SOME STATISTICAL ASPECTS OF THE CALIBRATION PROBLEM

by

DENNIS L. CLASON

B.S., California State College, Bakersfield, 1976

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1980

Approved by:

James L. Hess
Major Professor

Acknowledgements

I would like to thank all those who have made this possible, my parents, teachers, and friends. Special credit is due to my major professor, Dr. James Hess, and the members of my committee, Dr. George Milliken and Dr. Kenneth Kemp. I would also like to thank the graduate students in the Statistics Department, for helping provide an environment conducive to growth. Finally, I would like to thank my wife, Kim, for being understanding and supportive as I finished this.

TABLE OF CONTENTS

I. INTRODUCTION

Consider the situation in which we have two variables, X and Y, whose relationship is known to be linear over a certain range. Suppose further than Y is something relatively easy to measure, while X can only be measured with some difficulty. Equivalently, we could assume that the measurement of Y is nondestructive and measurement of X is destructive. Assuming that X represents the quantity which is of interest, the problem is to predict the value of X which generated some observed value Y. This problem is referred to as "inverse prediction" or the "calibration problem" in the literature. Before examining the statistical problem, let us consider two specific situations in which the problem arises.

In biology and chemistry many compounds of interest to the experimenter absorb specific light wavelengths strongly. By using a spectrophotometer, the amount of absorption can be determined quite easily. An effect referred to as Beer's law which relates the concentration of solute to the transmittance can be easily derived from physical considerations. Beer's law states that

$$T = 10^{-\varepsilon c l} \tag{I.1}$$

where T = fractional transmittance

ε = molar extinction coefficient

c = molar concentration of solute

l = length of light path in cm

Rather than working with transmittance, researchers prefer to use the optical density Y, defined as

$$Y = -\log_{10} T.$$

In terms of optical density, Beer's law becomes

(I.2) $$Y = \beta c$$

if we let $\beta = \varepsilon l$. We now have a linear model in the variable of interest, that is solute concentration, which places us in the framework described above.

As a second example, suppose we have a spring, a ruler, and a set of known masses. According to Hook's law

$$Y = km$$

(I.3) where Y = displacement of spring

m = weight of i^{th} mass

k = spring constant

Again we have a linear model in the weight of the object, and given an object of unknown mass, we want to "predict" the mass on the basis of the spring displacement.

II. POINT ESTIMATION OF X

There have been four point estimators proposed in the literature. All four estimators assume a fixed X type model with $E[Y] = \alpha + \beta X$. If α and β were known it would be reasonable to use

(II.1) $$\overset{\circ}{X} = (Y-\alpha)/\beta \ .$$

If we assume a normal error distribution it is easy to show that $\overset{\circ}{X}$ is the maximum likelihood estimator (MLE) in this case.

Retaining the normality assumption, but assuming the parameters are to be estimated by an experiment, we find the MLE is

(II.2) $$\hat{X} = (Y-\hat{\alpha})/\hat{\beta} \ ,$$

where $\hat{\alpha}$ and $\hat{\beta}$ are the ML estimates of α and β. $\hat{X}$ is then a MLE by the invariance property of MLE's. $\hat{X}$ is referred to as the "Classical Estimator." It is consistent, in the sense that under fairly broad conditions $X \xrightarrow{L} N(X,\sigma^2/\beta^2)$, which is the distribution of $\overset{\circ}{X}$. We state the conditions and results more precisely as a theorem.

Theorem: Suppose that a sequence of standards (X_j,Y_j) $j=1,\ldots,n$ are used to establish a sequence of calibration lines $\hat{Y}_n = \hat{\alpha}_n + \hat{\beta}_n X$, further, assume $Y_j \overset{ind}{\sim} N(\alpha + \beta X_j,\sigma^2)$, $j=1,\ldots,n$. Let V_i $i=1,\ldots,m$ be a set of unknowns whose corresponding X's are to be estimated. Assume that $V_i \overset{ind}{\sim} N(\alpha + \beta X_i,\sigma^2)$, the Y_j and V_i are independent and that the sequence of standards satisfies the regularity condition $\lim_{n\to\infty} \sup_j X_j^2/\Sigma_i(X_i-\bar{X})^2 = 0$. Then the following hold

1) $\hat{\beta}_n \xrightarrow{P} \beta$

2) $\hat{\alpha}_n \xrightarrow{P} \alpha$

3) $S_n^2 \xrightarrow{P} \sigma^2$

4) $\hat{X}_n^k = (Y_k - \hat{\alpha}_n)/\hat{\beta}_n \xrightarrow{L} N(X_k, \sigma^2/\beta^2)$

Before proving the theorem we will examine the assumptions for reasonableness. The bulk of the assumptions are those of the simple linear regression model with normal error. If this model is deemed unreasonable we should probably discard all Normal Theory as being equally unreasonable. This leaves the assumptions of independence between the standards and the unknowns, and the regularity condition. We feel that the independence assumption is reasonable since the standards are generally prepared in the laboratory while the observations stem from some population or process being observed. The regularity condition requires that we not observe Y at any single value of X_i infinitely often, which does not seem too restrictive. We will not prove the theorem.

Proof: 1) From linear model theory we know that

$$\hat{\beta}n \sim N(\beta, \sigma^2/\Sigma_i(X_i - \bar{X})^2),$$

hence $E(\hat{\beta}_n - \beta)^2 = VAR(\hat{\beta}_n) = \sigma^2/\Sigma_i(X_i - \bar{X})^2$. Now, $\text{limit}_{n\to\infty}\sigma^2/\Sigma_i(X_i-\bar{X})^2 = \text{limit}_{n\to\infty}\dfrac{\sigma^2}{\sup X_j^2} \cdot \dfrac{\sup X_j^2}{\Sigma_i(X_i-\bar{X})^2} = 0$, if $\sup X_j^2 = 0$, which is reasonable, since if $\sup_j X_j^2 = 0$, α and β are not estimable. Thus $\hat{\beta}_n$ is converges in mean square, and hence it also converges in probability.

