1

COMPARING VARIANCES OF SEVERAL MEASUREMENT METHODS USING A RANDOMIZED BLOCK DESIGN WITH REPEAT MEASUREMENTS: A CASE STUDY

Ajit C. Tamhane and Anthony J. Hayter

Northwestern University
Georgia Institute of Technology

Abstract: In this paper we consider the problem of comparing variances of several measurement methods in a randomized block design with repeat measurement methods. The analysis is presented in the context of an actual consulting study, which motivated this paper. We demonstrate how in a practical data analysis, a combination of informal graphical methods and formal inferential methods (multiple comparison methods) are employed to detect outliers, identify patterns and draw conclusions with confidence.

Keywords and phrases: Multiple comparisons of variances, graphical methods, normal theory methods, robust methods, mixed model, audiological measurements

1.1 Introduction

Professor Panchapakesan's contributions have been instrumental for a number of advances in the fields of ranking and selection, and multiple comparison procedures. Practical applications of these methods are still lagging, however. Most published applications use simple one-way layout designs. To spread the applications of these methods it is important to publish actual case studies involving more complex designs used in practice. The present paper is a contribution in this direction.

We consider the problem of comparing variances of several measurement methods/instruments when repeat measurements are made with each method on a randomly selected sample of subjects/items. Such a design is frequently used in laboratory experiments in physics, chemistry, biology and psychology, to name a few. The purpose here is to illustrate how a combination of graphical and formal inference (in particular, multiple comparison) methods aids in the analysis of data. The following example from our consulting experience is typical of such studies, and will be used to illustrate the various methods.

Example: The insertion gain of a hearing aid is defined as the sound pressure level (SPL) measured at the eardrum of the wearer with the hearing aid in place minus the SPL at the eardrum with no aid in place, the stimulus being the same under both conditions. For clinical measurement of insertion gain, the stimulus is presented over a nearby loudspeaker and the response is measured by a probe microphone in the ear canal of the subject. The standard practice was to locate the loudspeaker in the ear-level horizontal plane of the subject. It was claimed that loudspeaker locations above the horizontal plane would yield more precise (less variable) results. A study was conducted at the Department of Communication Studies and Disorders at Northwestern University to check this claim. The study compared the following loudspeaker locations:

- Location 0: $0°$ azimuth, $0°$ elevation (Standard/Control)
- Location 1: $45°$ azimuth, $0°$ elevation (New)
- Location 2: $0°$ azimuth, $90°$ elevation (New)
- Location 3: $45°$ azimuth, $45°$ elevation (New)

There were 10 subjects with five replicate measurements of insertion gain (obtained by measuring the SPL with and without the hearing aid in a random order in pairs) at each of the four loudspeaker locations. The order of the locations presented was randomized for each subject. Measurements of insertion gains were made at 6000 Hz frequency. Table 1.2 gives the raw data. The measurements are presented in the order they were taken. The investigator was primarily interested in comparing the within-subject variances for different measurement methods (loudspeaker locations). The locations apparently affect the variability of measurements because the measurement error depends on the angle of incidence of the sound waves at the ear drum.

The outline of the paper is as follows. Section 2 gives graphical analyses for detecting outliers and identifying patterns in the data. Section 3 shows how the conventional normal theory and related robust methods can be adapted to compare the variances in the present setting. Section 4 applies these methods to the data under study. Section 5 gives some concluding remarks.

1.2 Graphical Analyses and Descriptive Statistics

Before proceeding with detailed analyses it is important to examine the data for any excessive or systematic variability and to determine whether the data on any subjects should be discarded or modified. We first made box plots (shown in Figure 1.1) for each subject of his/her median centered measurements (using the separate median for each set of five subject $\times$ method measurements). Note that because of the median centering, four out of the twenty centered measurements for each subject are forced to be zero. From these plots it appears that subjects 3 and 6 have a much higher variability than other subjects. To examine the reasons for this high variability, run charts were made for the two subjects using the methods as labels. These plots are shown in Figures 1.2 and 1.3 for subjects 3 and 6, respectively. We see that there is something unusual going on with these two subjects: for subject 3, the second and fifth measurements are the largest while the remaining three are the smallest for each method. For subject 6, the pattern is more pronounced: measurements form two distinct clusters for each method with the second and the fourth measurements close to zero while the remaining three measurements are much larger, about 10 to 20. Such systematic patterns were not found for any other subjects. It could not be determined why these systematic patterns occurred for subjects 3 and 6, and not for others. Because of these systematic patterns (more than for reasons of high variability), these two subjects were discarded from further analyses. Note that these systematic patterns could not be revealed by the box plots.

