The Korean Communications
in Statistics Vol. 7, No. 3, 2000
pp. 885-897

Latent Variable Fit to Interlaboratory Studies

Gyeongbae Jeon[1)]

Abstract

The use of an unweighted mean and of separate tests is part of the current practice for analyzing interlaboratory studies, and we hope to improve on this method. We fit, using maximum likelihood (ML), a rather intricate, multi-parameter measurement model with the material's true value as a latent variable in a situation where quite serviceable regression and ANOVA calculations have already been developed. The model fit leads to both a weighted estimate of the overall mean, and to tests for equality of means, slopes and variances. Maximum likelihood tests for difference among variances poses a challenge in that the likelihood can easily become unbounded. Thus the major objective become to provide a useful test of variance equality.

Keywords : Interlaboratory studies, Latent variable, Maximum likelihood estimates

1. Interlaboratory studies

There are two ways in which interlaboratory studies are conducted. The first is designed to monitor laboratories to see good agreement between the results obtained by different laboratories, by allowing each laboratory to compare its results with those obtained by other laboratories and take remedial action, if necessary. It is often referred to as "proficiency testing." Here, the laboratories themselves are of primary concern. The second is concerned not so much with the laboratories as with the method of measurement. Tests performed with presumably identical materials, in presumably identical circumstances usually do not yield identical results. This is attributed to unavoidable random error inherent in every test procedure; the factors that may influence the outcome of a test cannot all be completely controlled. Many different factors contribute to the variability in application of a test method, including the operator, equipment used, calibration of the equipment, and environment.

Data from an interlaboratory study typically include two or three determinations from each of a number of laboratories all measuring the same characteristic on each of a number of carefully prepared materials. Although the data are in simple format, the statistical issues are not that simple.

1) Lecturer, Department of Statistics, Sungkyunkwan University, Seoul, Korea
E-mail: gjeon@channeli.net

Committee E11 of the American Society for Testing Materials(ASTM) has issued a software program with its standard E691 deals with the evaluation of methods of measurement in terms of reproducibility and repeatability. As a supplement to E691, J. Mandel proposed additional calculations using an unweighted average to estimate the materials's property and fitting a row-linear model to examine what mathematical model underlies the data. Proctor(1991) used nonlinear, generalized least squares to fit to sample covariances of the latent model (2.4) below. Fuller(1987) also noted that the model (2.4) was the psychometric single factor model. His development of maximum likelihood estimators follows closely that of Lawley and Maxwell(1971). Pantula and Fuller(1986) derived algorithm computing the maximum likelihood estimator and the estimated covariance matrix of the estimators of the factor model under factor vector is distributed as an normal random vector. Fuller and Proctor became bogged down in problems of non-convergence and solutions falling on the boundary of the parameter space. Thus the stage is set for our work to develop fitting methods and hypotheses tests.

2. Latent variable models

A classical measurement error model decomposes the recorded measurement X into a true value T and a random error E, *i.e.*,

$$X=T+E, \tag{2.1}$$

where E(E) = 0 and var(E) = σ^2, a constant. Elaborations of the model arise from specifying the true value T to accommodate a number of materials and a number of laboratories along with allowing for non-constant variance.

The first elaboration treats laboratories as a random effect and each material as fixed and given. The purpose of the model is to characterize the measurement method's precision. The model used is:

$$X_{ijk}= \tau_j+ l_{ij}+ \varepsilon_{ijk}, \tag{2.2}$$

where τ_j is the true value for the material j, l_{ij} is a random effect of the laboratory and e_{ijk} is a random effect for the "r" replicate. The random effects are taken to have zero means, to be independent and to have variances $\sigma^2_{l_j}$ and $\sigma^2_{e_j}$, depending only on the material.

