance of a tightly controlled environment. The magnitude of uncontrolled variation is clearly much smaller than that found in the field and may even appear to be nonexistent. Under these conditions some researchers have tended to ignore the statistical principles that they have used almost religiously in field experimentation. This will require re-education and, in some cases, the application of statistical techniques which are unfamiliar to the agricultural scientist.

References

Bishop, Yvonne M.M., Stephen E. Fienberg and Paul W. Holland (1976). Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: The MIT Press.

Halick, J.W. and K.K. Keneaster (1956). The use of a starch-iodine-blue test as a quality indicator of white raw milled rice. Cereal Chemistry 33, 315-319.

Hovenkamp-Hermelink, J.H.M., J.N. de Vries, P. Adamse, E. Jacobsen, B. Witholt and W.J. Feenstra (1988). Rapid estimation of the amylose/amylopectin ratio in small amounts of tuber and leaf tissue of the potato. Potato Research 31, 241-246.

Hunter, William G. and Warren F. Lamboy (1981). A Bayesian analysis of the linear calibration problem (with discussion). Technometrics 23, 323-350.

Juliano, B.O. (1971). A simplified assay for milled-rice amylose. Cereal Science Today 16, 334.

Oman, Samuel D. (1988). Confidence regions in multivariate calibration. *Annals of Statistics* 16, 174-187.

Williams, E. J. (1969). A note on regression methods in calibration. *Technometrics* 11, 189-192.

Williard, H.H., L.L. Merritt and J.A. Bean (1965). *Instrumental Methods of Analysis*. New York: Van Nostrand Reinhold.

Table 1. Summary statistics for the estimation of amylose concentration ignoring the amylopectin interference.

True amylose concentration	Mean estimated amylose concentration	Mean relative bias (Percent)	Mean true amylopectin concentration
0.0*	10.3(0.1)*	-	42.5*
2.5	12.6(0.1)	403.0	40.0
5.0	14.7(0.1)	193.4	37.5
7.5	16.9(0.2)	124.7	35.0
10.0	19.1(0.1)	90.7	32.5
12.5	21.0(0.2)	68.1	30.0
15.0	22.8(0.1)	52.2	27.5
17.5	24.9(0.2)	42.0	25.0

* Concentrations are in μg/ml. Estimated standard errors based on 33 observations are given in parantheses.

Table 2. Estimated absorptivity values $\hat{a}_{ij}$.

λ	Amylose ($\hat{a}_{1j}$)	Amylopectin ($\hat{a}_{2j}$)
620 nm	0.0233 (0.0010)	0.0056 (0.0002)
560 nm	0.0171 (0.0009)	0.0072 (0.0002)
	$\hat{\rho} = 0.88$	$\hat{\rho} = 0.96$

Table 3. Estimated starch concentrations for mixture #48.

	Amylose	Amylopectin
True concentration	10.00*	33.50
Estimated concentration	10.95	33.88
Estimated standard error	0.73	1.46
Relative bias	9.50%	1.13%

* Table entries are in μg/ml unless otherwise indicated.

Table 4. Summary statistics for the estimation of amylose
concentration accounting for the amylopectin interference.

True amylose concentration	Mean estimated amylose concentration	Variance of mean estimated concentration	Mean relative bias (Percent)
0.0*	0.2(0.04)*	0.02(0.01)	-
2.5	2.9(0.03)	0.02(0.01)	15.8
5.0	5.7(0.05)	0.04(0.01)	14.1
7.5	8.3(0.06)	0.10(0.03)	10.2
10.0	10.8(0.07)	0.11(0.04)	8.3
12.5	13.2(0.09)	0.15(0.06)	5.9
15.0	15.7(0.10)	0.23(0.07)	4.4
17.5	18.3(0.09)	0.17(0.06)	4.4

*All concentrations are in μg/ml. Estimated standard errors based
11 sets of estimates are given in parantheses.