In order to compare the within-subject variabilities of the methods, box plots were made of the same median centered measurements but now stratified by the method. Note that this forces eight of the 40 centered measurements (for eight subjects) to be zero. The result is shown in Figure 1.4. This plot provides a preliminary answer to the question under study, namely an indication that the current method has a higher variability than the new methods, with method 3 having the least variability. We shall investigate this suggestion more fully and determine whether any differences are significant.

Summary statistics for all cells (means $\overline{y}_{ij}$, variances s_{ij}^2 and logs of variances $x_{ij} = \ln s_{ij}^2$) as well as the corresponding statistics for the row (method) margins are shown in Table 1.3. (Subjects 3 and 6 are not included in the calculation of the marginal statistics for the methods.) These summary statistics generally confirm the findings of the graphical displays.

To study the structure of the cell means a mixed model analysis of variance (treating the methods fixed and the subjects random) of the data (omitting subjects 3 and 6) is given in Table 1.1. Using the F-tests as guidelines (because of the violation of the homoscedasticity assumption; however, due to the balanced design the F-tests should be quite robust), we see that all three effects

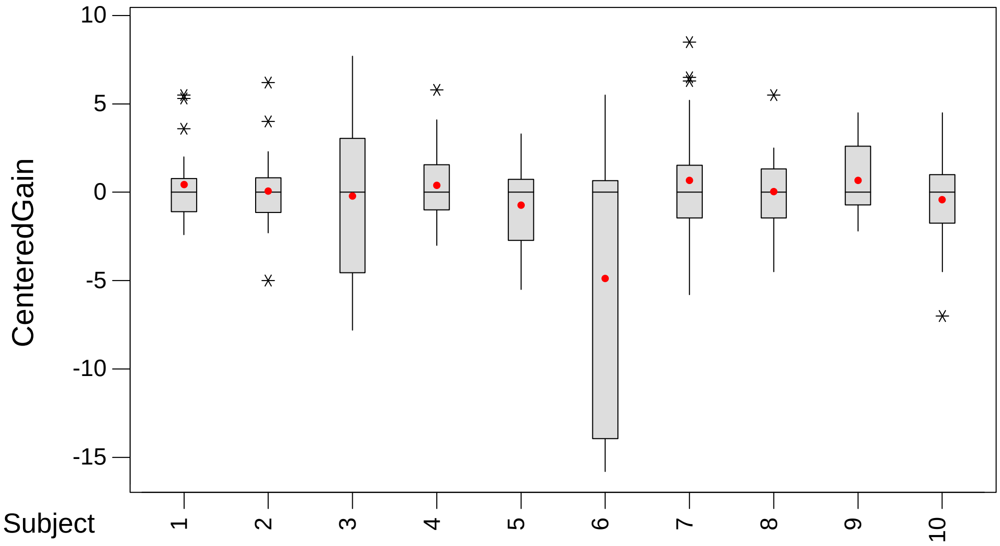

Figure 1.1: Box Plots of Median-Centered Insertion Gains for Subjects

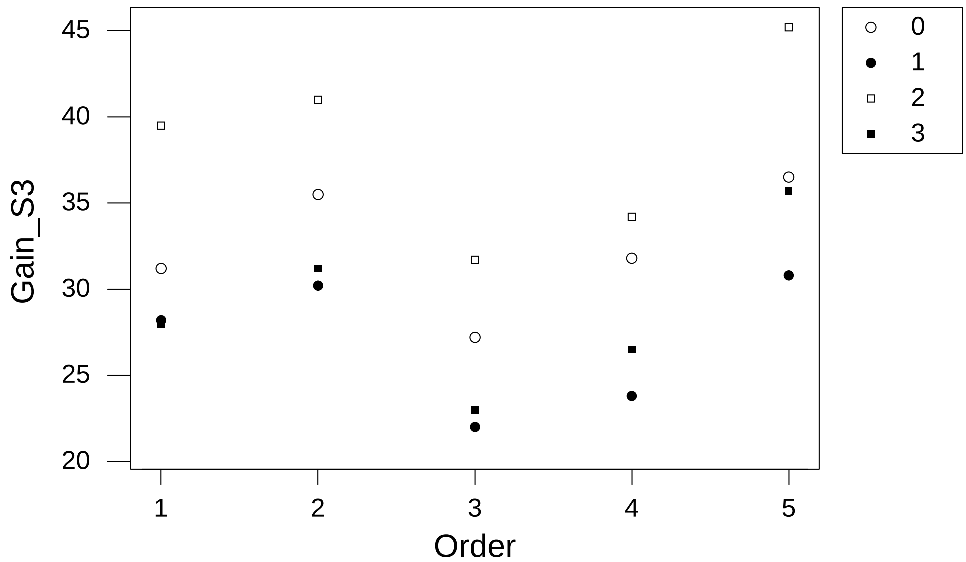

Figure 1.2: Run Chart of Insertion Gains for Subject 3

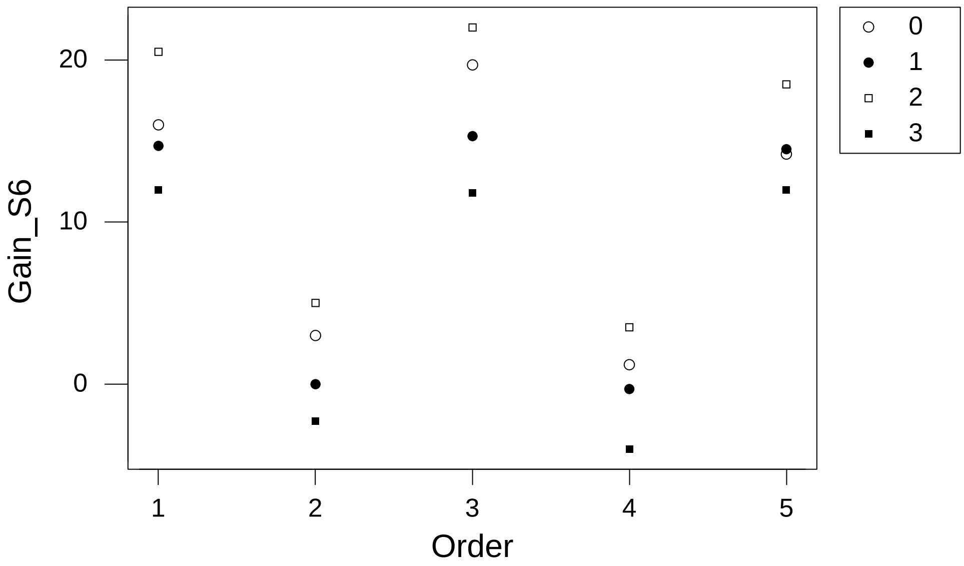

Figure 1.3: Run Chart of Insertion Gains for Subject 6

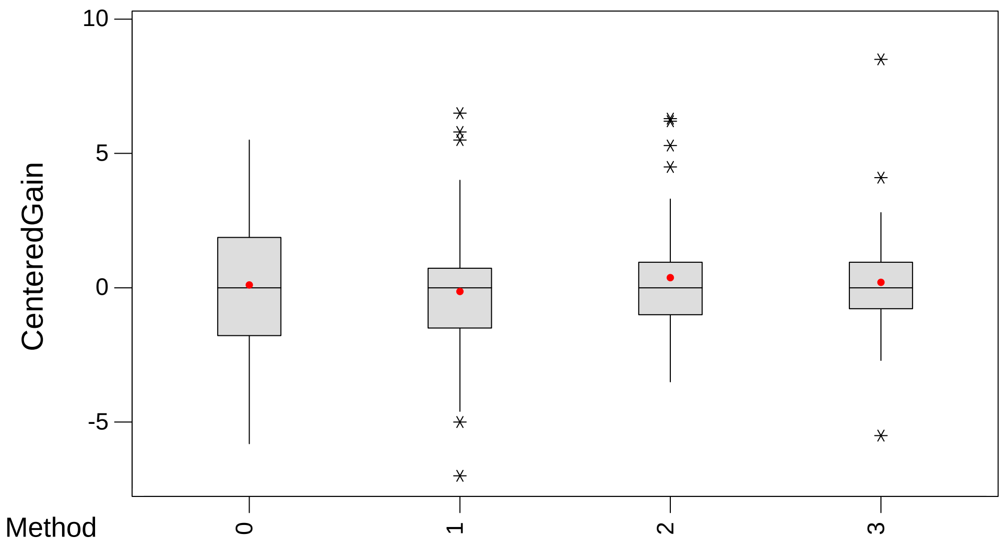

Figure 1.4: Box Plots of Median-Centered Insertion Gains for Methods

Table 1.1: Analysis of Variance

Source	SS	DF	MS	F	p
Methods	2225.8	3	741.92	35.80	0.000
Subjects	2593.2	7	370.45	17.87	0.000
Methods×Subjects	435.3	21	20.73	3.13	0.000
Error	848.8	128	6.63		
Total	6103.0	159			

