USING RANKS TO PERFORM EXACT AND ESTIMATED EXACT TESTS IN DESIGNED EXPERIMENTS

Scott J. Richter

Mark E. Payton

USING RANKS TO PERFORM EXACT AND ESTIMATED EXACT TESTS IN DESIGNED EXPERIMENTS

Scott J. Richter, Department of Mathematics, Western Kentucky University
Mark E. Payton, Department of Statistics, Oklahoma State University

ABSTRACT

A procedure is studied that uses rank transformed data to perform exact and estimated exact tests which is an alternative to the commonly used F-ratio test procedure. First, a common parametric test statistic is computed using rank transformed data, where two methods of ranking - ranks taken of the original observations, and ranks taken after aligning the observations - are studied. Significance is then determined using either the exact permutation distribution of the statistic or an estimate of this distribution based on a random sample of all possible permutations. Simulation studies compare the performance of this method to both the normal theory parametric F-test and the traditional rank transform procedure. Power and nominal type-I error rates are compared under conditions when normal theory assumptions are satisfied as well as when these assumptions are violated. The method is studied for a two factor factorial arrangement of treatments in a completely randomized design and also for a split-unit experiment.

1. INTRODUCTION

In experiments to determine if one or more factors have an effect on a response, the researcher typically can choose between one of two classes of analyses: parametric and nonparametric. Parametric procedures exist for both simple and complex experiments, but the validity of inferences made using these procedures depends on a set of unknown assumptions. The most common of these in the analysis of designed experiments is the assumption of normally distributed populations with equal variances. However, it is generally unknown to what extent the validity of the inferences suffers when the assumptions are not satisfied. Many nonparametric procedures, on the other hand, require less stringent assumptions, such as independent samples and observations, which can often be controlled by the experimenter. Furthermore, most of these methods depend on the exact permutation distribution of the test statistic for making inferences. However, due to the complexity of deriving the exact sampling distributions when sample sizes are large, most nonparametric methods rely on the asymptotic distribution of the test statistic. In addition, there exist few nonparametric procedures for analyzing complex experimental designs, and most of those that do exist are very limited in application.

Conover & Iman (1976) addressed this situation by proposing the procedure of performing parametric procedures on the ranks of the data when the parametric assumptions were suspected to be violated. Many studies of the "rank transform" procedure, however, have shown it to be non-robust and lacking in power in some situations, most notably in experiments where interaction is present (see Blair, et al. (1987), Sawilowsky, et al. (1989), Akritas (1990) and

Thompson & Ammann (1990)).

An adjustment to the usual rank transform, known as "ranking after alignment", was first proposed by Hodges & Lehmann (1962). This adjustment has been found to make the rank transform procedure more robust and more powerful in some situations, especially in designs with interaction. However, asymptotic sampling distributions are still used for tests of significance, and very few studies of the small sample properties are available. Fawcett & Salter (1984) and Groggel (1987) investigated the aligned rank procedure for testing main effects in a randomized block design. Conover & Iman (1976) examined the aligned rank procedure for testing for interaction in a two factor factorial experiment, using small effect magnitudes. Higgins et al. (1990) and Higgins & Tashtoush (1994) considered the aligned rank procedure for testing main effects and interaction in a two factor factorial experiment and also for testing main and sub-unit effects and interaction in a split-unit experiment.

In this paper, the performance of both the usual rank transform and the aligned rank transform is investigated when the exact permutation distribution of the sampling distribution of the test statistic is used. Simulation studies compare the performance of these methods to the parametric F-ratio test procedure when testing main effects and interaction in factorial and split-unit experiments.

2. Estimating Exact Distributions

For complex designs with large sample sizes, the exact distribution of the test statistic will be estimated based on a random sample of all possible permutations of the data. This method was first proposed by Dwass (1957) as "the most logical" way to obtain an approximation to Fisher's method of randomization, and tests based on this method of determining significance have become known as "Randomization Tests" (Manly, 1991 ; Edgington, 1995). This technique, when used on the actual observations, has the somewhat undesirable property that a possibly unique sampling distribution must be constructed for each set of data. In addition, two researchers performing a randomization test independently on the same set of data would likely obtain slightly different p-values. For a large (say 20,000) random sample of permutations, however, it is unlikely that two independent tests would arrive at different conclusions regarding significance. For example, for estimating the cumulative probability associated with the 95^{th} percentile of a sampling distribution based on a random sample of 20,000 permutations, the expected error of estimation, with 99% confidence, would be about .004, or .4%. Thus, very precise estimates of the exact critical values of the sampling distribution can be attained. Applied to rank transformed data, however, a unique sampling distribution would need to be derived only for each possible sample size. Thus, it is possible to create tables of critical values, given a particular sample size.

