Dr. Peter H. Westfall  
Editor: The American Statistician  
Texas Tech University  
Lubbock, TX 79409  

October 2, 2007  
RE: Manuscript MS07-086R  

Dear Dr. Westfall;

Thank you for forwarding the comments of the associate editor (AE) and your decision (9/26/2007) concerning manuscript MS07-086R. We would like to address the comments, since we believe that in each instance where our work was considered in ‘error’, the AE’s arguments are false. This is clarified in the discussion presented below.

If the ‘errors’ attributed by the AE did occur in the manuscript, we can completely understand your agreement with the question of whether “TAS readers really need to see this.” However, in the light of the AE’s technical misunderstanding, we believe that this conclusion is not warranted.

We can understand how conceptual issues in Statistics (one of which, we believe this paper addresses) can be readily misunderstood, even by knowledgeable and experienced statisticians. The fundamental issue presented in this manuscript is indeed counter-intuitive and appears to contradict common practice. However, progress in this respect can be made by the ‘give and take’ process of peer review. It is with this perspective that we ask you to reconsider your decision.

**Associate Editor’s Comments and Discussion**

**Associate Editor’s Comment 1.**  
*I personally found this an interesting subject for a paper, although I am not sure how well I think it is directed towards the American Statistician audience. I had some problems following it, but that may be because I think the manuscript contains at least 2 substantial errors.*

**Discussion of Associate Editor’s Comment 1.**  
We feel that the simple problem identified in the introduction is one for which statisticians commonly believe that they know what to do. This manuscript illustrates that the foundations usually employed for the solution are incomplete/ or may be commonly misunderstood. We consider this as an issue of broad interest for the readers of TAS.
Associate Editor’s Comment 2.
The first major error is that you use the fact that equation (2) gives biased estimates to motivate the transition to section 2. Let’s look at that again. We can rewrite (2) as

$$\hat{Y}_i = (1 - k_i) \hat{\mu} + k_i Y_i.$$  

Here $k_i$ is fixed so if the expected value of the left side is different from your value of 5, then either the expected value of $Y_i$ has to be different from 5 or the expected value of $\hat{\mu}$ has to be different from 5. Clearly, neither of those can be different from 5, so you must have made a computational mistake.

Discussion of Associate Editor’s Comment 2.
The bias of the predictor in equation (2) was a surprise, even for us when we first encountered it. In fact, it appears that $k_i$ is fixed and this is what leads one down the path followed by the AE to conclude that the predictor is unbiased. If this were correct in all cases, our statements in the manuscript would clearly be wrong!

The mistake in the AE’s reasoning has been shared by many statisticians over the years (including ourselves). This is counter-intuitive and appears to contradict prevailing statistical theory (which we feel deserves the attention of TAS readers). It arises from the fundamental ambiguity introduced by not distinguishing between a sample selection, and a subject in a population. This ambiguity is evident in the expression for

$$Y^*_i = \sum_{s=1}^{N} U_{is} Y_{i,s}$$ (page 6), representing the random variable corresponding to response for the $i^{th}$ selected subject. When response error is associated with $Y^*_i$, is it related to the position of the subject in a permutation ($i$, which we refer to as an index of the measurement condition), or to the subject (identified by the label $s$)? Either case is possible (as we illustrate in Tables 2 and 3). The AE’s conclusion applies when response error is related to the measurement conditions (positions), but not when it is associated with the subject.

An easy way to see the error in the AE’s logic is to examine the statement that “$k_i$ is fixed”. Recall that $k_i = \frac{\sigma^2}{\sigma^2 + \sigma_i^2}$. Clearly, $i$ is ‘fixed’ since it refers to the position in a permutation (commonly used as the sample index, i.e., $i = 1, \ldots, n$). However, what is meant by $\sigma_i^2$ is not clear when the response error variance depends on the subject, since we don’t know which subject is ‘realized’ in position $i$. For this reason, the statement that “$k_i$ is fixed” is false when the response variance depends on the subject.