Another way Model (2.2) has been found to fit poorly is in differences among laboratories. Laboratories may have additive biases. They may be more or less sensitive in tracking the materials. This is sometimes called scale bias. Finally, they may show more or less scatter around these systematic effects. The model with these features is:

$$X_{ijk}= \mu_i+ \beta_i \tau_j+ \Delta_{ij}+ \delta_{ij}+ \varepsilon_{ijk}, \tag{2.3}$$

where we not only consider the M material effects as fixed but also consider the L laboratory effects as fixed. The term Δ_{ij} is a non-linearity deviation for laboratory i with material j.

ε_{ijk} are random effects for replicates with zero means and independent from the other terms. They may have different variances $\sigma^2_{\varepsilon_{ij}}$ by laboratories and by materials. δ_{ij} are taken to be random with zero mean and independent with different variance $\sigma^2_{\delta_{ij}}$ by laboratories and materials. The μ_i represent laboratory additive effects and their differences suggest possible laboratory biases.

We consider latent variable models for the interlaboratory study setting and present a full maximum likelihood approach when both laboratories and materials are assumed fixed. With the base model given by below, eight variations on the model concerning the equality of the μ_i's, β_i's and σ_i^2's are considered.

The model we consider is a revised version of (2.3) in which there is only one replicate observation(r=1) and no provision for non-linearity. It is

$$X_{ij} = \mu_i + \beta_i \tau_j + \varepsilon_{ij}, \tag{2.4}$$

where i=1,2,...,L indexes laboratories or tests and j=1,2,...,M indexes specimens of materials or subjects. The response variable X_{ij} is an observation, a test result, by laboratory i on material j. The μ_i component is an additive effect of laboratory i. The parameters μ_i and β_i describe the measurement bias and measurement sensitivity characteristics of laboratory i. To preserve these meanings and to remove a problem of identification, we designate some one laboratory as "Number L" and set $\mu_L = 0$ with $\beta_L = 1$ for it. The τ_j represent the material's unknown fixed true values. It is assumed that ε_{ij} is normally distributed with mean zero and variance σ_i^2. It is further assumed that

$$E(\varepsilon_{ij}) = 0 \text{ for all } i, j$$

$$E(\varepsilon_{ij}^2) = \sigma_i^2 \text{ for all } j.$$

Further, $E(\varepsilon_{ij}\varepsilon_{i'j}) = 0$ for all j and for all pairs (i, i')

and $E(\varepsilon_{ij}\varepsilon_{ij'}) = 0$ for all i and for all pairs (j, j').

That is, laboratories are independent but may have different variances. We also assume that ε_{ij} is distributed independently of τ_j. The variances are here assumed not to depend on the material. This condition would seem rather unrealistic. Relationships between level of a material and variances are commonly experienced (Horwitz, Kamps, and Boyer 1980). In fact the E691 Standard requires reporting a separate standard deviation for each material. There are two reasons that justify our assumption. We begin data analysis by checking for outliers and for variance heterogeneity by materials and take whatever power transform is needed to homogenize variances. Secondly, whatever variance heterogeneity among materials might remain is not expected to affect, to any noticeable extent, our application of the methods we develop for laboratory means and

laboratory slopes and laboratory variances. That is, material variances are, roughly speaking, ancillary to the test statistics and estimates for laboratories. Certainly, the differences among the τ_j's themselves may not be ancillary, but their tie to the error variances will nearly be so.

The model specification for X_{ij} in (2.4) is similar to that in regression theory in that it is a linear combination of other variables. Here, however, τ_j, which plays the role of the independent variable, is not directly observed. We can distinguish between two kinds of models for the τ_j's. In one we conside τ_j to be random, as subjects may be a random sample in psychometric studies. In the other we consider τ_j to be nonrandom quantities that differ from one material to another. Here in the interlaboratory study setting, τ_j values are properly considered fixed. That is, the test materials were specifically chosen to span a range of levels for the interlaboratory experiment. The components τ_j are usually latent quantities, representing true values for the materials, but some interlaboratory study are conducted with, so-called reference materials and then the τ_j would be given. For the interlaboratory study that most of our data come from, the τ_j are not known.