3. Simulation Study for a Completely Randomized Two-Factor Factorial Experiment

3.1. Procedure.

Simulated data sets were generated to examine the performance of the three methods: the parametric F-test procedure (FT), the exact rank transform test procedure (RT), and the exact aligned rank transform test procedure (ART). The following model was used to generate the observations:

$$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + e_{ijk},$$

where A_i is the effect of the i[th] level of treatment A, i=1,2,3,4; B_j is the effect of the j[th] level of treatment B, j=1,2,3; AB_{ij} is the effect of the interaction between the i[th] level of factor A and the j[th] level of factor B, and e_{ijk} is the random error effect, k=1,2. For the ART, observations were aligned in the following manner: when testing interaction, an aligned observation was $AY_{ijk}=Y_{ijk}$- (sample mean)$_i$ - (sample mean)$_j$; when testing for main effects, an aligned observation for testing effect A was $Ay_{ijk}=Y_{ijk}$ - (sample mean)$_j$, and for testing effect B the aligned observation was $Ay_{ijk}=Y_{ijk}$ - (sample mean)$_i$. Standard normal and exponential (μ=3) distributions were used to model the error distributions. Effect sizes (denoted by "c" in the tabled results) are in standard deviation units, and range in magnitude from 0.5 (very small) to 3.5 (very large). Critical values for both rank tests were estimated by calculating the value of the test statistic for a random sample of twenty thousand permutations of the ranks of the data. Ten thousand samples were generated, and the proportion of test statistic values greater than or equal to the critical values for the respective sampling distributions was calculated. Thus, for estimating a nominal type I error rate of 0.05, the maximum error of estimation is 0.0056, with 99% confidence (values outside of this range are in bold in the tables that follow).

3.2 Results.

Normally distributed errors, equal variances (see Tables 1 & 2): The ART consistently showed power almost equal to that of the F-test. The ART often had slightly inflated nominal type I error rates, but the inflation was never severe, and did not appear to be affected by the magnitude of the modeled effects. The RT tended to compare favorably in most cases, but showed poor power when both main effects and interaction were present in the model, especially for testing interaction. In addition, for all models the RT had nominal type I error rates that inflated as the magnitude of the effects increased. For a more detailed study of the performance of the RT when the parametric assumptions are satisfied, see Blair et al. (1987).

Exponentially distributed errors (see Tables 3 & 4): Both rank tests had superior power to the F-test. A notable exception was the model which had both main effects and interaction present, where again the RT had less power for testing interaction than in other models. Even though for most models the power of the RT was about the same as the FT (except when effect magnitudes became very large, where the FT usually had more power), it was still outperformed by the ART. Interestingly, for small sample sizes (n=2 observations per cell), when the error distributions were non-normal, the nominal type I error rates for the RT did not show a tendency to inflate as the magnitudes of the effects increased.

Normally distributed errors, unequal variances (see tables 5 & 6): This was a much more serious problem than the lack of normality. The power for all methods was less than in the equal variance case, and this decrease in power became more severe as the degree of heterogeneity between variances increased. However, both rank tests consistently outperformed the FT in the power category, except for the RT in the previously discussed model. The FT did, however,

Kansas State University

often have slightly higher power for very small effect magnitudes. In addition, the ART usually had more power for testing interaction than the RT. Examination of nominal type I error rates for testing interaction when none was modeled revealed that these rates were inflated for all three methods, with more severe inflation occurring when the variances were more unequal. This indicated that variance heterogeneity actually tended to be falsely interpreted as interaction more often than would be expected. The ART seemed to be the most sensitive to this false interaction, which is not surprising since the alignment procedure isolates the effect of interaction, followed by the FT and then the RT. Thus, it is not surprising that the ART showed more power when interaction was actually modeled. The RT was the least sensitive to the presence of interaction.

