Statistics 850 **yandell** **Spring 1995**

Homework # 7—Due Mon 10 Apr

NOTE: You CANNOT use the computer for this assignment. All you need is on the attached printout, plus the ability to square numbers.

Last Thursday I got an email from a colleague, Sam Weerahandi, at BellCore who is concerned about analysis of variance when variances are unequal. He sent me a small dataset and some discussion. Since it was so timely, you get to see it for homework!

There are four groups and a total of 31 observations. As you will see, the standard deviations range from 1.5 to 5.7. One might choose to ignore this variability, but read on!
1. Write done the usual assumptions and report the results of the usual analysis of variance.
2. BY HAND, make a "dot plot" of the data, using letter symbols for group. Comment.
Now drop the assumption that variances are equal.
3. Note that lack of obvious relationship between mean and variance by group (PLOT THEM).
4. Instead, use the inverse of group variance as weight. Briefly justify this choice (why might this be reasonable). Now briefly critique it (why is it silly in this problem). How does the use of estimates of group variances affect the p-value? [Hint: examine the new SDs.]
5. Replace the observations by their ranks and rerun the usual analysis of variance. [This test is an approximation of the Kruskal-Wallis test.] Comment on whether this a reasonable approach for this problem. [Hint: what assumption(s) are helped by using ranks?]
6. The exact test differs from all of these. It is based on the randomization principle. That is, if there are no group differences, then all assignments of group labels to the data are equally likely. That is, one could (in theory) examine every permutation of the 31 responses (with 9 A's, 7 B's, 8 C's and 7 D's) and compute the F statistic for each one. The p-value is then the proportion of F values that are as extreme or larger than the one observed. According to my source, the "right" p-value for the raw data is .030, or for the ranks (exact Kruskal-Wallis) is .06.
Comment on the disparity among p-values (you now have 5 different ones!). What is your conclusion about differences among groups? [Again, be brief. There is no "right" answer to my question, anyway!]

Now think about inference on the variances themselves.
7. Calculate the ratio of the largest to smallest variances. This is Hartley's F-max test. ["Liberal" (using sample size 9, 4 groups) critical values for 5% and 1% are, respectively, 7.18 and 11.7. See Milliken and Johnson Table A.1.] This is a very easy test to perform. Unfortunately, it relies heavily on the normal assumption, and is best for balanced data. Interpret results, with cautions for this data set.
8. Conduct Levene's test for unequal variance (see SAS printout). This test is not sensitive to departures from normality (based on simulation studies – see Milliken and Johnson), and can be used for small samples. Interpret results.
9. Comment briefly on the dilemna of testing for equal variance before conducting analysis of variance. How is this problem lessened (or greatened) by increasing sample size?