Associate Editor’s Comment 3.
At the bottom of page 4 you give the impression that the subscripts are driving the analysis. Subscripts are what we want or need them to be. They never run the show. If the subscripts you have do not reflect reality, you need to find a model that does.
Discussion of Associate Editor’s Comment 3.
We agree that subscripts should not “run the show”. Nevertheless, the subscript \( i \) typically used in model (1): \( Y_i = \mu + B_i + E_i \) is not adequate to distinguish between the interpretation of \( W_i^* \) (in equation (8) given by \( Y_i^* = \mu + B_i + W_i^* \)) and the interpretation of \( \tilde{W}_i \) (in equation (10) given by \( \tilde{Y}_i = \mu + B_i + \tilde{W}_i \)). The basic cause of the problem is the loss of the identifiability of subjects in the expression \( Y_i^* = \sum_{s=1}^{N} U_{is} Y_{sk} \). Therefore, care must be taken in interpretation. The use of appropriate subscripts can help, as for example, by distinguishing \( \sigma^2_{se} \) from \( \sigma^2_{i} \), as we illustrate in the manuscript.

Associate Editor’s Comment 4.
I was a little upset about how you were defining the sampling on page 6. You start off talking about the marginal probabilities of \( U_{is} \) which led me to suppose that they must be independent. Then near the bottom of the page you get around to saying that sampling is without replacement. Similarly, my response to your statement that \( \xi \) indicates expectation with respect to sampling was, "what does that mean?" (See below.)

Discussion of Associate Editor’s Comment 4.
The AE assumed that the indicator random variables, \( U_{is} \), for \( i = 1,\ldots,N \) and \( s = 1,\ldots,N \) are independent based on the statement on page 6:

“Sampling is introduced via a set of indicator random variables, \( U_{is} \), that take on a value of one with probability \( 1/N \) if subject \( s \) is selected in position \( i \), \( i = 1,\ldots,N \), and zero otherwise.”

This assumption may be due in part to us not stating explicitly early on in the manuscript that sampling is without replacement. If we had done so, the AE may not have made the false assumption that the \( U_{is} \) are independent. The dependence between these random variables is clearly mentioned seven lines later, where we state:

“Assuming that sampling is without replacement implies that \( E_{\xi}(U_{is}U_{js},r) \) is zero if \( i = i^* \), \( s \neq s^* \) or \( i \neq i^* \), \( s = s^* \), but is equal to \( 1/N (N-1) \) if \( i \neq i^* \), \( s \neq s^* \), where \( \xi \) indicates expectation with respect to sampling.”

We feel that this concern is not substantive, and can be readily addressed by modifying the first statement as follows:

“Sampling without replacement is introduced via a set of correlated indicator random variables, \( U_{is} \), that take on a value of one with probability \( 1/N \) if subject \( s \) is selected in position \( i \), \( i = 1,\ldots,N \), and zero otherwise.”
Earlier on page 6 you mention that for simplicity you will assume only one measure for each subject. If that is the case, why are you burdening us with notation for multiple observations?

Discussion of Associate Editor’s Comment 5.
Indeed we mention on page 6 that there could be more than one measure per subject. In the next sentence, we state that “for simplicity, subsequently assume that only one measure is made on each sample subject.” This is not an unfair burden on the reader, since the remainder of the manuscript does not use the subscript $k$. However, this point is not consequential to the thrust of the manuscript. We could simplify the notation further if you and the AE consider this to be necessary.

Associate Editor’s Comment 6.
What I am going to label your second major mistake, I have not worked out the correct answers for. I believe you have a far too simplified view of the variance of the product of two random variables, even when they are independent. $W_i^*$ is a sum of the product of random variables. While the expected value of $W_{sk}$ is 0, which simplifies matters, the expected value of $U_{is}$ is most certainly not equal 0. Moreover, the random variables you are summing over are not independent. I feel like there is absolutely no reason to believe the variances you are stating on page 7.

Discussion of Associate Editor’s Comment 6.
This comment is likely a consequence of omission of the derivation of the expression for

$$\Sigma = \text{var}_{\mathbb{R}^n} (W^*) = \sigma_e^2 I_n$$

followed from the definitions of the random variables in

$$W^* = \begin{pmatrix} W_1^* & W_2^* & \cdots & W_n^* \end{pmatrix}'$$

where $W_i^* = \sum_{s=1}^{N} U_{is} W_{sk}$. We felt that a proof of this result was a technical issue that might detract the attention from the main ideas of the manuscript, and did not present the details. It is a special case of a result presented in the Appendix of a paper by Stanek and Singer (2004, JASA) that we reference in the manuscript. Since there is no clustering here, the derivation is simpler.