To determine estimates of the unknown parameters, we use maximum likelihood which minimizes -log *lik*, where

$$lik(\mu_1,\cdots,\mu_L,\beta_1,\cdots,\beta_L,\tau_1,\cdots,\tau_M,\sigma_1^2,\cdots,\sigma_L^2)=$$
$$\frac{1}{(2\pi)^{\frac{LM}{2}}}\prod\frac{1}{(\sigma_i^2)^{\frac{M}{2}}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{M}\frac{(X_{ij}-\mu_i-\beta_i\tau_j)^2}{\sigma_i^2}\right\} \tag{2.5}$$

This *lik* is the likelihood function obtained under the assumption that the observations have a multivariate normal distribution.

2.1 Maximum likelihood estimation

The log likelihood function for the L by M observations X_{ij}, assuming a multivariate normal distribution is;

$$\log lik = constant - \frac{M}{2}\sum_{i=1}^{L}\log\sigma_i^2-\frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{M}\frac{(X_{ij}-\mu_i-\beta_i\tau_j)^2}{\sigma_i^2} \tag{2.6}$$

Maximization (2.6) is equivalent to minimizing

$$F(\mu_1,\cdots,\mu_L,\beta_1,\cdots,\beta_L,\tau_1,\cdots,\tau_M,\sigma_1^2,\cdots,\sigma_L^2)=-\log lik \tag{2.7}$$

For maximum likelihood estimates, we need the first-order partial derivatives with respect to parameters μ_i, β_i, τ_i, and σ_i^2 and set these equations to zero. Because of the necessity of confirming the correctness of the measurement method, the standards organizations require at least

5 laboratories, and usually there are 8 or more participating. There will also be as many or more materials. Thus the number of observations $n=ML$ will fairly exceed the number of parameters which is at most $3L+M-2$. These likelihood equations cannot in general be solved algebraically, and thus iterative methods are applied to calculate or approximate the maximum likelihood estimates numerically.

To start the iterative procedure we require reasonably good initial estimates. We base initial estimates on the results of regressing X_{ij} on $\hat{\tau}_j = \overline{X}_{.j}$. The iterative maximum likelihood procedure seems to converge satisfactorily for models having the same error variance for every participating laboratory. But we cannot obtain maximum likelihood estimates for models having different error variances for each laboratory because the likelihood function (2.5) is unbounded. Anderson and Rubin(1956) showed that the likelihood function does not have a maximum. To show this fact, note that if $\mu_L = 0$, $\beta_L = 1$ and $\tau_j = X_{Lj}$ then $(X_{Lj} - \mu_L - \beta_L \tau_j)$ is zero so σ_L^2 can be made arbitrarily small without changing the quantity in the exponent of (2.5). As σ_L^2 approaches zero, the likelihood is unbounded. Thus the likelihood function has no maximum, and maximum likelihood estimates do not exist.

The likelihood with all variances equal is well behaved and the calculation of maximum likelihood estimates is relatively routine. The likelihood with unequal, unconstrained variances is unbounded with some variance estimates falling on the boundary of the parameter space ($\sigma_i^2 = 0$ for some i). Thus it is not possible to obtain a meaningful test of

$$H_0: \ \sigma_1^2 = \cdots = \sigma_L^2$$

$$H_1: \text{ not all } \sigma_i^2 \text{ equal.}$$

However, in interlaboratory study applications it is unlikely that the full generality inherent in H_1 is necessary. Likely departures from H_0 are those in which the laboratories are segregated into subgroups defined by commonality of variances. For example, with $L=5$ it would be unusual to have 5 unique variances $\sigma_i^2 \neq \sigma_j^2$, $i \neq j$, $i, j = 1, \cdots, 5$. A more likely occurrence is one in which, the variances are segregated into two groups with common variance within groups, e.g., $\sigma_1^2 = \sigma_4^2$, $\sigma_2^2 = \sigma_3^2 = \sigma_5^2$ with $\sigma_1^2 \neq \sigma_2^2$.