The problem of nominal type I error rate inflation was not limited only to the test for interaction, however. When only one main effect was modeled along with an interaction effect, the nominal type I error rates for testing the unmodeled main effect were also inflated for all methods. Thus, it is apparent that variance heterogeneity can produce very erratic behavior in the analysis.

4. SIMULATION STUDY FOR A SPLIT-UNIT EXPERIMENT

4.1. Procedure

Simulated data sets were generated to examine the performance of the three methods: the parametric F-test procedure (FT), the exact rank transform test procedure (RT), and the exact aligned rank transform test procedure (ART). A split-unit experiment with main units in a randomized complete block design was considered. The following model was used to generate the observations:

$$Y_{ijk} = B_i + M_j + BM_{ij} + S_k + SM_{jk} + E_{ijk},$$

where B_i is the random effect of the i^{th} block, i=1,2,3; M_j is the fixed effect of the j^{th} level of the main unit treatment, j=1,2,3,4; BM_{ij} is the random effect of the interaction between the i^{th} block and the j^{th} level of the main unit treatment, S_k is the fixed effect of the k^{th} level of the sub-unit treatment, k=1,2,3; SM_{jk} is the fixed effect of the interaction between the j^{th} level of the sub-unit treatment with the k^{th} level of the main unit treatment, and E_{ijk} is the random sub-unit error effect. The random effect BM_{ij} was used as error to test for the effect of the main unit treatment, while the random effect E_{ijk} was used as error to test both the sub-unit treatment effect, S_k, and the interaction effect, SM_{jk}. Standard normal (both with homogeneous and heterogeneous variances), exponential (μ=3) and uniform [-3,3] distributions were used to model the error distributions. Ten thousand samples were generated, and the proportion of test statistic values greater than or equal to the critical values for the respective sampling distributions was calculated.

For the aligned rank procedure, three different methods of aligning were used, depending upon the effect being tested. For testing main unit treatment effect, the observations were aligned by subtracting estimates of both block and sub-unit treatment effects. For testing sub-unit treatment effect, estimates of both block and main unit treatment effects were subtracted from each observation. Finally, for testing interaction, the observations were aligned by subtracting block, main unit and sub-unit effect estimates.

4.2. Results

Normally distributed main-unit and sub-unit errors (see Tables 7 & 8): In this situation, all random effects were modeled as identically distributed standard normal distributions. The three methods performed almost identically to the previous study of the two-way layout in a completely randomized design. Both rank tests consistently had power almost equal to that of the F-test. As in the completely randomized case, the RT again showed poor power for testing interaction when both main and sub-unit main effects and interaction were present in the model. When only main and sub-unit effects were in the model, the RT again had type I error rates that inflated as the magnitude of the effects increased. This behavior was not as evident for other models, however.

Exponentially distributed errors (see Tables 9 & 10): When the sub-unit error effect was exponentially distributed, both rank tests had more power than the F-test for all models. When all fixed effects were in the model, the power of the ART was clearly superior to the other two, although the drop-off in power for the RT was not as severe as had been observed in previous situations.

Heterogeneous errors (see Tables 11-14): Two cases were considered. One of the errors was modeled as normally distributed with heterogeneous variances, while the other was modeled as normally distributed with homogeneous variances. In each case, the block effect was modeled as having a standard normal distribution. For all models, a ratio between the largest and the smallest variances of 30:1 (very large) was considered. As in the completely randomized case, unequal error variances turned out to be a more serious problem than the lack of normality. However, while in the completely randomized case, the performance of the rank tests was generally better than that of the F-test, in the split-unit case the results were mixed.

The power of all tests was lower when the main units had heterogeneous variances, and the power became worse as the degree of the heterogeneity increased. When only main unit and sub-unit treatment effects were present, the rank tests had better power for testing for main unit treatment effect, but slightly less power for testing for sub-unit treatment effect. In addition, the RT had nominal type I error rates that increased steadily with increasing effect magnitudes. When all effects were present, the FT had the best power, with the ART close behind and the RT a distant third.

The rank tests performed consistently better than the FT when then sub-unit error effect had unequal variances. When the ratio of largest to smallest variance was 30:1, the rank tests had more power. For all methods, there was also a slight nominal type I error rate inflation for testing the interaction effect, which became more severe as the variance ratio increased. Surprisingly, the RT showed less inflation than the either the FT or the ART. When only both main and sub-unit effects were modeled, the rank tests were much more powerful, with some nominal type I error rate inflation for testing interaction evident for all methods. However, while the FT and the ART nominal rates remained constant as the magnitude of the effects increased, the RT showed its familiar inflation as an increasing function of effect magnitude. When all fixed effects were in the model, the ART had much more power than the other two methods for testing interaction.