We illustrate the special case of the derivation in an appendix to this letter, showing that

$$\Sigma = \text{var}_{\mathbb{R}^n} (W^*) = \sigma_e^2 I_n$$. Please see Stanek and Singer (2004) for the original proof. The derivations could be included in a revised manuscript.

Associate Editor’s Comment 7.
After finding this second problem, I only skimmed the rest of the manuscript. The covariance matrices you end up with seem like they are probably of the proper form (when the components are properly identified). But I didn't notice anything in the paper that made me feel, "TAS readers really need to see this."

Discussion of Associate Editor’s Comment 7.
In light of the response to each of the AE’s comments, we would appreciate if you could reconsider the decision on the disposition of the manuscript and its value for TAS
readers. We feel that this fundamental statistical issue, though counter-intuitive and seemingly contradicive to the prevailing statistical theory, should be brought to TAS readers’ attention. We look forward to your response to our comments, and hope that these comments are of value in making your decision.
Appendix. Simplification of the Variance Under a Two-stage Random Permutation Model with Response Error to a Single Stage Random Permutation Model with Response Error

(simplified from Appendix A of Stanek and Singer 2004).

In the submitted manuscript (MS07-086R), \( \mathbf{Y}' \) is a vector of dimension \( n \times 1 \) corresponding to random variables for a simple random without replacement sample. In Appendix A of Stanek and Singer (2004), (assuming single stage sampling), the vector \( \mathbf{Y}' \) is of dimension \( N \times 1 \) representing a full set of random variables for a permutation of units in the population. We develop an expression for the variance of \( \mathbf{Y}' \) for this full \( N \times 1 \) vector, and then note that the variance of the sample random variables may be taken without loss of generality as the upper \( n \times n \) quadrant of the variance matrix.

We avoided matrix notation in the manuscript to keep the presentation simple. We illustrate here how to use matrix notation to develop expressions for the variance similar to Appendix A of Stanek and Singer (2004). Following this exposition, we include another sketch of the variance development based on summation notation (that may be more accessible to some readers). First, let \( \mathbf{U} = \left( \mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_N \right) \) denote the \( N \times N \) matrix with columns \( \mathbf{U}_s = \left( U_{1s}, U_{2s}, \ldots, U_{Ns} \right)' \) that define indicator random variables that generate the permutations of \( \mathbf{y} = \left( y_1, y_2, \ldots, y_N \right)' \), the vector of unobservable latent values for the labeled subjects. The vector representing the latent values of a permutation of subjects is \( \mathbf{Y}' = \mathbf{Uy} \). Let \( \mathbf{W}' = \left( \mathbf{W}_{1k}, \mathbf{W}_{2k}, \ldots, \mathbf{W}_{nk} \right)' \) be a vector of subject-specific response errors (for \( k = 1 \)), and we define \( \mathbf{W}' = \left( \mathbf{W}_1', \mathbf{W}_2', \ldots, \mathbf{W}_N' \right)' = \mathbf{UW}' \) to be an \( N \times 1 \) vector, including all positions in a permutation (instead of dimension \( n \times 1 \) ) defined in the submitted manuscript MS07-
With these assumptions, we represent random variables for a random permutation with response error as:

\[ Y^* = Y + W^*. \]

We evaluate the variance of \( Y^* \) using the full set of random variables, making use of the conditional expansion of the variance, i.e.,

\[
\text{var}_{\xi R} (Y^*) = E_{\xi} \left[ \text{var}_{R\xi} (Y^*) \right] + \text{var}_{\xi} \left[ E_{R\xi} (Y^*) \right].
\]

Given \( \xi \), and using \( W^* = UW^* \) and \( Y^* = Y + W^* \), it follows that

\[
\text{var}_{R\xi} (Y^*) = \left[ U \right] \text{var}_{R\xi} \left( W^* \right) \left[ U \right]' \quad \text{since} \ Y \text{ is non-stochastic. Since} \ \text{var}_{R\xi} \left( W^* \right) = \bigoplus_{s=1}^{N} \sigma_{se}^2
\]

and \( E_{R\xi} (Y^*) = Y \), it follows that

\[
\text{var}_{\xi R} (Y^*) = E_{\xi} \left[ U \left( \bigoplus_{s=1}^{N} \sigma_{se}^2 \right) U' \right] + \text{var}_{\xi} [Y].
\]