The expectation that variances will likely segregate into such groups provides the rational for constructing various null and alternative hypotheses which incorporate grouping of the variances. These have the advantages of both practical relevance and boundedness of the likelihood provided all of the groups contain at least two laboratories.

In the following sections a number of alternative hypotheses are formulated depending on different grouping criteria. Because the groupings are not known it is necessary to consider all possible groupings of a particular type. We propose and investigate the suitability of using

Bonferroni corrections to address the problem of multiple groupings. The unboundedness of the likelihood can also be addressed directly by placing meaningful lower bounds on the smallest variance. This method is investigated first, then we discuss methods based on grouping.

We recognize that the various test statistics we propose do not have the usual asymptotic χ^2 distributions generally associated with likelihood ratio tests because of the ordering of the variances implied by the groupings. Therefore, the suitability of the usual large-sample likelihood ratio theory, combined with Bonferroni adjustment should be further studied.

2.2 Method A of Formulating Alternative Hypothesis

For homogeneity of variances the null hypothesis of particular interest is:

$$H_0: \ \sigma_1^2 = \cdots = \sigma_L^2 = \sigma_0^2$$

Our first suggested approach to resolve the problem of unbounded likelihood was to take a lower bound for variance estimates at a small positive value. We took a laboratory in turn and fixed its variance at a reasonably small value. Since it would be unrealistic that any laboratory should be more than ten times as precise as the others, we take the small value as one tenth of the average of the all starting variances. Let $\sigma_i^{2(0)}$ denote the variance starting values, the small value $\tilde{\sigma}^2$ is defined as $\frac{1}{10}\sum_{i=1}^{L}\sigma_i^{2(0)}/L$ and is called the lower bound we use. This lower bound represents an extreme amount of departure when all laboratories are using the same method.

For the L^{th} laboratory, $\hat{\mu}_L$ and $\hat{\beta}_L$ are set to zero and one respectively. If laboratory k has its variance fixed at $\tilde{\sigma}^2$, then a hypothesis of some interest is that, apart from σ_k^2, the σ_i^2's are all equal. Let $\sigma_{0(k)}^2$ denote this common value, the null hypothesis is written:

$$H_{0A_k}: \ \sigma_i^2 = \sigma_{0(k)}^2 \geq \tilde{\sigma}^2.$$

The minimum value of the -log likelihood under $H_{0(k)}$ is found and denoted by $L(\hat{\omega}_k)$.

$$L(\hat{\omega}_k) = \min F(\underline{\zeta}, \ \sigma_{0(k)}^2), \ \underline{\zeta} = (\mu_1, \cdots, \mu_L, \beta_1, \cdots, \beta_L, \tau_1, \cdots, \tau_M)$$

where $$F(\underline{\zeta}, \ \sigma_{0(k)}^2) = constant + \frac{M}{2}\left(\log \tilde{\sigma}^2 + (L-1)\log \sigma_{0(k)}^2\right) +$$

$$\frac{1}{2}\left(\frac{1}{\tilde{\sigma}^2}\sum_{j=1}^{M}(X_{kj} - \mu_k - \beta_k \tau_j)^2 + \frac{1}{\sigma_{0(k)}^2}\sum_{\substack{i=1 \\ i\neq k}}^{L}\sum_{j=1}^{M}(X_{ij} - \mu_i - \beta_i \tau_j)^2\right).$$