Kansas State University

Investigation of the nominal type I error rates when the main or sub-unit variances were unequal revealed a problem of inflated nominal type I error rates similar to that of the completely randomized experiment (see Tables 11 & 13). When the main unit variances were heterogeneous, nominal type I error rates for testing the main unit treatment effect were often larger than expected. When the sub-unit variances were heterogeneous, nominal type I error rates for tests for sub-unit treatment and interaction effects were always inflated. However, heterogeneous main unit variances did not adversely affect the nominal levels of the sub-unit tests, and vice-versa. Once again, the inflation of the nominal rates for the RT was often a function of the magnitude of the modeled effects, while the inflation of the nominal rates for the FT and the ART seemed to be independent of the effect magnitude. Once more this indicates that when error variances are heterogeneous, test results may be misleading, especially when testing for interaction. This was not a problem when one of the underlying populations was skewed (exponentially distributed).

5. Conclusion and Summary

The exact aligned rank procedure appears to be the overall best choice for performing tests in a general factorial experiment. When the error distribution was symmetric and error variances were homogeneous, the ART was nearly as powerful as the F-test, with an almost negligible difference in power between the two methods. For a skewed error distribution, the ART was clearly more powerful than the F-test. When the error variances were heterogeneous, both methods had problems maintaining nominal type I error levels for testing interaction, but the ART showed superior power for detecting main effects and interaction.

Although the results were not as consistent as for the completely randomized case, the exact aligned rank procedure appears to be viable alternative to the normal theory F-test for performing tests in a split-unit factorial design, and is certainly a better choice than the rank transform method. Once more, when the error distributions were normal and error variances were homogeneous (situations in which the F-test is known to work well), the ART was always nearly as powerful, with usually an almost negligible difference in power between the two methods. For exponential error distributions, the ART was clearly more powerful than the F-test. Uniformly distributed errors were also examined for several models. The results were nearly identical to the case for normally distributed errors, with the F-test having the most power, followed closely by the ART and then the RT. The ART again often had slightly inflated nominal type I error rates for testing interaction. When the error variances were heterogeneous, both methods tended to have problems maintaining nominal type I error levels for interaction, although this problem was less severe in the split-unit case, while the ART usually had superior power for detecting main effects. Although the FT outperformed the ART in some cases, even when parametric assumptions were violated, the ART had superior power in most cases, and tended to enjoy a greater power advantage when it was the more powerful test, especially when the assumptions of normality and homogeneity of variance were violated. Even though the simulation results indicate that a nonexistent interaction effect can be introduced when error variances are unequal, this phenomenon occurs for both the FT and the ART. Since typically the analysis is performed without the benefit of definite knowledge of the nature of the error variances, and since the ART generally has more power than the FT when variances are unequal, the ART seems a logical choice over the FT.

One issue that deserves comment is the choice of estimator used for aligning observations. The mean was used in this study, but an argument could be made for using a more robust measure, especially when the error distribution is skewed. Higgins and Tashtoush (1994) examined the use of the trimmed mean and the median, but concluded that the in gain power did not necessarily override the greater ease of implementation of the procedure using the mean. Also, no matter which estimator of location is used, the performance of the test may be affected by the properties of that estimator for the underlying error distribution. This may explain, for example, the inflated type I error rates observed for samples from skewed distributions, where the mean is probably not the most robust measure of location.

The problems with the rank transform method in two-way factorial designs are not alleviated by using the exact permutation distribution of the test statistic computed on the ranks. Based upon the results of this and other studies, the rank transform procedure should not be used to analyze data in a factorial arrangement, due to the serious type I error rate inflations caused by the transformation of data to ranks, and also to the poor power exhibited for some models. This implies that the rank transform procedure should be avoided in any design that allows for interaction between factors, including split-unit experiments.

References

Akritas, M.G. (1990). The rank transform method in some two-factor designs. *Journal of the American Statistical Association, 85(409)*, 73-78.