2) Again, from linear model theory

$$E(\hat{\alpha} - \alpha)^2 = (\Sigma X_j^2/n\Sigma_i(X_i - \bar{X})^2\sigma^2.$$

Hence,

$$\text{limit}_{n\to\infty}E(\hat{\alpha}_n-\alpha)^2 \leq \text{limit}_{n\to\infty}(n \sup X_j^2/n\Sigma_i(X_i-\bar{X})^2)\sigma^2$$

$$\leq \sigma^2\text{limit}_{n\to\infty}\sup X_j^2/\Sigma_i(X_i-\bar{X})^2$$

$$= 0.$$

So $\hat{\alpha}$ converges in mean square to α.

3) $S^2 = \frac{1}{n}\Sigma(Y_i-\hat{Y}_i)^2 = \frac{1}{n}\Sigma(Y_i-\hat{\beta}X_i)^2$

(For the sake of clarity we assume the data has been centered.)

$$= \frac{1}{n}\Sigma(\beta X_i+\varepsilon_i-\hat{\beta}X_i)^2$$

$$= \frac{1}{n}\Sigma\varepsilon_i^2 - \frac{2}{n}(\hat{\beta}-\beta)\Sigma X_i\varepsilon_i + \frac{(\hat{\beta}-\beta)^2}{n}\Sigma X_i^2$$

Now by using the fact that $\hat{\beta} = \frac{\Sigma X_i Y_i}{\Sigma X_i^2} = \frac{\Sigma X_i(X_i\beta+\varepsilon_i)}{\Sigma X_i^2}$

it is easy to show that

$$S^2 = \frac{1}{n}\Sigma\varepsilon_i^2 - \frac{(\hat{\beta}-\beta)}{n}\Sigma X_i^2.$$

By Kolmogorov's Strong Law of large numbers $\frac{1}{n}\Sigma\varepsilon_i^2 \xrightarrow{a.s.} \sigma^2$, since $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$. Thus we need only show that $\frac{(\hat{\beta}-\beta)}{n}\Sigma X_i\varepsilon_i \xrightarrow{P} 0$. But

$$\frac{(\hat{\beta}-\beta)}{n}\Sigma X_i^2 = \frac{(\Sigma X_i\varepsilon_i)^2}{n(\Sigma X_i^2)^2}\Sigma X_i^2 .$$

Now $X^2 \xrightarrow{P} 0$ if and only if $X \xrightarrow{P} 0$, so we will consider the square root of this quantity

$\frac{\Sigma X_i \varepsilon_i}{\sqrt{n}\sqrt{\Sigma X_i^2}} \sim N(0,\sigma^2/n)$ so by Chebyshev's inequality $\frac{(\hat{\beta}-\beta)}{n}\Sigma X_i^2 \xrightarrow{P} 0$,

and hence $S^2 \xrightarrow{P} \sigma^2$.

4) Finally we want to show that $\hat{X}_n^k \xrightarrow{L} N(X_k,\sigma^2/\beta^2)$. Now since $\hat{\alpha} \xrightarrow{L} \alpha$ and $\hat{\beta} \xrightarrow{L} \beta$ we can replace $\hat{\alpha}$ and $\hat{\beta}$ by α and β without changing the limiting distribution of $\hat{X}$, hence the limiting distribution of $\frac{V_k-\hat{\alpha}}{\hat{\beta}}$ is the same as that of $(V_k-\alpha)/\beta = \overset{\circ}{X}$, and $\overset{\circ}{X} \sim N(X,\sigma^2/\beta^2)$.

ged.

As a corollary to this theorem, if we have a sequence of m independent observations of V at a single value of X, and the conditions of the theorem hold, $\hat{X}_n^k \xrightarrow{L} N(X_k,\sigma^2/m\beta^2)$ and $\hat{X}_n^k \xrightarrow{P} X_k$ where $\bar{X}_n^k = (\bar{V}_m-\hat{\alpha}_n)/\hat{\beta}_n$. This follows because the sequence $\bar{V}_m \to \alpha+\beta X_k$ almost surely, by Kolmogorov's Strong Law of Large Numbers.

This theorem can be used to obtain large sample confidence intervals of X given an observation of Y. This will be discussed further in the chapter on interval estimation.

Eisenhart (1939) noted that a regression could be found for X or Y, then this would yield a direct estimate of X from Y. He discarded this estimator because the partition of the Total Sum of Squares in the ANOVA table failed to make sense. Krutchkoff (1967) pointed out that the linear model

(II.3) $\quad Y_i = \alpha + \beta X_i + \varepsilon_i \qquad \varepsilon \overset{iid}{\sim} N(0, \ ^2)$

could be reparameterized as

(II.4) $\quad X_i = \gamma + \delta Y_i + \eta_i \qquad \eta_i \overset{iid}{\sim} N(0,\sigma^2/\beta^2)$

where $\gamma = -\frac{\alpha}{\beta}$, $\delta = 1/\beta$, $\eta_i = -\frac{\varepsilon_i}{\beta}$. His proposal was to estimate the parameters of this "inverse" regression model, then use the "inverse estimator"

$$\text{(II.5)} \qquad \check{X} = \hat{\gamma} + \hat{\delta}Y,$$

where $\hat{\gamma}$ and $\hat{\delta}$ are the least squares estimates of γ and δ respectively.

Krutchkoff (1967) then proceeded to do a mammoth simulation study comparing $\check{X}$ and $\hat{X}$. Using the estimated Mean Squared Error (MSE) as his goodness criterion Krutchkoff claimed

> "Method B ($\check{X}$) has a smaller MSE than method A ($\hat{X}$) <u>for all X</u> (author's emphasis). This being the case, any criterion based on the MSE with favor the Inverse Method B over the Classical approach, Method A." (Krutchkoff 1967)

Unfortunately, his simulation study fails to back his claim. Using design parameters of $\alpha = 0$, $\beta = .5$, $\sigma^2 = .1$, 10 observations at $X = 0$, 10 observations at $X = 1$, predicting the value $X = 2$ Krutchkoff estimated the MSE of $\check{X}$ to be .077, and the estimated MSE of $\hat{X}$ was .064. The estimated standard error was .001 for both estimators. So clearly $\check{X}$ does not have a uniformly smaller MSE than $\hat{X}$.