The alternative states that:

$$H_{1A_k}: \ \sigma_i^2 \geq \tilde{\sigma}^2 \ , (i \neq k).$$

The minimum under H_{1A_k} can be found and is denoted by $L(\hat{\Omega}_k)$:

$$L(\hat{\Omega}_k) = \min F(\underline{\zeta}, \sigma_i^2 \mid i=1,\cdots,L;\ i\neq k)$$

where $$F(\underline{\zeta}, \sigma_i^2) = constant + \frac{M}{2}\left(\log \tilde{\sigma}^2 + \sum_{\substack{i=1\\ i\neq k}}^{L} \log \sigma_i^2\right) +$$

$$\frac{1}{2}\left(\frac{1}{\tilde{\sigma}^2}\sum_{j=1}^{M}(X_{kj}-\mu_k-\beta_k\tau_j)^2 + \sum_{\substack{i=1\\ i\neq k}}^{L}\sum_{j=1}^{M}\frac{(X_{ij}-\mu_i-\beta_i\tau_j)^2}{\sigma_i^2}\right).$$

If one of the variance estimates $\hat{\sigma}_i^2$ $(i=1,\cdots,k-1,k+1,\cdots,L)$ was found equal to, say $\hat{\sigma}_i^2 = \tilde{\sigma}^2$, it suggests that likelihood under H_{1A_k} is unbounded. This means that fixing laboratory k's variance at $\tilde{\sigma}^2$ is not applicable and another laboratory must have its variance fixed. Each laboratory has its variance fixed at a lower bound in turn. This would generate L values of $L(\hat{\omega}_k)$ and the corresponding L values of $L(\hat{\Omega}_k)$ for $k=1,2,\cdots,L$. Since any one of the laboratories could have the fixed variance, we must apply arguments from the topic of multiple comparison tests to adjust the calculation of a significance probability. If more than one $L(\hat{\Omega}_k)$ and $L(\hat{\omega}_k)$ is finite for each k, we should choose the test statistic chi-squared value as the maximum over all L differences of $L(\hat{\Omega}_k)$ and $L(\hat{\omega}_k)$. That is, we pick the one that has the largest χ^2 or smallest p-value and conduct the test at significance level $\frac{\alpha}{L}$ by applying the Bonferroni multiple comparison procedure. If the test with largest χ^2 is rejected at an adjusted significance level $\frac{\alpha}{L}$, this gives evidence that there exist precision differences among laboratories.

2.3. Method B of Formulating Alternative Hypothesis

Our next approach was to pair the participating laboratories. That is, laboratories $2i$-1 and $2i$ for $i=1,2,\cdots,\frac{L}{2}$ were assigned common variance σ_{pi}^2 if the number of laboratories L is even. Then we can obtain the maximum likelihood estimates for $\frac{L}{2}$ common variances. If L were odd then there would be $\frac{L-3}{2}$ common variances and one for a triple.

The null hypothesis of particular interest is that the σ_i^2's are all equal. Let σ_0^2 denote the common value for all laboratories. A reasonable hypothesis test is

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_L^2 = \sigma_0^2 \text{ or } H_0: \sigma_{p1}^2 = \sigma_{p2}^2 = \cdots = \sigma_0^2$$

The $L(\hat{\omega})$ under the null hypothesis H_0 is found by:

$$L(\hat{\omega}) = \min\ F(\underline{\zeta},\ \sigma_0^2),\ \underline{\zeta} = (\mu_1, \cdots, \mu_L, \beta_1, \cdots, \beta_L, \tau_1, \cdots, \tau_M)$$

where $$F(\underline{\zeta},\ \sigma_0^2) = constant + \frac{ML}{2}(\log \sigma_0^2) + \frac{1}{2}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^{L}\sum_{j=1}^{M}(X_{ij} - \mu_i - \beta_i \tau_j)^2\right).$$

If L is even, the alternative hypotheses H_{1B} states:

$$H_{1B}: \text{not all } \sigma_{pi}^2 \text{ are equal } \left(i = 1, \cdots, \frac{L}{2}\right)$$

