

Abstract

When utilizing a general two-way layout, an investigator is often interested in whether or not the two factors interact. If interaction is not present, then the investigator can explore the main effects of the factors separately. If, however, interaction is present, investigation of the main effects may not be important, and the investigator will be more interested in examining the nature of the interaction. In this project, we developed several nonparametric test statistics for testing for interaction in the two-way layout with multiple replications per cell. Null distributions were derived for these statistics, and then the proposed statistics were compared with existing procedures through Monte Carlo power studies.

Introduction

When there are multiple replications per cell, the general model for the two-way layout is

$$Y_{ijk} = \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad \text{and} \quad k = 1, \dots, K,$$

where two factors U and V , having I and J levels respectively, are being investigated and K is the number of replications per cell. The error terms, denoted by the ε_{ijk} 's, are assumed to be independent and identically distributed with common median θ . α_i is the effect of the i^{th} level of factor U , β_j is the effect of the j^{th} level of factor V , and γ_{ij} is the effect of the interaction between the i^{th} level of factor U and the j^{th} level of factor V . Thus, we are considering the general two-way layout with an equal number of replications per cell.

Statistics for Detecting Interaction

Recent research by Hartlaub, Dean, and Wolfe (1999) resulted in the development of procedures to test for interaction in the two-way layout with one observation per cell. An invariance problem with their statistics was solved by Lehman, Wolfe, Dean, and Hartlaub (2001) who proposed symmetrized procedures. The symmetrized procedures, S-SA (symmetrized statistics aligned by averages) and S-SM (symmetrized statistics aligned by medians) are based on the statistics CRA and RCA, and CRM and RCM, respectively (from Hartlaub, Dean, and Wolfe (1999)). S-SA and S-SM performed well in the one replication per cell setting. Thus, we propose three extensions of these statistics which utilize the technique of crossed comparisons (see Tukey 1991) to detect interaction in the multiple replications per cell setting. Our proposed statistics utilize a common idea for eliminating nuisance effects (common methods that were used in previous research), aligning with averages or medians to remove one of the nuisance effects (row or column) and ranking within the columns or rows to remove the other.

Our first extension, CASSA, involves computing cell averages and applying the existing procedure S-SA on those cell averages. Alternatively, S-SM can be applied to the cell medians to yield CMSSM. The second extension, SSACA, is based on aligning with averages, ranking within rows or columns, calculating the average rank for each cell, and applying S-SA on those average ranks. Similarly, cell medians can be used to align the data, and median ranks for each cell can be used with existing one observation per cell procedures to yield SSMCM. The final extension, APCSSAD and APCSSMD utilize *all possible comparisons* (APC) along with division by a constant (D) for scaling purposes.

Procedure: APCSSAD

Step 1. Calculate APCCRA. Align within the columns using column averages, and then rank within the rows. Now compute the $J(J-1)/2$ crossed comparisons denoted V_{jj}' .

$$V_{jj}' = \sum_{1 \leq i < i' \leq I} \sum_{k_1=1}^K \sum_{k_2=1}^K \sum_{k_3=1}^K \sum_{k_4=1}^K \{(r_{ijk_1} + r_{i'j'k_2}) - (r_{i'jk_3} + r_{ij'k_4})\}^2.$$

APCCRA is the maximum of the V_{jj}' 's.

Step 2. Calculate APCCRAD. Divide APCCRA by $K^4 I(I-1)/2$, the number of summands in APC-CRA, to obtain a scaled version of the crossed comparisons for the maximum column comparison. That is,

$$APCCRAD = \frac{2 * APCCRA}{K^4 I(I - 1)}.$$

Step 3. Calculate APCRCA. Repeat Step 1, with alignment in the rows and ranking in the columns. APCRCA is computed by taking the maximum of $I(I - 1)/2$ possible row comparisons.

Step 4. Calculate APCRCAD. APCRCAD is computed by dividing APCRCA by $K^4 J(J - 1)/2$.

Step 5. Standardization. APCCRAD and APCRCAD are standardized by subtracting the appropriate null mean and dividing by the appropriate null standard deviation. Find

$$APCCRAD^* = \frac{APCCRAD - E_0(APCCRAD)}{\sqrt{V_0(APCCRAD)}} \text{ and } APCRCAD^* = \frac{APCRCAD - E_0(APCRCAD)}{\sqrt{V_0(APCRCAD)}},$$

where $E_0(APCCRAD)$ and $E_0(APCRCAD)$ are the null means of APCCRAD and APCRCAD respectively, and $V_0(APCCRAD)$ and $V_0(APCRCAD)$ are the null variances of APCCRAD and APCRCAD respectively.

Step 6. Calculate APCSSAD. APCSSAD is the maximum of $APCCRAD^*$ and $APCRCAD^*$.

Alternatively, using medians to align the data instead of averages in Procedure APCSSAD yields APCSSMD.