It can be shown that $\check{X}$ has finite variance for $n > 6$. If we write the classical estimator as $\hat{X} = \bar{X} + (Y - \bar{Y})/\hat{\beta}$, clearly then $VAR(\hat{X}) = VAR[(Y - \bar{Y})/\hat{\beta}]$. But $Y - \bar{Y}$ and $\hat{\beta}$ are independent normal random variables, hence their ratio has a Cauchy distribution, and the variance and hence the MSE of $\hat{X}$ must be infinite for any finite n.

So as we have seen, the Inverse estimator has a finite variance and hence finite MSE, while the Classical estimator has infinite MSE. In fact, it has been shown by Williams (1969) that a UMVU (Uniform Minimum Variance Unbiased) estimator exists based on the sufficient statistics Y, $\bar{Y}$, and $\hat{\beta}$, and is

(II.6) $$\tilde{X}_{mv} = \bar{X} + (Y - \bar{Y})g(\hat{\beta})$$

$$\text{where } g(\hat{\beta}) = \begin{cases} \frac{1}{\phi^2}\,[\exp(\hat{\beta}^2/2\phi^2)][1-\Phi(\hat{\beta}/\phi)] & \text{if } \beta > 0 \\ -\frac{1}{\phi^2}\,[\exp(\beta^2/2\phi^2)]\Phi(\hat{\beta}/\phi) & \text{if } \beta < 0 \end{cases}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal distribution, and $\phi = \sigma^2/\Sigma_i(X_i - \bar{X})^2$.

But the variance of $\tilde{X}_{mv}$ is also infinite. Thus using the MSE as the criterion we would find the inverse estimator preferable to either the UMVUE or the Classical estimator. But at the same time we would also find any arbitrary constant, say 0, to be a better estimate than $\hat{X}$. Williams (1969a) says that this line of argument shows that the MSE is not an appropriate criterion in this type of problem.

If the MSE is not an appropriate criterion, the question is then what is appropriate? Brown (1979) suggests that Integrated Mean Square Error (IMSE) is more appropriate. We agree, since the IMSE is in a sense averaging the MSE over all relevant values of X. That is, the IMSE of an estimator is defined to be

$$\text{IMSE } (\tilde{X}) = \int_A E\{(\tilde{X} - x)^2\}\, dW(x)$$

where the expectation is with respect to Y at $X = x$, and $W(x)$ is a weighting function. Without loss of generality we may assume that $W(\cdot)$ is a distribution function over A, where A is the region we wish to calibrate.

Given this new criterion we can develop a linear predictor $\tilde{X} = \lambda_0 + \lambda y$, which minimizes the IMSE. The MSE of $\tilde{X}$ is defined to be

$$(II.7) \qquad E(\tilde{X} - X)^2 = E(\lambda_0 + \lambda Y - X)^2$$

$$= \lambda\sigma^2 + (\lambda_0 + \lambda\alpha)^2 + 2(\lambda_0 + \lambda\alpha)(\lambda\beta - 1)X + (\lambda\beta - 1)^2X^2.$$

Thus, the IMSE of $\tilde{X}$ is

$$(II.8) \qquad IMSE(\tilde{X}) = \int_A E(\tilde{X} - X)^2 \, d\, W(x)$$

$$= \lambda^2\sigma^2 + (\lambda_0 + \lambda\alpha)^2 + 2(\lambda_0 + \lambda\alpha)(\lambda\beta - 1)w_1 + (\lambda\beta - 1)^2 w_2$$

where $w_j = \int_A x^j dW(x)$, $j=0,1,2$.

Now if we take the partial derivatives of the IMSE with respect to λ_0 and λ we find

$$(II.9) \qquad \frac{\partial IMSE(\tilde{X})}{\partial \lambda_0} = 2\lambda_0 + 2\lambda(\alpha + \beta w_1) - w_1$$

$$\frac{\partial IMSE(\tilde{X})}{\partial \lambda_0} = (\sigma^2 + \alpha^2 + w_1\alpha\beta + \beta^2 w_2)\lambda$$

$$+ 2(\alpha + w_1\beta)\lambda_0 - 2(\alpha w_1 + \beta w_2).$$

Setting both partial derivatives equal to zero and solving the resulting linear system of equations we find that

$$(II.10) \qquad \hat{\lambda}_0 = w_1 - \lambda(\alpha + w_1\beta)$$

$$\hat{\lambda} = \beta(w_2 - w_1^2)/[\sigma^2 + (w_2 - w_1^2)\beta^2]$$

If we let $M = w_1$, the mean of the weighting distribution, and $V = w_2 - w_1^2$, the variance of the weighting distribution we can write

$$(II.11) \qquad \hat{\lambda}_0 = M - \lambda(\alpha + M\beta)$$

$$\hat{\lambda} = V\beta/(\sigma^2 + V\beta^2).$$

Note that $\hat{\lambda}_0$ and $\hat{\lambda}$ are not statistics, since they still depend upon the unknown parameters α, β and σ^2. Brown suggests that the Least Squares estimates of α, β, σ^2 be used in place of the parameters. The IMSE optimal estimator is

$$(II.12) \qquad \tilde{X}_0 = \hat{\lambda}_0 + \hat{\lambda}y = \{M\hat{\sigma}^2 + \hat{\beta}V(y - \hat{\alpha})\} / (\hat{\sigma}^2 + V\hat{\beta}^2).$$

An interesting result occurs when the first two moments of $W(x)$ are equal to the first two moments of the fixed X's, that is, when $M = \bar{X}$ and $V = n^{-1}\Sigma(X_i - \bar{X})^2$. Then, since the LS estimates are

$$(II.13) \qquad \hat{\beta} = n^{-1}\Sigma(X_i - \bar{X})(Y_i - \bar{X})/n^{-1}\Sigma(X_i - \bar{X})^2 = S_{xy}/S_{xx} \qquad \text{say}$$

$$\hat{\sigma}^2 = n^{-1}\Sigma(Y_i - \bar{Y})^2 - \hat{\beta}^2 n^{-1}\Sigma(X_i - \bar{X})^2 = S_{yy} - S_{xy}^2 .$$