The $L(\hat{\Omega})$ under the alternative hypothesis H_{1B} is found by:

$$L(\hat{\Omega}) = \min\ F\left(\underline{\zeta},\ \sigma_{pi}^2 \mid i = 1, \cdots, \frac{L}{2}\right)$$

where $$F(\underline{\zeta},\ \sigma_{pi}^2) = constant + \frac{M}{2}\left(2\sum_{i=1}^{\frac{L}{2}}\log \sigma_i^2\right) +$$

$$\frac{1}{2}\left(\sum_{i=1}^{\frac{L}{2}}\sum_{j=1}^{M}\frac{(X_{2i-1,j} - \mu_{2i-1} - \beta_{2i-1}\tau_j)^2 + (X_{2i,j} - \mu_{2i} - \beta_{2i}\tau_j)^2}{\sigma_{pi}^2}\right).$$

If L is odd, the alternative hypothesis H_{1B} states:

$$H_{1B}: \text{not all } \sigma_{pi}^2 \text{ are equal. } \left(i = 1, \cdots, \frac{L-1}{2}\right)$$

The $L(\hat{\Omega})$ under the alternative hypothesis H_{1B} is found by:

$$L(\hat{\Omega}) = \min\ F\left(\underline{\zeta},\ \sigma_{pi}^2 \mid i = 1, \cdots, \frac{L-1}{2}\right)$$

where $$F(\underline{\zeta},\ \sigma_{pi}^2) = constant + \frac{M}{2}\left(3\log \sigma_{p1}^2 + 2\sum_{i=2}^{\frac{L-1}{2}}\log \sigma_{pi}^2\right) +$$

$$\frac{1}{2}\left(\sum_{i=1}^{3}\sum_{j=1}^{M}\frac{(X_{ij} - \mu_i - \beta_i \tau_j)^2}{\sigma_{pi}^2} + \right.$$

$$\left.\sum_{i=2}^{\frac{L-1}{2}}\sum_{j=1}^{M}\frac{(X_{2i,j} - \mu_{2i} - \beta_{2i}\tau_j)^2 + (X_{2i+1,j} - \mu_{2i+1} - \beta_{2i+1}\tau_j)^2}{\sigma_{pi}^2}\right)$$

Let R be the number of possible combinations of pairs. As the number of participating laboratories goes larger, the number of possible combinations of pairs will be much larger. Rather than have some particular combination designated for test or rather than pick one combination at random, we suggest running all possible pairings and finding the minimum $L(\hat{\Omega}_k)$ under the

alternative hypothesis H_{1B_k} for $k=1,\cdots,R$. We pick the one that has the largest χ^2 or smallest p-value. To implement this approach and to apply overall tests of significance for comparing variances, we must apply a Bonferroni multiple comparison test at significance level $\frac{\alpha}{L}$.

2.4. Method C of Formulating Alternative Hypothesis

Instead of developing overall tests of significance for comparing variances in error of measurement, let's concentrate on identifying whether any single laboratory has a different error variance from the other laboratories. We would thus be particularly interested in comparing the imprecision of a single laboratory k, or σ_k^2, with imprecision of the others which are taken to equal one another.

Under the given condition:

$$\sigma_i^2 = \sigma_{0(k)}^2, \quad (i=1,\cdots,k-1,k+1,\cdots L) \tag{2.8}$$

where $\sigma_{0(k)}^2$ is unknown common variance of participating laboratories except laboratory k. Let σ_0^2 denote common variance when the σ_i^2's are all equal. The null and alternative hypotheses are formulated as:

$$H_0: \ \sigma_k^2 = \sigma_{0(k)}^2 \text{ or } H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_L^2 = \sigma_0^2$$

and $\quad H_{1C_k}: \ \sigma_k^2 \geq \sigma_{0(k)}^2$ with the given condition (2.8) is retained.