Now \( U \left( \bigoplus_{s=1}^{N} \sigma_{se}^2 \right) U' \) is a diagonal matrix with diagonal elements interchanged for all positions on the diagonal. As a result, defining \( \sigma_{e}^2 = \sum_{s=1}^{N} \sigma_{se}^2 / N \), we have

\[
E_{\xi} \left[ U \left( \bigoplus_{s=1}^{N} \sigma_{se}^2 \right) U' \right] = \sigma_{e}^2 I_N.
\]

We evaluate \( \text{var}_{\xi} (Y) = \text{var}_{\xi} \left[ Uy \right] \) next. Note that \( y \) is non-stochastic, and \( U \) is a permutation matrix. Using a vec expansion, it is straightforward to show that

\[
\text{var}_{\xi} \left[ \text{vec} \left( U \right) \right] = \frac{1}{N-1} P_N \bigotimes P_N \quad \text{where} \quad P_a = I_a - \frac{J_a}{a}. \ As \ result, \ \text{var}_{\xi} (Y) = \sigma^2 P_N.
\]

Combining these results, \( \text{var}_{\xi R} (Y^*) = \sigma_{e}^2 I_N + \sigma^2 P_N \). The upper \( n \times n \) matrix is the variance of the sample vector, as indicated.

**Alternative Derivation Using Summation Notation**
We claim that \( \text{var}_{\xi_R}(W^*_i) = \frac{1}{N} \sigma^2_s \), and \( \text{cov}_{\xi_R}(W^*_i, W^*_j) = 0 \) when \( i \neq i^* \). First, we consider the expression for \( \text{cov}_{\xi_R}(W^*_1, W^*_2) \).

\[
\text{cov}_{\xi_R}(W^*_1, W^*_2) = \text{cov}_{\xi_R} \left( \sum_{s=1}^{N} U_{1s} W_{sk}, \sum_{s' = 1}^{N} U_{2s'} W_{sk} \right)
= \sum_{s=1}^{N} \sum_{s' = 1}^{N} \text{cov}_{\xi_R}(U_{1s} W_{sk}, U_{2s'} W_{sk})
= \sum_{s=1}^{N} \sum_{s' = 1}^{N} E_{\xi_R}(U_{1s} U_{2s'} W_{sk} W_{sk})
\]

since \( E_{\xi_R}(U_{1s}, W_{sk}) = E_{\xi_R}(U_{2s'}, W_{sk}) = 0 \) because

\[
E_{\xi_R}(U_{1s}, W_{sk}) = E_{\xi} \left( E_{\xi|\xi}(U_{1s}, W_{sk}) \right) = E_{\xi} \left( U_{1s} E_{\xi|\xi}(W_{sk}) \right) = E_{\xi}(U_{1s} 0).
\]

On the other hand

\[
\text{cov}_{\xi_R}(W^*_1, W^*_2) = \sum_{s=1}^{N} \sum_{s' = 1}^{N} E_{\xi_R}(U_{1s} U_{2s'} W_{sk} W_{sk}) = 0,
\]

since, if \( s' = s \), then \( U_{1s} U_{2s'} = 0 \), and

\[
E_{\xi_R}(U_{1s} U_{2s'} W_{sk} W_{sk}) = E_{\xi} \left( E_{\xi|\xi}(U_{1s} U_{2s'} W_{sk} W_{sk}) \right)
= E_{\xi}(U_{1s} U_{2s'} 0) = 0
\]

Next, we consider \( \text{var}_{\xi_R}(W^*_1) \). We can express

\[
\text{var}_{\xi_R}(W^*_1) = E_{\xi_R} \left( \sum_{s=1}^{N} U_{1s} W_{sk} \right)^2
= \sum_{s=1}^{N} \sum_{s' = 1}^{N} E_{\xi_R}(U_{1s} W_{sk} U_{1s'} W_{sk})
= \sum_{s=1}^{N} E_{\xi_R}(U_{1s} W_{sk}^2) = \sum_{s=1}^{N} \frac{1}{N} \sigma^2_s
\]

since, if \( s' \neq s \) then \( U_{1s} U_{1s'} = 0 \) and

\[
E_{\xi_R}(U_{1s} W_{sk}^2) = E_{\xi} \left( U_{1s} E_{\xi|\xi}(W_{sk}^2) \right) = E_{\xi}(U_{1s} \sigma^2_s) = \frac{1}{N} \sigma^2_s.
\]

Reference