From (II.12) $\tilde{X}_0 = \{M\hat{\sigma}^2 + \hat{\beta}V(Y - \hat{\alpha})\}/(\sigma^2 + V\hat{\beta}^2)$

but $V = {}_{xx}$, $\hat{\beta} = S_{xy}/S_{xx}$, and $\hat{\alpha} = Y - \hat{\beta}\bar{X}$ so that

$$(II.14) \qquad \tilde{X} = \frac{\bar{X}(S_{yy} - S_{xy}^2/S_{xx}) + (S_{xy}/S_{xx})(S_{xx})Y - \bar{Y} + (S_{xy}/S_{xx})\bar{X}}{S_{yy} - S_{xy}^2/S_{xx} + S_{xx}(S_{xy}^2/S_{xx}^2)}$$

$$= \frac{\bar{X}\, X_{yy} - \bar{X}\, S_{xy}^2/S_{xx} + S_{xy}\,(Y - \bar{Y}) + \bar{X}\, S_{xy}^2/S_{xy}}{S_{yy}}$$

$$= \bar{X} - (S_{xy}/S_{yy})(Y - \bar{Y}) = \check{X}.$$

So, if we are willing to regard the sample distribution of the X's as a "prior distribution" of the values we expect X to take on, then the inverse estimator is the optimal linear estimator of X, in terms of IMSE.

Brown then presents some limiting results, in terms of $Q = \beta^2/\sigma^2$ and V, for the optimal ($\tilde{X}_0$), classical ($\hat{X}$) and inverse ($\check{X}$) estimators. He first

considers the case where $Q \to \infty$, and all estimators have an IMSE = 0. He then considers the case of $Q \to 0$, that is, either $\beta^2 \to 0$ or $\sigma^2 \to \infty$. In either case, Q represents in some sense the signal to noise ratio, and one must question the utility of any technique when the signal, that is the calibration line, is buried by the noise level. For this reason we consider his limiting results of $IMSE(\hat{X}) = \infty$, $IMSE(\tilde{X}_0) = V$, and $IMSE(\check{X}) = V + (\bar{X} - M)^2$ to be of limited significance at best. As $V \to \infty$, $IMSE(\hat{X}) = Q^{-1}$, $IMSE(\tilde{X}_0) = Q^{-1}$ and $IMSE(\check{X}) = \infty$. This result bears out Berkson's (1969) conclusion that although $\check{X}$ may "beat" $\hat{X}$ over any finite interval, as the interval to be calibrated grows larger, $\hat{X}$ tends to do better over the whole range of values than X does. Finally, when $\check{X} = \tilde{X}_0$ he notes that $\text{limit}_{V \to S_{xx}, M \to \bar{X}} IMSE(\tilde{X}_0) = V/(1+VQ) < \text{limit}_{V \to S_{xx}, M \to \bar{X}} IMSE(\hat{X}) = 1/Q$.

The final point estimator we will consider was developed independently by Winslow (1976) and Naszodi (1978). Winslow developed the estimator under the assumption that the distribution of $\hat{\beta}$ is a truncated normal distribution. We will follow Naszodi's development, which makes milder assumptions about the distribution of $\hat{\alpha}$ and $\hat{\beta}$.

The derivation begins by expanding the classical estimator $\hat{X}$ in a Taylor series α, β and y. Assuming that the error distribution is symmetric it can be shown that

$$(II.15) \qquad E\{\hat{X}\} = X + (X - \bar{X})\sigma^2/\{\beta^2 \Sigma (X_i - \bar{X})^2\} + R$$

where R is a remainder term. Naszodi shows that this remainder is infinite in the case of a finite sample from a normal distribution, however, as seen in the section on the classical estimator the estimate converges and the bias tends to zero as n tends to infinity. Naszodi then avoids the convergence problem by truncating the error distribution. Then we can estimate

the bias in $\hat{X}$ and remove it to obtain a "practically unbiased estimator." We define the estimator as

(II.16) $$X^{+} = \hat{X} - \hat{B}(\hat{X})$$

where $\hat{B}(\hat{X})$ estimates the bias in $\hat{X}$, that is $\hat{B}(\hat{X})$

$$= (\hat{X} - \bar{X})\hat{\sigma}^2/(\hat{\beta}^2(X_i - \bar{X})^2).$$

Upon substitution and simplification we obtain the result

(II.17) $$X^{+} = \bar{X} + (Y - \bar{Y})/[\hat{\beta} + \hat{\sigma}^2/(\hat{\beta}^2 + \Sigma(X_i - \bar{X})^2].$$

If we rewrite (II.17) so that the regression function is in terms of Y we find that

(II.18) $$Y^{+} = \{\hat{\beta} + \sigma^2/(\hat{\beta}^2\Sigma(X_i - \bar{X})^2)\ (X - \bar{X}) + \bar{y}$$

which, like the inverse estimator, is a case of fitting a so-called wrong regression line to the data.

III. CONFIDENCE INTERVALS FOR X

In this chapter we will develop several interval estimates for X. Two of them have confidence levels which hold for each individual interval, the other three have confidence levels which hold simultaneously for all X in some interval. The choice of the type of interval to use lies with the user.

In Chapter II we found a limiting distribution for $\hat{X}$. This leads us to use

$$(III.1) \qquad (\hat{X} - X)/(\sigma^2/\hat{\beta}^2) \xrightarrow{L} N(0,1)$$

as a pivot, to form the $100(1 - \alpha)\%$ asymptotic confidence interval

$$(III.2) \qquad \hat{X} - Z_{\alpha/2}(\hat{\sigma}^2/\hat{\beta}^2)^{1/2} < X < \hat{X} + Z_{\alpha/2}(\hat{\sigma}^2/\hat{\beta}^2)^{1/2}.$$

This is the interval often given in applied statistics texts (e.g., Bennett and Franklin (1963)). The user should be cautious of applying this interval to small samples, since it is an asymptotic result.

Fieller (1954) derived the following result which can be used in some instances to obtain a conservative $100(1 - \alpha)\%$ confidence interval for X. The interval is conservative in that the true confidence coefficient is not less than $1 - \alpha$.