The $L(\hat{\omega})$ under the null hypothesis H_0 is found by:.

$$L(\hat{\omega}) = \min \ F(\underline{\zeta}, \ \sigma_0^2), \quad \underline{\zeta} = (\mu_1,\cdots,\mu_L,\beta_1,\cdots,\beta_L,\tau_1,\cdots,\tau_M)$$

where $$F(\underline{\zeta}, \ \sigma_0^2) = constant + \frac{ML}{2}(\log \sigma_0^2) + \frac{1}{2}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^{L}\sum_{j=1}^{M}(X_{ij}-\mu_i-\beta_i\tau_j)^2\right).$$

The $L(\hat{\Omega}_k)$ under the alternative hypothesis H_{1C_k} is found by:

$$L(\hat{\Omega}_k) = \min \ F(\underline{\zeta}, \ \sigma_k^2, \ \sigma_{0(k)}^2)$$

where $$F(\underline{\zeta}, \ \sigma_k^2, \sigma_{0(k)}^2) = constant + \frac{M}{2}(\log \sigma_k^2 + (L-1)\log \sigma_{0(k)}^2) +$$

$$\frac{1}{2}\left(\frac{1}{\sigma_k^2}\sum_{j=1}^{M}(X_{kj}-\mu_k-\beta_k\tau_j)^2 + \frac{1}{\sigma_{0(k)}^2}\sum_{\substack{i=1\\ i\neq k}}^{L}\sum_{j=1}^{M}(X_{ij}-\mu_i-\beta_i\tau_j)^2\right).$$

For some data observations, convergence may not be achieved if the data suggest that another laboratory has a smaller variance than the initially designated common variance of the grouped laboratories. If this happens, the likelihood is unbounded and a minimum under alternative

hypothesis H_{1C_k} does not exist. Thus we run all possible groups and find the groups satisfying the constraint $\sigma_k^2 \geq \sigma_{0(k)}^2$. If more than one group converged, we would pick the one that has the largest χ^2 or the smallest p-values. We would thus apply the Bonferroni multiple comparison test at significance level $\frac{\alpha}{L}$.

2.5. Method D of Formulating Alternative Hypothesis

This method combines features of Methods B and C. We pair two laboratories and assume them to have a common variance while the other variances are free to differ. If we consider the two as "reference" laboratories, for the time being, then we might be interested in comparing the imprecision of the remaining individual laboratories and a statistical test of significance to do this would of course be of some utility. Denote these reference laboratories as $a1$ and $a2$ and their common variance as $\sigma_{(a1,a2)}^2$. Let σ_0^2 denote the common variance when the σ_i^2's are all equal. The null and alternative hypotheses are formulated as:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_L^2 = \sigma_0^2$$

$$\text{and } H_{1D}: \sigma_i^2 \geq \sigma_{(a1,a2)}^2, \quad \text{for } i \neq a1 \text{ nor } a2$$

where $\sigma_{(a1,a2)}^2$ is the common variance of $a1$ and $a2$.

The $L(\hat{\omega})$ under the null hypothesis H_0 is found by:

$$L(\hat{\omega}) = \min F(\underline{\zeta}, \sigma_0^2), \quad \underline{\zeta} = (\mu_1, \cdots, \mu_L, \beta_1, \cdots, \beta_L, \tau_1, \cdots, \tau_M)$$

where $$F(\underline{\zeta}, \sigma_0^2) = constant + \frac{ML}{2}(\log \sigma_0^2) + \frac{1}{2}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^{L}\sum_{j=1}^{M}(X_{ij} - \mu_i - \beta_i \tau_j)^2\right).$$

The $L(\hat{\Omega})$, under the alternative hypothesis H_{1D}, is obtained by

$$L(\hat{\Omega}) = \min F(\underline{\zeta}, \sigma_{(a1,a2)}^2, \sigma_i^2 \mid i = 1, \cdots, L: i \neq a1 \ nor \ a2)$$

where

$$F(\underline{\zeta}, \sigma_{(a1,a2)}^2, \sigma_i^2 \mid i = 1, \cdots, L: i \neq a1 \ nor \ a2) = constant + \frac{M}{2}\left(2\log \sigma_{(a1,a2)}^2 + \sum_{\substack{i=1 \\ i \neq a1,a2}}^{L} \log \sigma_i^2\right)$$