as significant. However, the main effects of the methods and subjects dominate over their interaction effect. Thus the structure of the cell means appears to be additive. This can be verified by fitting an additive model to the data and studying the residuals.

One could make more plots to detect further quirks and patterns in the data, but one should be cautious of not crossing the boundary between prudently selected graphics and data-dredging. So we will stop here and summarize our findings thus far: subjects 3 and 6 were discarded because they exhibited excessive and, more importantly, systematic variation. After omitting these two subjects, the remaining subjects form a relatively homogeneous group, both with regard to their means and SDs. The data on these subjects indicate that method 3 has less within- subject variability than method 0. Methods differ in their mean levels (with method 2 giving the highest readings and method 1 giving the lowest readings for most subjects) and the structure of the means is roughly additive.

1.3 Formal Statistical Analyses

1.3.1 Model

Let there be $M \geq 2$ new measurement methods, whose variances are to be compared with a control method, labeled 0. Suppose we have available $N \geq 2$ randomly selected subjects and we make $n_{ij} \geq 2$ repeat measurements on the jth subject using the ith method ($0 \leq i \leq M$). The following mixed-effects model (Scheffé 1959, Ch. 8) is proposed for the data:

$$y_{ijk} = m_{ij} + e_{ijk} \quad (0 \leq i \leq M, 1 \leq j \leq N, 1 \leq k \leq n_{ij}).$$

Here $(m_{0j}, m_{1j}, \ldots, m_{Mj})$ for $j = 1, \ldots, N$ are assumed to be i.i.d. random vectors from an $(M+1)$-variate distribution (not necessarily normal) with mean vector $(\mu_0, \mu_1, \ldots, \mu_M)$ and covariance matrix $\mathbf{\Sigma} = \{\sigma_{ii'}\}$. Furthermore, for each method i, the measurement errors e_{ijk} are assumed to be i.i.d. (not necessarily normal) with mean 0 and variance σ_i^2, and they are independent of the m_{ij}. Note that in this model any two observations y_{ijk} and $y_{i'j'k'}$ are correlated (assuming a general $\mathbf{\Sigma}$ with nonzero off-diagonal elements) iff $j = j'$. All parameters in the model are unknown. The primary interest lies in comparing the imprecisions of the methods as measured by the σ_i^2.

1.3.2 Multiple Comparison Procedures

Let $\overline{y}_{ij\cdot}$ and the $\overline{e}_{ij\cdot}$ be the cell means of the y_{ijk} and e_{ijk}, respectively. Then the

$$z_{ijk} = y_{ijk} - \overline{y}_{ij\cdot} = e_{ijk} - \overline{e}_{ij\cdot}$$

are distributed independently of the $(m_{0j}, m_{1j}, \ldots, m_{Mj})$ with

$$E(z_{ijk}) = 0, \mathrm{Var}(z_{ijk}) = \sigma_i^2 \left(\frac{n_{ij} - 1}{n_{ij}} \right), \mathrm{Cov}(z_{ijk}, z_{ijk'}) = -\frac{\sigma_i^2}{n_{ij}}. \qquad (1)$$

Note that the widths of the box plots in Figure 1.4 are roughly proportional to $\mathrm{SD}(z_{ijk}) = \sigma_i \sqrt{(n_{ij} - 1)/n_{ij}}$. Furthermore, the z_{ijk} $(1 \leq k \leq n_{ij})$ are mutually independent for $0 \leq i \leq M$ and $1 \leq j \leq N$. Hence for each method i, the within-subject sample variances

$$s_{ij}^2 = \frac{\sum_{k=1}^{n_{ij}} z_{ijk}^2}{n_{ij} - 1} \qquad (2)$$

are mutually independent with

$$E(s_{ij}^2) = \sigma_i^2 \text{ and } \mathrm{Var}(s_{ij}^2) = \sigma_i^4 \left(\frac{2}{n_{ij} - 1} + \frac{\gamma_{ij}}{n_{ij}} \right),$$

where γ_{ij} is the kurtosis of the distribution of the e_{ijk} (see Scheffé (1959, Section 3.8)). Therefore standard procedures based on the independence assumption can be used to make inferences on the σ_i^2 from the s_{ij}^2. We briefly discuss such normal theory procedures in Section 3.2.1 and robust procedures (which do not assume that the e_{ijk} are normally distributed) in Section 3.2.2.