Theorem (Fieller's Theorem)

Let $(a,b)' \sim N_2[\binom{\alpha}{\beta}, \sigma^2\begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix}]$. A conservative $100(1 - \alpha)\%$ confidence interval for a parameter $p = E(a)/E(b)$ estimated by $m = a/b$ is

(III.3) $\{M - k^2v_{12} + k[v_{11} - 2mv + m^2v_{22} - k^2(v_{11}v_{22} - v_{12}^2)]^{1/2}/(1-k^2v_{22})$

where $k = st/b$, s^2 is an estimate of σ^2 independent of $(a,b)'$ based on ν degrees of freedom, and t is the $100(1 - \alpha/2)$ percentile of the t distribution with ν degrees of freedom.

This method will only yield finite intervals when $1 - k^2v_{22} > 0$ and the discriminant $d = (v_{11} - 2mv_{12} + m^2v_{22} - k^2(v_{11}v_{22} - v_{12}^2)) > 0$. This is because the confidence interval yielded by this technique is actually the region of the real line over which a quadratic function is negative. When $1 - k^2v_{22} > 0$ and $d > 0$ the quadratic is negative over a finite interval. In order to insure that we obtain only useful intervals Graybill (1976) recommends the following procedure. Obtain $\hat{X}$, then test $H_o:\beta = 0$ vs $H_a:\beta \neq 0$ with size α. If H_o is rejected, then this method will yield a useful interval. If H_o is not rejected, then no useful interval exists.

To apply Fieller's Theorem we obtain estimates of α and β from a calibration experiment. We know that $(\hat{\alpha},\hat{\beta})'$ is independent of our future observation, Y, and

$$(\hat{\alpha},\hat{\beta})' \sim N_2\left[\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \frac{\sigma^2}{n\Sigma(X_i-\bar{X})^2}\begin{pmatrix}\Sigma X_i^2 & -n\bar{X}\\ -n\bar{X} & n\end{pmatrix}\right].$$

We recall that $Y \sim N(\alpha+\beta X,\sigma^2)$, hence the vector $(Y - \hat{\alpha},\hat{\beta})$ must have a normal distribution, so

$$(Y - \hat{\alpha},\hat{\beta}) \sim N_2\left[\begin{pmatrix}\beta X\\ \beta\end{pmatrix}, \frac{\sigma^2}{n\Sigma(X_i-\bar{X})^2}\begin{pmatrix}1+\Sigma X_i^2 & n\bar{X}\\ n\bar{X} & n\end{pmatrix}\right].$$

In the framework of Fieller's Theorem $p = X$, p is estimated by $\hat{X} = (Y - \hat{\alpha})/\hat{\beta}$ and $v_{11} = 1 + \Sigma X_i^2/n\Sigma(X_i - \bar{X})^2$, $v_{12} = +n\bar{X}$, $v_{22} = 1/\Sigma(X_i - \bar{X})^2$.

Scheffé (1973) showed that for quite general linear models $y = f(x) + \epsilon$ it is possible to obtain simultaneous $100(1 - \alpha)\%$ confidence intervals

for X given an observation of Y. His method is based on two assumptions

(III.4) $\underline{Y} = [\underline{G(x)}]\underline{\beta} + \underline{\varepsilon}$

where $\underline{G(X)} = \begin{bmatrix} g_1(X_1) & \cdots & g_p(X_1) \\ \vdots & & \\ g_1(X_n) & \cdots & g_p(X_n) \end{bmatrix}$

$g_i(\cdot)$ are known functions with continuous derivatives over the calibration interval. $\underline{\varepsilon} \sim N(0,\sigma^2 I)$, $\nu\hat{\sigma}^2/\sigma^2 \sim \chi^2(\nu)$, and $\hat{\sigma}^2$ is independent of $\hat{\underline{\beta}}$.

Scheffé further assumes that $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ will be estimated by least squares.

The method in its general form is a graphic procedure. After the parameters have been estimated the calibration curve $\hat{Y} = \underline{g}'(X)\hat{\underline{\beta}}$ is drawn, then the tables supplied by Scheffé (1973) are used to draw confidence bands for the regression. One then obtains the confidence interval for X given Y = y by projecting the intersections of the line Y = y and the confidence bands onto the X-axis. It is not necessary to be able to analytically invert either the calibration curve or the confidence bands. In fact, obtaining an inversion may be difficult and time consuming; this method renders it unnecessary. However, in the case of simple linear regression (i.e., $\underline{g}(X) = [1,X]'$) the inversion is possible and not too difficult. The intervals are based on the probability statement

(III.5) $P\{|\hat{Y}_0 - Y_0| \leq (2F)^{1/2}\hat{\sigma}^2[1+n^{-1}+(X-\bar{X})^2/\Sigma(X_i-\bar{X})^2]^{1/2}\} = 1 - \alpha$

where F is $1 - \alpha$ quantile of the $F_{(2,\nu)}$ distribution.

This leads to the confidence interval for Y given X = x

$$(III.6)\quad \hat{Y}_0 - (2F)^{1/2}\hat{\sigma}[1 + n^{-1} + (X_0 - \bar{X})^2/\Sigma(X_i - \bar{X})^2]^{1/2} < Y_0$$
$$< \hat{Y}_0 + (2F)^{1/2}\hat{\sigma}[1 + n^{-1} + (X_0 - \bar{X})^2]\,\Sigma(X_i - \bar{X})^2]^{1/2}$$

Now given an observation Y_0, if we recall that $\xi\{Y_0\} = \alpha + \beta X_0$ and $Y_0 = \bar{Y} + \beta(X_0 - \bar{X})$ we can substitute these for Y_0 and $\hat{Y}_0$ into (III.6) and rearrange the terms to solve for X_0 obtaining

$$(III.7)\quad [\bar{X} + \hat{\beta}(Y_0 - \bar{Y})/\lambda] - [(2F)^{1/2}\hat{\sigma}/\lambda][\lambda(1 + n^{-1}) + (Y_0 - \bar{Y})^2/$$
$$\Sigma(X_i - \bar{X})^2]^{1/2} < X_0 < [\bar{X}_0 + \hat{\beta}(Y_0 - \bar{Y})/\lambda] + [(2F)^{1/2}\hat{\sigma}/\lambda]$$
$$[\lambda(1 + n^{-1}) + (Y_0 - \bar{Y})^2/\Sigma(X_i - \bar{X})^2]^{1/2}.$$

$$\text{where } \lambda = \hat{\beta}^2 - \frac{2F\hat{\sigma}^2}{\Sigma(X-\bar{X})^2}.$$

The interval above is an algebraic representation of the projection method used for more complex models. The interval width is the horizontal distance between the hyperbolic confidence bands at $Y = Y_0$. Clearly then, the interval width is an increasing function of $|Y_0 - \bar{Y}|$.

