# NOTE M: AN ACTUAL PROBLEM REQUIRING UNBIASED STANDARD DEVIATIONS

## CHARLES D. HENDRIX, STATISTICAL CONSULTANT

This was concerned with "stack test data", as in "smokestacks". But more than just smokestacks... all sorts of point emissions. The pollutants were small particles, various metals, dioxins, and more. The issue was "how good are the EPA sampling/assay methods" for each pollutant. Data were collected by taking pairs of simultaneous samples (N = 2) near the top of many different stacks across different plants and industries. By doing this we were able to measure the variation "due to sampling and assay of the samples using the EPA methods". The physical samples, of course, had to be obtained by professional samplers. These folks are paid well to go to the tops of stacks and take samples using well-defined protocols. Sometimes we managed to get not just N = 2, but N = 3 and N = 4 and even (once) N = 8 simultaneous samples. The "big deal" was concerned with "how good are the EPA methods"... "good" being in terms of the inherent variation due to sampling and assay of the samples. Also, to build equations for "estimated Sigma (unbiased) = f(average concentration).

As you would suspect, the EPA had one position on how "good" and "industry" had another. When all is said and done, the regulations were written to compensate for the inherent variation in the EPA methods. Arguments about this had gone on for several years. I got invited to the party. The further I dug into it the more I realized there were two biases that neither side had recognized. Both were in the same direction. The first was the one cited here. The other one came from the fact that the relationship between "std dev" and "average" was (as best we could tell) linear on Ln-Ln paper. Hence in order to get a straight-line relationship (needed to make predictions) we had to fit the data in the Ln-Ln plane and then convert back to "real units". Right at this point there's another "statistical bias", and it was very large... in the range of 25% - 50%. The trick was to get all of this together (huge amounts of data from multiple sources... get the data "blessed" by all interested parties.... do the work for each of the pollutants... and then gain agreement that "we did it right".

To make it more interesting, some of the data came in triads (2 df) and some came as quads (3 df) and some came with even more df. But the majority of the data came in pairs (1 df)

Variances are additive, but the cited standard deviations must not be averaged. In the first place, they are biased estimates. In the second place the bias changes with the number of df in each sample.

To make it more interesting, the game here was concerned with finding a 0relationship between "estimated Sigma" and the sample averages. Imagine a graphic of the individual estimates of Sigma vs the respective averages. So we could not just "pool" the 1 df estimates of Sigma. The procedure evolved into (1) convert the biased estimates of Sigma into unbiased estimates and then run a weighted regression (weights are the respective df for the samples) to find the relationship. As you can imagine, the variation about the line of regression was large, but with sufficient data (and a very wide range of

averages) the relationship could be established. Substantial issues about environmental regulations hinged on this relationship. So to verify that the completed relationship was unbiased (and that confidence intervals on "Sigma" were about as expected) this entire procedure was simulated (Monte Carlo) extensively. This was the culmination of a multi-year project with a lot of interested parties.

So this is the one instance in which I tangled with unbiased estimates of Sigma... all from small samples of data. Once in 40+ years. But it was one of those things where knowing the presence of the bias and "fixing it" was crucial.

I hope I never have to go through one like this again. The approx. 20% difference in "unbiased vs. biased" was enough to (1) upset some people whom "count" and to please some other people who also "count". Meetings with the interested parties were agonizing. Explaining that bias was pure torture. Using a weighted regression evolved into a major "trust me".

The test method is in http://crtd.asme.org/imw/remap.pdf

Note: This was posted on the news group sci.stat.math in August 2006.