$$+ \frac{1}{2}\left(\frac{1}{\sigma_{(a1,a2)}^2}\sum_{i=a1,a2}\sum_{j=1}^{M}(X_{ij} - \mu_i - \beta_i \tau_j)^2 + \sum_{\substack{i=1 \\ i \neq a1,a2}}^{L}\sum_{j=1}^{M}\frac{(X_{ij} - \mu_i - \beta_i \tau_j)^2}{\sigma_i^2}\right).$$

For some data observations, convergence may not be achieved if the data suggest that another

laboratory has a smaller variance than the initially designated common variance $\hat{\sigma}^2_{(a1,a2)}$ of the paired laboratories. If this happens, the likelihood is unbounded and a minimum under the alternative hypothesis H_{1D} does not exist. Thus we would run all possible pairs $\left(R=\frac{L(L-1)}{2}\right)$ and find pairs satisfying the constraints $\sigma^2_i \geq \sigma^2_{(a1,a2)}$ for $i \neq a1$ nor $a2$. If more than one pair converged, we would pick the one that has the largest χ^2 or the smallest p-value. We would thus apply the Bonferroni multiple comparison test at significance level $\frac{2\alpha}{L(L-1)}$ for the alternative Method D.

2.6 Testing Hypothesis Formulation

Let H_0 be any specific hypothesis concerning the parametric structure of the model and let H_1 be an alternative less specific hypothesis. We can then test H_0 against H_1 by means of the likelihood ratio technique. Let $L(\hat{\omega})$ be the minimum of -log *lik* under H_0 and $L(\hat{\Omega})$ be the minimum under H_1. Then $T=2(L(\hat{\omega})-L(\hat{\Omega}))$. Under H_0 (and when $\hat{\Omega}$ and $\hat{\omega}$ are in reasonable proximity to one another), this T is distributed, in large samples, as a chi-square distribution with degrees of freedom equal to the difference in the number of independent parameters estimated under H_0 and H_1. Method D uses L-1 variance parameters under H_1 and a common variance parameter under H_0, and thus the chi-square test statistics from Method D have L-2 degrees of freedom. Method B uses $\frac{L}{2}$ or $\frac{L-1}{2}$ variance parameters under H_1 and a common variance parameter under H_0, and thus the test statistics T from Method B have $\frac{L-2}{2}$ or $\frac{L-3}{2}$ degrees of freedom. Likewise Method C uses 2 variance parameters under H_1 and a common variance parameter under H_0, thus test statistics from Method C has one degree of freedom. In case of Method A, the hypotheses H_0 and H_{1A_k} are not hierarchical. That is, the test statistic based on the change of likelihood from H_0 to H_{1A_k} could have close to a chi-squared distribution with L-*1* degrees of freedom but it may not. We prefer to use the chi-squared test statistic based on the reduction in likelihood from H_{0A_k} to H_{1A_k} for which $\hat{\omega}_k \subset \hat{\Omega}_k$ and which will give a chi-square distribution with L-*2* degrees of freedom, to test for variance inequalities.

We are questioning about the way laboratory performances differ and how they are tied to the parameters. Differences among the μ_i's which might be called additive biases are wide spread and can be easily explained by calibration problems. Differences by β_i's which are sometimes called scale biases also arise from relatively simple conditions. The difference among the σ^2_i's

[6] Mandel, J. (1995). Analyzing interlaboratory data according to ASTM Standard E-691, Paper presented in Atlanta at Symposium.

[7]Pantula. S. G. and Fuller, W. A. (1986). Computational algorithm for the factor model, Communications in Statistics, Simulation, Vol 15 (1) 227-259

[8] Proctor, C. H. (1991). Quick fitting covariance models to interlaboratory trial data. Paper given in Atlanta, 11, ASA, Unpublished.