Scheffé's (1973) procedure is based on a graphic application of his simultaneous confidence interval procedure, which uses the F - distribution to form the probability statements. An alternative method uses the bivariate t distribution (see Graybill (1976), Trout and Swallow (1979)). This method yields intervals which hold simultaneously with confidence level $1 - \alpha$ for all X in a predetermined interval. Unlike Scheffé's intervals, these intervals have a constant width. The drawback to this "Uniform confidence band" method is its requirement of tables of the quantity

$$(III.8)\quad \int_{-D}^{D}\int_{-AD}^{AD} (2\pi)^{-1}(1-p^2)^{-1/2}\left[1 + \frac{u^2 - 2puv + v^2}{(n-2)(1-p^2)}\right]^{-n/2} dudv = 1-\alpha$$

for various values of n, p, A and α. These are available in Bowden and Graybill (1966) and Trout and Swallow (1979).

To find the uniform confidence bands we define

(III.9) $$g = [1 + n^{-1} + (X_L - \bar{X})^2/\Sigma(X_i - \bar{X})^2]^{1/2}$$

$$h = [1 + n^{-1} + (X_u - \bar{X})^2/\Sigma(X_i - \bar{X})^2]^{1/2}$$

$$p = [1 + n^{-1} + (X_L - \bar{X})(X_u - \bar{X})/\Sigma(X_i - \bar{X})^2]/gh$$

$$A = g/h$$

where $[X_L, X_u]$ is the interval over which confidence intervals are wanted.

Then a set of $100(1 - \alpha)\%$ confidence intervals can be based on the probability statement

(III.10) $$P\{|Y_0 - \hat{Y}_0| \geq \delta\hat{\sigma} | X_0 \epsilon [X_L, X_u]\} = 1 - \alpha$$

where $\delta = gD$.

Table 1 from Trout and Swallow (1979) gives D for various values of A, n-2, α and $|p|$. Now we can invert the interval by using the same method used to invert Scheffé's interval. This yields the confidence interval

(III.11) $$(Y - \hat{\alpha})/\hat{\beta} - \hat{\sigma}\delta/\hat{\beta} < X_0 < (Y - \hat{\alpha})/\hat{\beta} + \hat{\sigma}\delta/\hat{\beta}$$

for all X in the range $[X_L, X_u]$.

The final interval estimate of X is also a simultaneous procedure, making use of the Bonferroni inequality and the Fieller's theorem interval developed earlier. Like the uniform bands, it holds for all Y in a predetermined interval. We will show this in the following theorem.

Theorem: A set of $100(1 - \alpha)\%$ simultaneous confidence intervals for the calibration of X over the range of measurement $[Y_L, Y_u]$ is

$L < X_0 < U$,

(III.12)
$$L = (b_1 - a_1)Y_0/(Y_u - Y_L) + (Y_u a_1 - Y_L b_1)/(Y_u - Y_L)$$

$$U = (b_2 - a_2)Y_0/(Y_u - Y_L) + (Y_u a_2 - Y_L b_2)/(Y_u - Y_L)$$

where $a_1 < X_L < a_2$ is $100(1 - \alpha)\%$ confidence interval for X at Y_L and $b_1 < X_u < b_2$ is a $100(1 - \alpha_u)\%$ confidence interval for X at Y_u, and $\alpha_L + \alpha_u = \alpha$. The intervals at Y_L and Y_u are found by the Fieller's theorem method.

Proof: By the premises (a_1, a_2) and (b_1, b_2) are confidence intervals for X_L and X_u at confidence levels $1 - \alpha_L$ and $1 - \alpha_u$, respectively. By the Bonferroni inequality then, both statements hold simultaneously with confidence level $1 - \alpha_L - \alpha_u = 1 - \alpha$. Now, suppose that (L,U) is not a confidence interval for X_0 with confidence level at least $(1 - \alpha)$. Then the true $100(1 - \alpha)\%$ confidence interval is (C,D), $C < L$, $D > U$. Now suppose $Y_0 = Y_L$. Then (a_1, a_2) is not a $100(1 - \alpha)\%$ confidence interval for X_L. But by the premise (a_1, a_2) has confidence level $1 - \alpha_L > 1 - \alpha$, which is a contradiction. Hence (L,U) must have confidence level $(1 - \alpha)$, and the statement holds for all Y_0 in the interval $[Y_L, Y_u]$ simultaneously.

qed.

Graphically, this method is finding a calibration line for X, then putting confidence intervals for X at the boundaries of a region of interest. Now the points (a_1, Y_L), (a_2, Y_L), (b_1, Y_u) and (b_2, Y_u) are added, and the

trapezoid connecting those four points is drawn in. To find the confidence interval at some specified y in $[Y_L, Y_u]$, the line Y = y is drawn and the intersection of this line and the trapezoid is projected on the X-axis.

These last methods (the Uniform Band and the Bonferroni) are of particular use when confidence intervals are wanted only if X falls in a particular region. For example, in a medical lab, if the test results indicate that nothing abnormal is happening we may not want confidence intervals. On the other hand, if extreme values are indicated, a confidence interval may be desirable. These methods allow us to tailor the interval for the region of interest, while Scheffé's method fits intervals for all values of Y. It is for this reason that we can obtain sharper intervals than Scheffé, when we are willing to restrict our interest to a subset of the real line.

IV. A PRACTICAL EXAMPLE

As a numeric example of the preceeding techniques, we will use the following data given by Afifi and Azen (1972). The data are lactic acid concentration in millimoles/liter of solution (mM) in a known solution (X), and lactic acid concentration (mM) of the solution as determined by a meter (Y). Each reading of Y represents an independently prepared standard solution of concentration X.