Blair, R.C., Sawilowsky, S.S., & Higgins, J.J. (1987). Limitations of the rank transform statistic in tests of interaction. *Communications in Statistics: Computation and Simulation*, *B16,* 1133-1145.

Conover, W.J., & Iman, R.L. (1976). On some alternative procedures using ranks for the analysis of experimental designs. *Communications in Statistics, A5*, 1349-1368.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics, 28*, 181-187.

Edgington, E.S. (1995). *Randomization Tests, 3rd ed.* Marcel Dekker, New York.

Fawcett, R.F. & Salter, K.C. (1984). A Monte Carlo study of the F-test and three tests based on ranks of treatment effects in randomized block designs. *Communications in Statistics, B13,* 213-225.

Groggel, D.J. (1987). A Monte Carlo study of rank tests for block designs. *Communications in Statistics, 16*(3), 601-620.

Higgins, J.J., Blair, R.C. & Tashtoush, S. (1990). The aligned rank transform procedure. Proceedings of the 1990 Kansas State University Conference on Applied Statistics in Agriculture, 185-195.

Higgins, J.J. & Tashtoush, S. (1994). An aligned rank transform test for interaction. *Nonlinear World, 1*, 201-211.

Hodges, J.L., & Lehmann, E.L. (1962). Rank methods for combination of independent

Kansas State University

experiments in analysis of variance. *Annals of Mathematical Statistics, 27*, 324-335.

Manly, B.F.J. (1991). *Randomization and Monte Carlo Methods in Biology*. Chapman & Hall, New York.

Sawilowsky, S.S., Blair, R.C. & Higgins, J.J. (1989). An investigation of the type I error and power properties of the rank transform procedure in factorial ANOVA. *Journal of Educational Statistics, 14(3)*, 255-267.

Thompson, G.L., & Ammann, L.P. (1990). Efficiencies of interblock rank statistics for repeated measures designs. *Journal of the American Statistical Association, 85(410),* 519-528.

Tables

Table 1. Proportion of rejections at $\alpha = .05$, normally distributed errors with equal variance. A and B main effects present (a2=b1=c, a3=b2= -c).

n = 2

Test for:	Statistic	c = 0.5	1.5	2.5	3.5
Factor A	F	.210	.968	1.00	1.00
	FR	.199	.942	1.00	1.00
	FAR	.199	.959	1.00	1.00
Factor B	F	.329	.999	1.00	1.00
	FR	.317	.996	1.00	1.00
	FAR	.319	.998	1.00	1.00
Interaction	F	.050	.050	.050	.050
	FR	.054	.054	.054	**.068**
	FAR	.056	.056	.056	.056

n = 10

Test for:	Statistic	c = 0.5	1.5	2.5	3.5
Interaction	F	.049	.049	.049	.049
	FR	.051	**.134**	**.671**	**.997**
	FAR	.050	.050	.050	.050

Table 2. Proportion of rejections at $\alpha = .05$, normally distributed errors with equal variance. A, B and interaction effects present (ab11=c, b1=ab41= -c).

n = 2

Test for:	Statistic	c 0.5	1.5	2.5	3.5
Factor A	F	.066	.213	.527	.830
	FR	.066	.132	.193	.218
	FAR	.065	.153	.252	.290
Factor B	F	.139	.780	.997	1.00
	FR	.134	.652	.940	.994
	FAR	.140	.732	.989	1.00
Interaction	F	.069	.260	.655	.931
	FR	.066	.153	.230	.264
	FAR	.075	.251	.617	.909

Table 3. Proportion of rejections at $\alpha = .05$, identically exponentially distributed errors. A and B main effects present (a2=b1=c, a3=b2=-c).

n = 2

Test for:	Statistic	c 0.5	1.5	2.5	3.5
Factor A	F	.066	.246	.574	.828
	FR	.083	.314	.621	.834
	FAR	.086	.335	.665	.877
Factor B	F	.084	.386	.762	.943
	FR	.119	.497	.825	.956
	FAR	.113	.485	.839	.966
Interaction	F	.055	.055	.055	.055
	FR	**.058**	**.059**	**.059**	**.057**
	FAR	**.074**	**.074**	**.074**	**.074**

n = 10

Test for:	Statistic	c 0.5	1.5	2.5	3.5
Factor A	F	.172	.898	1.00	1.00
	FR	.329	.985	1.00	1.00
	FAR	.332	.993	1.00	1.00
Factor B	F	.251	.977	1.00	1.00
	FR	.477	.999	1.00	1.00
	FAR	.463	1.00	1.00	1.00
Interaction	F	.048	.048	.048	.048
	FR	.053	**.060**	**.078**	**.121**
	FAR	**.061**	**.061**	**.061**	**.061**