Normal Theory Procedures

Under the normality assumption, $\gamma_{ij} = 0, s_{ij}^2 \sim \sigma_i^2 \chi_{\nu_{ij}}^2 / \nu_{ij}$ and the minimum variance unbiased estimate of σ_i^2 is

$$s_i^2 = \frac{\sum_{j=1}^{N} \nu_{ij} s_{ij}^2}{\sum_{j=1}^{N} \nu_{ij}} \sim \sigma_i^2 \chi_{\nu_i}^2 / \nu_i,$$

where $\nu_{ij} = n_{ij} - 1$ and $\nu_i = \sum_{j=1}^{N} \nu_{ij}$. The usual normal theory procedures can be applied to the s_i^2 to make inferences about the σ_i^2. For example, Hartley's (1950) $F_{\max}$ test can be used if all pairwise comparisons between the σ_i^2 are of interest. An illustration of a normal theory procedure for comparing the new method variances σ_i^2 $(1 \le i \le 3)$ with σ_0^2 is given in Section 4.

Robust Procedures

It is well-known that normal theory procedures for variances are not robust when the assumption of normality is violated. Box (1953) proposed a robust procedure for testing the equality of variances in a one-way layout. This procedure randomly divides each sample into subsamples of equal sizes and calculates an estimate of the treatment variance from each subsample. In the present problem the necessity of forming random subsamples is obviated because for each method we have N "natural" subsamples, namely $\{y_{ijk}\ (1 \le k \le n_{ij})\}$ for $j = 1, \ldots, N$. Thus Box's procedure can be used as follows: Compute the subsample variances s_{ij}^2 using (2). Let

$$x_{ij} = \ln s_{ij}^2.$$

Then the x_{ij} are independent and, for large n_{ij}, are approximately normally distributed with

$$E(x_{ij}) \approx \ln \sigma_i^2 \text{ and } \mathrm{Var}(x_{ij}) \approx \frac{2}{n_{ij} - 1} + \frac{\gamma_{ij}}{n_{ij}};$$

see Scheffé (1959, Section 3.8). Which procedure is now appropriate depends on the sample sizes n_{ij} and the assumptions we are willing to make about the γ_{ij}. For instance, if $n_{ij} = n$ for all i, j and if we assume $\gamma_{ij} = \gamma$ for all i, j then the x_{ij} have approximately a constant variance. Therefore the usual ANOVA procedures can be used to compare the $\ln \sigma_i^2$. Multiple comparisons with a control can be carried out using Dunnett's (1955) method. This method is illustrated in the next section.

An alternative to the above procedure is to use a nonparametric procedure. Different approaches are possible, but a simple approach is as follows: Compute the Wilcoxon signed rank statistics for comparing the current method with each one of the new methods using the x_{ij} as the data values, and then refer these statistics to the critical point of the null distribution of the maximum of such correlated statistics. This test procedure due to Nemenyi (1963) is described in Section 2.3.2 of Chapter 9 of Hochberg and Tamhane (1987) and is illustrated in the following section. A test based on the signed ranks of $x_{0j} - x_{ij} = \ln(s_{0j}^2 / s_{ij}^2)$ will not only avoid the problems caused by lack of normality of the original data or lack of approximate normality of the x_{ij} caused by insufficiently large n, but will also be resistant to outliers in the orginal data.

1.4 Return to Example

We now return to the example of Section 1. Table 1.3 gives the cell variances, s_{ij}^2, and their logs, x_{ij}, for the data given in Table 1.2. As noted in Section 2, subjects 3 and 6 are deleted as outliers. All subsequent analyses are based on the remaining $N = 8$ subjects.