(IV.1)

X		Y
1mM	(4)	1.1mM, 0.7mM, 1.8mM, 0.4mM
3mM	(5)	3.0mM, 1.4mM, 4.9mM, 4.4mM, 4.5mM
5mM	(3)	7.3mM, 8.2mM, 6.2mM
10mM	(4)	12.0mM, 13.1mM, 12.6mM, 13.2mM
15mM	(4)	18.7mM, 19.7mM, 17.4mM, 17.1mM

$\Sigma X_i Y_i = 1769.6$ $\quad \bar{X} = 6.7$

$\Sigma X_i^2 = 1424$ $\quad \bar{Y} = 8.385$

$\Sigma Y_i^2 = 2220.21$ $\quad n = 20$

The least squares estimators of α and β are

(IV.2)
$$\hat{\alpha} = 0.1595$$
$$\hat{\beta} = 1.2277$$
$$\hat{\sigma}^2 = 1.07874.$$

Hence the classical estimator is

(IV.3) $$\hat{X} = (Y - \hat{\alpha})/\hat{\beta} = (Y - 0.1595)/1.2277.$$

The inverse parameter estimates are

$$\text{(IV.4)} \qquad \hat{\delta} = \Sigma(X_i - \bar{X})(Y_i - \bar{Y})/\Sigma(Y_i - \bar{Y})^2 = 0.0458$$
$$\hat{\gamma} = \bar{X} - \hat{\delta}\bar{Y} = 0.7936$$

and the inverse estimate is

$$\text{(IV.5)} \qquad \overset{\vee}{X} = \hat{\gamma} + \hat{\delta}Y = 0.7936 + 0.0458Y.$$

The inverse estimator is also the minimum IMSE estimator if we choose the weighting function to have mean 6.7 and variance 26.31. If we choose some other weighting function, say a uniform distribution over (1,15), then we have a mean M = 7.5 and variance V = 16.33. In this case, the optimal estimator is

$$\text{(IV.6)} \qquad \tilde{X}_0 = \hat{\lambda}_0 + \hat{\lambda}Y$$
$$= \{M\hat{\sigma}^2 + \hat{\beta}V(Y - \hat{\alpha})\}/(\hat{\sigma}^2 + V\hat{\sigma}^2)$$
$$= \{8.091 + 20.048(Y - 0.1595)\}/25.6921 \ .$$

The asymptotic 95% confidence interval is

$$\text{(IV.7)} \qquad \hat{X} - 1.96\sqrt{\hat{\sigma}^2/\hat{\beta}^2} < X < X + 1.96\sqrt{\hat{\sigma}^2/\hat{\beta}^2} \qquad \text{or}$$
$$X - 1.658 < X < X + 1.658.$$

Fieller's interval is based on the distribution of $(a,b)' = (Y - \hat{\alpha},\hat{\beta})' \sim N_2[(\beta X,\beta)',\sigma^2(\underline{X}'\underline{X})^{-1}]$

$$(X'X)^{-1} = \begin{bmatrix} 1.13530977 & 0.01273280 \\ 0.01273280 & 0.00190042 \end{bmatrix}$$

So $v_{11} = 1.13530977$, $v_{12} = 0.01273280$, $v_{22} = 0.0190042$

$k = st/b = (1.07874)^{1/2}(2.101)/1.2277 = 1.7774.$

The interval then is

$$(IV.8) \qquad \{\hat{X} - k^2 v_{12} \pm k[v_{11} - 2\hat{X}v_{12} + \hat{X}^2 v_{22} - k^2(v_{11}v_{22} - v_{12}^2)]^{1/2}\}/(1-k^2 v_{22})$$

$$= \{\hat{X} - .0402 \pm 1.7774[1.0972 - 0.255X + .019X^2]\}/.9400.$$

Suppose that $\hat{X} = \bar{X}$. Then the asymptotic interval is $5.04 < X < 8.36$. The interval based on Fieller's theorem is $4.56 < X < 9.16$. If $X = 12$, then the asymptotic interval is $10.34 < X < 13.66$. The Fieller's theorem interval is $9.17 < X < 16.27$.

We will calculate the Scheffé and Uniform width intervals. The calculation of Scheffé's interval requires $F_{(.05,2,18)} = 3.550$

$$\lambda = \hat{\beta}^2 - (2F_{(2,15)}\hat{\sigma}^2/\Sigma(X_i - \bar{X})^2) = 1.4927.$$

and the interval is

$$(IV.9) \qquad [\bar{X} + \hat{\beta}(Y - \bar{Y})/\lambda] \pm [(2F)^{1/2}\hat{\sigma}/\lambda]\{\lambda(1 + n^{-1}) + (Y_0 - \bar{Y})^2/$$

$$\Sigma(X_i - \bar{X})^2\}^{1/2}$$

$$= [6.7 + 1.2277(Y - 8.385)/1.4927] \pm 1.8883\{1.5668 + (Y - 8.385)^2/$$

$$526.2\}^{1/2}$$

Again, if $Y_0 = \bar{Y}$ then $\hat{X} = \bar{X}$, and this will be the narrowest point in this interval, which will be $4.3364 < X < 9.0636$. If $\hat{X} = 12$, then $Y = 14.8919$, and the confidence interval will be $7.29 < X < 12.05$. We note that this interval, like the Fieller's theorem interval is not symmetric about $\hat{X}$.

To calculate the uniform width intervals we must find

$$g^* = \left[1 + n^{-1} + \frac{(X_L - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]^{1/2} = 1.0544$$

$$h^* = (1 + n^{-1} + (X_u - \bar{X})^2/\Sigma(X_i - \bar{X})^2] = 1.0867$$

$$A^* = g^*/h^* = .9703$$

$$p^* = (1 + n^{-1} + (X_L - \bar{X})(X_u - \bar{X})/\Sigma(X_i - \bar{X})^2]/g^*h^* = .9144$$

when we take $(X_L, X_u) = (1,15)$. Now, from Bowden and Graybill's (1963) Table 3, $D_{(.95,10,1)} = .9144$, so $\delta^* = g^*D = 2.5306$, and the confidence interval is

$$\text{(IV.10)} \qquad \hat{X} \pm \hat{\sigma}\delta^*/\hat{\beta} = \hat{X} \pm 2.1409.$$

The interval for $\hat{X} = 6.7$ is $4.5591 < X < 8.8409$. For $X = 12$, the interval is $9.8591 < X < 14.1409$.

V. SUMMARY AND CONCLUSIONS

In this report we have considered the solutions currently available to the seemingly simple problem of establishing and using a calibration line. Most practitioners use either the MLE of X, $\hat{X}$ or the inverse estimator $\check{X}$. Neither group seems to be aware of the theoretical background of their estimator. Some instrumentation systems have a calibration feature built into their mocroprocessor unit, although they sometimes fail to tell the user what calibration technique is being used. Each of the available estimators is in some sense optimal. The MLE has been shown to be a Consistent Asymptotic Normal Estimator, while the Inverse estimator minimizes the IMSE when the weighting function is chosen properly. Perhaps now the pejorative term "wrong regression" for the inverse estimator will be allowed to die. The "practically unbiased estimator" has the appealing property of removing the bias from the MLE in small samples. Work remaining to be done in point estimation includes the derivation of either the exact distribution of X or of an asymptotic distribution so that confidence intervals can be developed for this estimator.

The results in the area of interval estimation are more extensive than those in point estimation, in the sense that more interval estimates are available. It is not known how far back the asymptotic interval dates, although the author is unaware of any published proof prior to this date. The oldest proven interval estimate of X is due to Fieller (1954). Scheffé's method of calibration appeared in 1973 and is the first set of simultaneous confidence intervals. Trout and Swallow (1979) added the Uniform Interval approach, while the Bonferroni Method has been modified from a technique used by Khorasani and Milliken (1980).

All the currently available interval estimates are based on the classical estimate, since they involve fitting the model

$$(V.1) \qquad Y_i = \alpha + \beta X_i + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

If the asymptotic distributions can be found for the other point estimators it may be possible to develop interval estimates based on these estimators. Until then, an experimenter who wants confidence intervals is restricted to using $\hat{X}$.

For practical use it is the author's opinion that for one at a time confidence intervals the practitioner should use the Fieller's theorem interval. This is recommended for two reasons, first because the interval is conservative in that the confidence level is not less than $1 - \alpha$. Despite its conservativity it manages to do quite well nonetheless (see Chapter IV) in most reasonable situations. Secondly, the only competing interval at this time is the asymptotic interval, and its behavior for small samples is not known.

If a set of simultaneous intervals are desired there are now three methods available. For some problems, Scheffé's method may yield totally useless results. If the user wants intervals for any possible value of X he has no choice but to use Scheffé intervals. However, if he needs intervals only over a particular range of values he will be better off with either the Bonferroni or Uniform techniques. In using the Bonferroni and Uniform techniques the practitioner should calculate both regions, and then choose the better region for that problem. Since the expected widths are unknown for both intervals there is currently no theoretical reason to prefer one technique over the other.

As is always the case when several competing techniques are available for use, the final decision on the methods to use lie with the practitioner.

The choice should not be made blindly, in ignorance of the properties of the other tools available. Hopefully the reader now has some reasons for using the methods he chooses to use.

LITERATURE CITED

Afifi, A. A., and Azen, S. P. 1972. Statistical Analysis: A Computer Oriented Approach, Academic Press, New York.

Bennett, C. A. and Franklin, N. L. 1963. Statistical Analysis in Chemistry and the Chemical Industry, Wiley, New York.

Berkson, J. 1967. Estimation of a linear function of a calibration line: consideration of a recent proposal. Technometrics 11:4.

Bowden, D. C. and Graybill, F. A. 1966. Confidence bands of uniform and proportional width for linear models. JASA 61:182-198.

Brownlee, K. A. 1967. Statistical Theory and Methodology in Science and Engineering, Wiley, New York.

Eisenhart, C. 1939. The interpretation of certain regression methods and their use in biological and industrial research.

Fieller, E. C. 1954. Problems in interval estimation. JRSS-B 16:175-185.

Finney. 1964. Statistical Methods in Bioassay.

Graybill, F. A. 1976. Theory and Application of the Linear Model. Duxbury North Scituate, Mass.

Khorasani, F. and Milliken, G. 1980. Simultaneous uniform confidence bands for a segmented lines regression model. To be published in Technometrics.

Krutchkoff, R. G. 1967. Classical and inverse methods of calibration. Technometrics 9:425.

Naszodi, L. J. 1978. Elimination of the bias in the course of calibration. Technometrics 20:201.

Saw, J. G. 1970. Letter to the Editor. Technometrics 12:937.

Scheffé, H. 1973. A statistical theory of calibration. Annals of Statistics 1:1-37.

Shukla, G. K. 1972. On the problem of calibration. Technometrics 14:547-553.

Williams, E. J. 1969a. A note on regression methods in calibration. Technometrics 11:189.

_____. 1969b. Regression models in calibration problems. Proc. 37th Session, Bulletin of the International Statistical Institute, Vol. 43, Book 1, pp. 17-28.

Winslow, G. H. 1976. Some statistical aspects of the calibration and use of linear measurement systems. Journal of the Institute of Nuclear Materials Management, Vol. 5, no. 1.

SOME STATISTICAL ASPECTS OF THE CALIBRATION PROBLEM

by

DENNIS L. CLASON
B. S. California State College Bakersfield, 1976

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas
1980

The problem of predicting the value of a carrier variable (X) which generated an observation (Y) is discussed. The problem is motivated by two hypothetical examples which illustrate the problem.

Four point estimators are given; the Classical estimator ($\hat{X}$), the Inverse estimator ($\check{X}$), the Linear Minimum Integrated Mean Square Error estimator ($\tilde{X}_c$) and a "Practically Unbiased Estimation" (X^+). Conditions are given so that $\tilde{X}_c = \check{X}$. $\hat{X}$ is shown to be a maximum likelihood estimate and an asymptotic distribution is given for $\hat{X}$.

Five confidence intervals are given. The first is based upon the asymptotic distribution of $\hat{X}$, the second relies on an application of Fieller's (1954) theorem. These two intervals have confidence levels which hold for the individual interval. The remaining intervals have confidence levels which hold for all X in a specified interval. These are based on Scheffe's (1973) prediction bands, the Uniform Confidence bands of Bowden and Graybill (1966), and a new interval using the Bonferroni inequality. Comparisons are made among these methods.

Also, a numerical example is given.