Table 4. Proportion of rejections at α = .05, identically exponentially distributed errors. A, B and interaction effects present (ab11=c, b1=ab41=-c).

n = 2			c			
	Test for:	Statistic	0.5	1.5	2.5	3.5
	Factor A	F	.049	.063	.097	.154
		FR	.054	.073	.094	.121
		FAR	.057	.080	.113	.151
	Factor B	F	.057	.155	.362	.610
		FR	.073	.224	.405	.576
		FAR	.072	.208	.420	.634
	Interaction	F	.058	.075	.113	.186
		FR	.059	.082	.109	.142
		FAR	.076	.100	.153	.234
n=10	Test for:	Statistic	0.5	1.5	2.5	3.5
	Factor A	F	.059	.167	.412	.707
		FR	.077	.238	.443	.616
		FAR	.075	.268	.549	.774
	Factor B	F	.113	.638	.961	1.00
		FR	.200	.832	.986	1.00
		FAR	.185	.841	.992	1.00
	Interaction	F	.065	.227	.592	.891
		FR	.089	.335	.634	.836
		FAR	.091	.412	.846	.984

Table 5. Proportion of rejections at α = .05, normally distributed errors with unequal variance. A and B main effects present (a2=b1=c, a3=b2= -c).

n = 2			c			
(30:1 ratio)	Test for:	Statistic	0.5	1.5	2.5	3.5
	Factor A	F	.108	.218	.475	.753
		FR	.096	.280	.562	.802
		FAR	.097	.279	.613	.874
	Factor B	F	.108	.313	.651	.887
		FR	.102	.380	.718	.914
		FAR	.105	.406	.757	.945
	Interaction	F	**.113**	**.113**	**.113**	**.113**
		FR	**.080**	**.099**	**.110**	**.111**
		FAR	**.134**	**.134**	**.134**	**.134**

Table 6. Proportion of rejections at $\alpha = .05$, normally distributed errors with unequal variance. A, B and interaction effects present (ab11=c, b1=ab41= -c).

n = 2			c			
(30:1 ratio)	Test for:	Statistic	0.5	1.5	2.5	3.5
	Factor A	F	.097	.110	.132	.167
		FR	.076	.090	.115	.146
		FAR	.082	.093	.118	.144
	Factor B	F	.090	.160	.291	.481
		FR	.075	.144	.275	.455
		FAR	.075	.153	.302	.506
	Interaction	F	.117	.132	.164	.211
		FR	.078	.086	.110	.140
		FAR	.135	.157	.193	.248

Table 7. Proportion of rejections at $\alpha = .05$, normally distributed errors with equal variance. MU and SU main effects present (m2=s1=c, m3=s2= -c).

		c			
Test for:	Statistic	0.5	1.5	2.5	3.5
MU Trt	F	.088	.474	.900	.994
	FR	.091	.467	.889	.993
	FAR	.096	.481	.897	.993
SU Trt	F	.500	1.00	1.00	1.00
	FR	.449	1.00	1.00	1.00
	FAR	.473	1.00	1.00	1.00
Interaction	F	.049	.049	.049	.049
	FR	.046	.047	**.077**	**.148**
	FAR	.049	.049	.049	.049

Kansas State University

Table 8. Proportion of rejections at α = .05, normally distributed errors with equal variance. MU, SU main effects and interaction effect present (ms11=-c, s1=ms41=c).

		c			
Test for:	Statistic	0.5	1.5	2.5	3.5
MU Trt	F	.052	.087	.168	.298
	FR	.057	.078	.114	.146
	FAR	.058	.087	.123	.155
SU Trt	F	.187	.942	1.00	1.00
	FR	.168	.875	.998	1.00
	FAR	.179	.911	1.00	1.00
Interaction	F	.079	.416	.894	.997
	FR	.070	.269	.497	.642
	FAR	.075	.383	.850	.991

Table 9. Proportion of rejections at α = .05, exponentially distributed sub-unit errors, normally distributed block effect and main unit errors. MU and SU main effects present (m2=s1=c, m3=s2= -c).