Our procedures were compared with three existing statistics: the F test, adjusted rank transform (RT), and De Kroon and Van Der Laan's statistic (DEKR). During the ranking process, the common ties correction of using average ranks was employed for all of the new procedures, DEKR, and RT.

Null Distribution Derivation, Monte Carlo Simulation, and Results

The null distributions for each of the three settings we examined, the $3*3*3$, $4*8*2$ and $5*6*4$, were derived for all statistics except the F statistic. For each setting, the statistics were compared with error terms drawn from the Normal (0,1), Uniform (-2,2), Exponential (1), or Cauchy (0,1) distributions. Monte Carlo simulation was used to check the null distribution critical values and to perform power comparisons for our nine statistics at the .05 alpha level. Multiple main effects were chosen for each factor, and three different types of interaction, product interaction, specific interaction, and Martin's interaction were studied.

Graphical displays were used to compare the rejection rates for each setting as Factor U and V effects change. The figures at the end of this report show some simulation results for specific interaction II, where the cells in row 1, column 1, and row 2, column 2 each had 2 added to all replications, while the cells in row 1, column 2, and row 2, column 1 each had 2 subtracted from all replications. In Figure 1, APCSSMD clearly is the best statistic. Procedures utilizing the median, such as APCSSMD, typically perform better in the Cauchy setting due to the heavy tails of the distribution. DEKR is unstable (sensitive to row effects), and the F statistic performs poorly. Previous studies have also noted these problems with DEKR and the F statistic.

In Figure 2, the F statistic performs best, as expected with normal error terms, but not much power is lost when using either APCSSAD or SSACA. Procedures utilizing averages tend to perform well with normal error terms. The RT and all other statistics are clearly inferior in this case. These graphs highlight some of the results we found, many of which were expected based on previous work. Notably though, the power estimates of our new procedures APCSSAD, APCSSMD, SSACA, and SSMCM, clearly indicate that further research may be worthwhile.

Suggestions for Future Research

Power comparisons for settings with more replications per cell should be made. Another form of RT based on $Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$ should be compared to the form of the RT studied here, which was based on $Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.}$. An investigation into the general two-way layout with an unequal number of replications per cell would be useful. Finally, block designs with a small number of treatment factors and a large number of blocks should be considered.

Conclusion

Based on the results of our power comparisons, several statistics can be immediately eliminated from use to detect interaction in the multiple replications per cell setting. CASSA and CMSSM fail to compete with the other statistics in detecting interaction. As seen in previous studies, DEKR suffers from the introduction of row effects and is thus not recommended for use in the multiple replications per cell setting. Other authors have found that the RT should not be used to detect general interaction and we concur based on our findings.

APCSSAD and APCSSMD perform marginally better than SSACA and SSMCM in the settings with Exponential and Cauchy error terms respectively, and these four statistics clearly are the best in these settings. When using APCSSAD or SSACA, not much power is lost compared to the F test when error terms are Normal or Uniform. In conclusion, since APCSSAD and APCSSMD do perform slightly better than SSACA and SSMCM, they are most strongly recommended for use in detecting interaction.

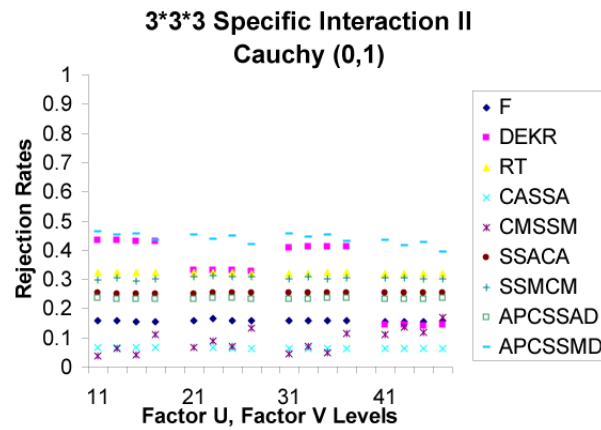


Figure 1:

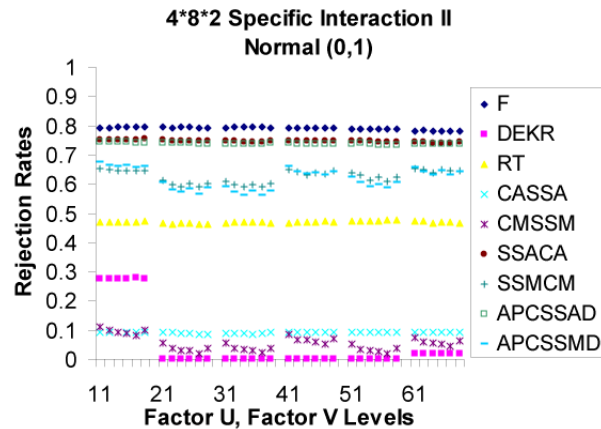


Figure 2: