Statistics 531 **Homework #1** **Due February 5, 2009**

1. Consider estimation of σ from sample ranges and sample standard deviations.

   (a) Problem 1.1 of the notes. The notes refers to the printed Course Notes by Professor Vardeman, which can be downloaded at

   `http://www.public.iastate.edu/~isuhwu/stat531/stat531.html`

   (b) Problem 1.2 of the notes (just check for $n = 2, 3, 4$, and 5). (Use Table A.1 on page 509 of V&J (the text) to find values of $d_2(n)$ and $c_4(n)$. The conclusion here for these sample sizes is in fact true for all sample sizes.)

   (c) Problem 1.5 of the notes.

2. Show how formula 5.27 of V&J follows from the result in the appendix of the notes. (optional)

3. In the notation used in class, suppose that $\sigma_x = 1$ and $\sigma_{\text{measurement}} = .5$.

   (a) Find an approximate variance-covariance matrix for the random vector $\begin{bmatrix} \widehat{\sigma}_x \\ s \end{bmatrix}$.

   (b) Compute the determinant of this matrix for the two cases $\{n = 5 \ \text{ and } \ m = 10\}$ and $\{n = 10 \ \text{ and } \ m = 5\}$. If the determinant is taken to be a measure of "size" for a matrix, which of the two data collection plans (pairs of sample sizes) seems to be better if σ_x is expected to be about twice the size of $\sigma_{\text{measurement}}$ and one is going to estimate them using the vector $\begin{bmatrix} \widehat{\sigma}_x \\ s \end{bmatrix}$?

4. Consider a range-based estimator $\widetilde{\sigma}_x$ and a balanced one-way ANOVA based estimator $\check{\sigma}_x$ introduced in the class. Use the delta method and find explicit standard errors for $\widetilde{\sigma}_x$ and $\check{\sigma}_x$.

5. Use the data from Operator A (only) in V&J Problem 2.5 and compute "MSTr" and "MSE." (You may use a One-Way ANOVA program to do this if you like.) Then compute $\check{\sigma}_x$ and a standard error for it. Finally, either employ hand calculations based on Section 1.5 of the notes, or use the Brandon Paris program to find 90% two-sided confidence limits for σ_x. What feature of the data in this problem should make you a little skeptical about the real world relevance of the calculations here?

   Note: All the "course" software can be found from the following web site

   `http://www.public.iastate.edu/~isuhwu/stat531/Software/531software.html`

6. Run the data of V&J Problem 2.5 through a Two-Way ANOVA program to obtain the mean squares referred to in Section 1.4 of the notes. Use these mean squares and estimate $\sigma, \sigma_{\text{reproducibility}}$ and σ_{parts} by hand. Then run Andy Chiang's program and based on the output you obtain, discuss what has been learned from these data about the sizes of part (or "process") variation, reproducibility variation, and repeatability variation. (Note that the "answers" in the back of the text are based on sample ranges, not standard deviations.)

7. Problem 1.7 of the notes. (See V&J Section 2.2.3.)

8. Problem 1.12 of the notes.

9. Below is a data set from a real calibration study (taken from a paper by John Mandel of NBS/NIST). Unfortunately, the units are not given in the paper. For sake of argument, suppose that they are ppm (parts per million) of some contaminant. x is the "truth"/gold-standard-measurement. y is the local laboratory measurement. All on $n = 14$ different specimens.

x	y
647	605
728	675
1039	965
1095	995
1116	1018
1194	1117
1557	1422
1594	1470
1896	1762
1983	1739
2136	1918
2192	1983
2224	2008
2244	2010

(a) Fit a simple linear regression model to these data. For a fixed specimen, what do you estimate "$\sigma_{\text{measurement}}$" to be at the local laboratory? What "conversion formula" do you recommend for translating "local lab measurements" to estimated "gold standard measurements"?

(b) A new specimen is measured as $y = 2000$ at the local laboratory. Give a 95% confidence interval for the "true"/gold standard value for this specimen derived from inverting prediction limits for y. Then give an approximate 95% confidence interval using the delta method formula from page 11 of the notes.