In this example, one-sided comparisons of new methods with the current method are of interest because new methods would be useful only if they reduced the variability as compared to the current method. To see if it is appropriate to analyze the data under the normality assumptions, we first made normal plots of the residuals z_{ijk} for each method. (Note that this strictly requires that the z_{ijk} be independent, which they are not. However, equation (1) gives the $\mathrm{Corr}(z_{ijk}, z_{ijk'}) = -1/(n-1) = -1/4$, which is fairly small.) To save space we have not shown the plots here. The plots do not show any gross violation of normality. Therefore we consider normal theory based $100(1-\alpha)\%$ simultaneous upper one-sided confidence limits:

$$\frac{\sigma_i^2}{\sigma_0^2} \leq \frac{1}{c}\frac{s_i^2}{s_0^2} \quad (i = 1, 2, 3).$$

Here c is the lower 100α percentage point of the smallest of M r.v.'s, $\chi_{\nu_i}^2/\chi_{\nu_0}^2 \ (1 \leq i \leq M)$, where the $\chi_{\nu_i}^2$ are independent chi- square r.v.'s with ν_i d.f. In the present example, $M = 3, \nu_i = 32$ and the s_i^2 are given in the last column of Table 1.3. The values of c have been tabulated by Gupta and Sobel (1962). For $1 - \alpha = 0.90, M = 3$ and common d.f. $= 32$ for all sample variances, we find $c = 0.5314$ from Table 3 in their paper. Therefore the upper confidence limits for σ_i^2/σ_0^2 for $i = 1, 2, 3$ are 1.858, 1.385 and 1.204, respectively. The ordering of these confidence limits are in agreement with Figure 1.4. Thus none of the ratios σ_i^2/σ_0^2 can be demonstrated to be less than 1, and hence none of the methods can be shown to significantly improve on the current method using a familywise type I error rate of $\alpha = 0.10$.

If one does not wish to make the normality assumption then a robust analysis can be based on the transformation $x_{ij} = \ln s_{ij}^2$ discussed in Section 3.2.2. In this case, approximate $100(1-\alpha)\%$ simultaneous upper one-sided confidence limits on $\ln(\sigma_i^2/\sigma_0^2)$ are given by

$$\ln(\sigma_i^2/\sigma_0^2) \leq \overline{x}_i - \overline{x}_0 + d s_x \sqrt{2/N} \quad (1 \leq i \leq M),$$

where d is the upper 100α percentage point of the maximum of M jointly distributed student t r.v.'s with $\nu = (M+1)(N-1)$ d.f. and common associated correlation coefficient $\rho = 1/2$. The values of d have been tabulated by Bechhofer and Dunnett (1988). For $1 - \alpha = 0.90, M = 3$ and $\nu = 28$, we find $d = 1.78935$ from Table 1 in their paper. Also, s_x, the pooled sample standard deviation of the x_{ij}, is calculated to be 0.733. Therefore simultaneous 90% upper one- sided confidence limits on $\ln(\sigma_i^2/\sigma_0^2)$ are calculated to be 0.667, 0.285 and -0.034 for $i = 1, 2, 3$, respectively. The corresponding confidence limits on σ_i^2/σ_0^2 are then 1.948, 1.330 and 0.967, respectively. These limits are not too different from the confidence limits calculated from the normal theory, but here we are able to show that method 3 is significantly better than the method 0. (This finding is in accord with the box plots shown in Figure 1.4.) It should be noted, however, that these robust confidence limits are based on approximate normality of the x_{ij} which holds only if the n_{ij} are large, whereas in the present example n_{ij} equal only 5.

If we do not wish to rely on the approximate normality of the x_{ij} then we can use the nonparametric tests referred to in Section 3.2.2. Note that the data x_{ij} are observed in blocks. The Wilcoxon signed rank statistics between method 0 and the three new methods are as follows: $W_1 = 13, W_2 = 22$ and $W_3 = 33$. It is well-known that the joint null distribution of W_1, W_2, W_3 is not distribution-free (see Hochberg and Tamhane (1987, p. 255)), and so the exact critical points of $\max(W_1, W_2, W_3)$ are not available. We can apply the Bonferroni method by comparing the one-sided marginal p-value of W_3 with $0.10/3 = 0.033$. This p-value is found to be 0.0195 from Table H in Lehmann (1975) for $N = 8$. Hence method 3 is shown to be significantly better than method 0 at a type I familywise error rate of $\alpha = 0.10$. Alternatively, we can calculate a large sample approximate critical value of $\max(W_1, W_2, W_3)$ given by (see equation (2.16) of Hochberg and Tamhane (1987, Ch. 9))