		c			
Test for:	Statistic	0.5	1.5	2.5	3.5
MU Trt	F	.066	.198	.470	.748
	FR	.074	.234	.513	.770
	FAR	.074	.240	.542	.801
SU Trt	F	.095	.543	.909	.989
	FR	.126	.657	.948	.996
	FAR	.125	.655	.952	.997
Interaction	F	.044	.044	.044	.044
	FR	.049	.049	.049	.055
	FAR	**.058**	**.058**	**.058**	**.058**

Table 10. Proportion of rejections at α = .05, exponentially distributed sub-unit errors, normally distributed block effect and main unit errors. MU, SU main effects and interaction effect present (ms11=-c, s1=ms41=c).

		c			
Test for:	Statistic	0.5	1.5	2.5	3.0
MU Trt	F	.054	.068	.096	.138
	FR	.055	.070	.094	.120
	FAR	.056	.074	.098	.132
SU Trt	F	.061	.220	.518	.778
	FR	.076	.282	.574	.778
	FAR	.076	.274	.582	.805
Interaction	F	.050	.080	.160	.288
	FR	.055	.094	.155	.227
	FAR	.064	.105	.198	.345

Table 11. Proportion of rejections at α=0.05, normally distributed errors, unequal main unit error variances. Ratio largest to smallest variance 30:1. MU and SU main effects present (m2=s1=c, m3=s2= -c).

		c				
Test for:	Statistic	0.0	0.5	1.5	2.5	3.5
MU Trt	F	**.083**	.088	.130	.223	.366
	FR	**.090**	.095	.151	.257	.405
	FAR	**.084**	.090	.142	.258	.407
SU Trt	F	.050	.509	1.00	1.00	1.00
	FR	.056	.422	1.00	1.00	1.00
	FAR	.050	.440	1.00	1.00	1.00
Interaction	F	.052	.052	.052	.052	.052
	FR	.051	.057	**.080**	**.107**	**.120**
	FAR	.050	.050	.050	.050	.050

Kansas State University

Table 12. Proportion of rejections at α=0.05, normally distributed errors, unequal main unit error variances. Ratio largest to smallest variance 30:1. MU, SU main effects and interaction effect present (ms11=-c, s1=ms41=c).

Test for:	Statistic	c 0.5	1.5	2.5	3.5
MU Trt	F	.084	.088	.097	.109
	FR	.091	.092	.094	.101
	FAR	.085	.087	.094	.103
SU Trt	F	.194	.936	1.00	1.00
	FR	.133	.691	.969	1.00
	FAR	.144	.777	.991	1.00
Interaction	F	.082	.421	.890	.996
	FR	.067	.152	.302	.458
	FAR	.070	.307	.735	.947

Table 13. Proportion of rejections at α=0.05, normally distributed errors, unequal sub-unit error variances. Ratio largest to smallest variance 30:1. MU and SU main effects present (m2=s1=c, m3=s2= -c).

Test for:	Statistic	c 0.0	0.5	1.5	2.5	3.5
MU Trt	F	.052	.063	.155	.350	.619
	FR	.055	.070	.184	.389	.625
	FAR	.052	.067	.191	.437	.701
SU Trt	F	**.074**	.095	.411	.911	.999
	FR	**.073**	.131	.666	.985	1.00
	FAR	**.068**	.114	.636	.984	1.00
Interaction	F	**.083**	**.083**	**.083**	**.083**	**.083**
	FR	**.065**	**.065**	**.074**	**.081**	**.083**
	FAR	**.105**	**.105**	**.105**	**.105**	**.105**

Table 14. Proportion of rejections at α=0.05, normally distributed errors, unequal sub-unit error variances. Ratio largest to smallest variance 30:1. MU, SU main effects and interaction effect present (ms11=-c, s1=ms41=c).

		c			
Test for:	Statistic	0.5	1.5	2.5	3.5
MU Trt	F	.053	.059	.079	.111
	FR	.057	.075	.095	.122
	FAR	.054	.073	.101	.135
SU Trt	F	.081	.159	.370	.682
	FR	.090	.240	.537	.816
	FAR	.078	.210	.510	.814
Interaction	F	.085	.102	.143	.219
	FR	.070	.107	.170	.242
	FAR	.108	.135	.193	.294