$$\frac{N(N+1)}{4} + \frac{1}{2} + z_{k,1/2,\alpha}\sqrt{\frac{N(N+1)(2N+1)}{24}},$$

where $z_{k,1/2,\alpha}$ is the upper α critical point of k equicorrelated standard normal variates with common correlation $= 1/2$. For $k = 3$ and $\alpha = 0.10$, we find $z_{k,1/2,\alpha} = 1.7336$ using Dunnett's (1989) program. By substituting $N = 8$, we obtain the desired critical value to be 30.88, which is exceeded by W_3. The other two methods are not significantly better than method 0.

One can use Steel's (1959) sign test procedure above if the assumption of the symmetric distribution of the $x_{ij} - x_{0j}$ is not valid. The three sign statistics are $S_1 = 3, S_2 = 5$ and $S_3 = 7$ (because only one subject, viz. subject 7, has less variability for method 0 than for method 3). Using Table 1 of Rhyne and Steel (1965) we find that the exact multiplicity adjusted p-value of $S_3 = 7$ equals $0.066 < 0.10$. Hence the conclusion drawn using the Wilcoxon signed rank is confirmed. Alternatively, we can calculate a large sample critical value of $\max(S_1, S_2, S_3)$ given by (see equation (2.7) of Hochberg and Tamhane (1987, Ch. 9))

$$\frac{N}{2} + \frac{1}{2} + z_{k,1/3,\alpha}\frac{\sqrt{N}}{2}.$$

For $k = 3$ and $\alpha = 0.10$, we find $z_{k,1/3,\alpha} = 1.7738$ using Dunnett's (1989) program available at the web site http://lib.stat.cmu.edu/general. By substituting $N = 8$, we obtain the desired critical value to be 7.01, which S_3 just fails to exceed.

1.5 Concluding Remarks

We have given an example of a case study that employed a combination of graphical methods and formal multiple comparison methods to arrive at practically useful conclusions. Graphical examination of the data is helpful in not only revealing the main patterns in the data, but also in identifying outliers that can vitiate the formal analyses. We demonstrated how standard multiple comparison methods for one-way layouts can be readily modified to a more complex (but balanced) design used here.

Based on the graphical displays and formal statistical analyses we can conclude that method 3 is more precise (less variable) than method 0 at $\alpha = 0.10$. It is interesting that only the normal theory method for comparing variances did not support this conclusion. Since this conclusion is based on the data from only eight subjects, it would be desirable to do further experimentation to confirm this finding. The investigator should be advised to find possible reasons for the systematic and excessive variability observed in subjects 3 and 6, and to avoid those reasons in future experimentation.

Acknowledgment: The authors wish to thank Mr. Larry Revit, a graduate student in the School of Communication Studies at Northwestern University, for allowing the use of the data that he gathered for his thesis research.

References

1. Bechhofer, R. E. and Dunnett, C. W. (1988). Tables of the multivariate Student's t distribution, *Selected Tables in Mathematical Statistics*, **11**, Providence, R.I.: American Mathematical Society.

2. Box, G. E. P. (1953). Non-normality and tests on variances, *Biometrika*, **40**, 318-335.

3. Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control, *Journal of the American Statistical Association*, **50**, 1096-1121.

4. Dunnett, C. W. (1989). Multivariate normal probability integrals with product correlation structure (Algorithm AS251), *Applied Statistics*, **38**, 564-579.

5. Gupta, S. S. and Sobel, M. (1962). On the smallest of several correlated F statistics, *Biometrika*, **49**, 509-523.

6. Hartley, H. O. (1950). On the maximum F-ratio as a shortcut test for heterogeneity of variances, *Biometrika*, **37**, 308-312.

7. Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*, New York: John Wiley.

8. Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

9. Nemenyi, P. (1963). Distribution-free multiple comparisons, unpublished doctoral dissertation, Priceton University, Princeton, NJ.

10. Rhyne, A. L. and Steel, R. G. D. (1965). Tables for a treatments versus control multiple comparison test, *Technometrics*, **7**, 293-306.

11. Scheffé, H. (1959). *The Analysis of Variance*, New York: John Wiley.

12. Steel, R. G. D. (1959). A multiple comparison sign test: Treatments vs. control, *Journal of the American Statistical Association*, **54**, 767-775.

Table 1.2: Insertion Gain (y_{ijk}) in Decibels

Methods	Subject									
	1	2	3	4	5	6	7	8	9	10
0	11.4	4.7	31.2	22.5	17.5	16.0	19.5	6.5	22.2	14.0
	11.5	4.5	35.5	19.0	11.5	3.0	14.3	12.0	21.0	7.5
	15.0	8.5	27.2	16.5	13.5	19.7	8.5	8.5	19.5	11.0
	10.0	3.8	31.8	18.6	18.0	1.2	15.0	2.0	18.0	7.5
	9.0	3.3	36.5	18.8	15.5	14.2	12.5	4.5	17.8	9.5
1	9.2	−1.0	28.2	22.8	14.0	14.7	16.0	7.5	16.5	0.0
	8.0	6.3	30.2	17.0	9.2	0.0	9.5	6.5	13.7	−2.5
	15.0	4.0	22.0	14.0	11.0	15.3	8.0	4.5	13.5	4.5
	10.3	2.5	23.8	15.5	16.0	−0.3	9.8	6.0	17.7	5.5
	9.5	4.0	30.8	17.3	13.8	14.5	8.2	4.0	12.8	4.5
2	16.8	16.0	39.5	26.2	20.3	20.5	25.3	13.0	25.2	19.0
	17.7	23.2	41.0	21.2	16.5	5.0	19.0	15.0	20.5	19.0
	23.8	17.0	31.7	23.5	18.5	22.0	17.2	12.5	20.7	17.0
	18.5	15.5	34.2	22.5	23.3	3.5	18.5	11.7	21.3	17.0
	19.2	17.3	45.2	24.3	20.0	18.5	20.8	10.7	18.5	18.0
3	11.3	4.7	28.0	19.3	15.5	12.0	20.0	3.8	13.5	7.0
	9.8	8.3	31.2	14.2	9.2	−2.3	11.5	5.0	10.7	4.5
	13.0	8.0	23.0	15.0	12.2	11.8	8.8	2.3	10.7	9.0
	11.0	6.2	26.5	15.2	15.2	−4.0	1.7	3.0	13.0	7.5
	10.7	7.0	35.7	17.0	14.7	12.0	10.8	1.7	9.7	7.0

Table 1.3: Cell Means $\overline{y}_{ij\cdot}$, Cell Variances s_{ij}^2 and Their Logs $x_{ij} = \ln s_{ij}^2$

Method	Subject										Average
	1	2	3	4	5	6	7	8	9	10	
0	11.40	4.96	32.44	19.08	15.20	10.82	13.96	6.70	19.70	9.90	12.613
	5.15	4.24	13.84	4.67	7.45	67.73	15.92	14.59	3.61	7.40	7.879
	1.639	1.445	2.628	1.541	2.008	4.216	2.768	2.680	1.284	2.001	1.921
1	10.40	3.16	27.00	17.32	12.80	8.84	10.30	5.70	14.84	2.40	9.615
	7.29	7.24	15.37	11.09	7.24	67.40	10.76	2.07	4.54	12.04	7.784
	1.987	1.980	2.732	2.406	1.980	4.211	2.376	0.728	1.513	2.488	1.932
2	19.20	17.80	38.32	23.54	19.72	13.90	20.16	12.58	21.24	18.00	19.030
	7.40	9.67	29.16	3.53	6.25	79.39	9.92	2.59	6.00	1.00	5.795
	2.001	2.269	3.373	1.261	1.833	4.374	2.295	0.952	1.792	0.000	1.550
3	11.16	6.84	28.88	16.14	13.36	5.90	12.56	3.16	11.52	7.00	10.218
	1.37	2.13	23.23	4.16	7.13	68.56	18.58	1.66	2.69	2.63	5.044
	0.315	0.756	3.145	1.426	1.964	4.228	2.922	0.507	0.990	0.967	1.231

The upper entry in each cell is $\overline{y}_{ij\cdot}$, the middle entry is s_{ij}^2 and the lower entry is $x_{ij} = \ln s_{ij}^2$. The entries in the 'Average' column are the corresponding averages, $\overline{y}_{i\cdot\cdot}$, s_i^2 and $\overline{x}_i$, over subjects (excluding subjects 3 and 6).