Kansas State University Libraries

# New Prairie Press

Conference on Applied Statistics in Agriculture        2016 - 28th Annual Conference Proceedings

# PAIRED COMPETITION ANALYSIS USING MIXED MODELS

Patrick Gallagher
*Purdue University*, gallaghp@purdue.edu

Bruce A. Craig
*Purdue University*, bacraig@purdue.edu

Tim Luttermoser
*Purdue Univeristy*, tlutterm@purdue.edu

Grzegorz Buczkowski
*Purdue University*, gbuczkow@purdue.edu

*See next page for additional authors*

Follow this and additional works at: https://newprairiepress.org/agstatconference

Part of the Agriculture Commons, and the Applied Statistics Commons

## Recommended Citation

## Author Information

Patrick Gallagher, Bruce A. Craig, Tim Luttermoser, and Grzegorz Buczkowski

# PAIRED COMPETITION ANALYSIS USING MIXED MODELS

Patrick Gallagher[1], Bruce A Craig[1], Tim Luttermoser[2], and Grzegorz Buczkowski[2]

[1]Department of Statistics, Purdue University, [2]Department of Entomology, Purdue University

## *Abstract*

Urban and rural colonies of odorous house ants (*Tapinoma sessile*) have very different social structures. Urban colonies are very large with hundreds of cohabiting queens, while rural colonies are small with only one queen. To investigate whether worker ant aggressiveness varies across these two colony types, an experiment was performed using an aggression assay, in which 50 ants from each of two colonies were placed in a petri dish and allowed to fight. The response was the total number of dead ants within 24 hours. Because the ants were all the same species and not marked by colony, the number of dead ants per colony could not be determined.

A total of 138 colony pairings, involving six urban (U) and seven rural (R) colonies, were used in the experiment. Interest was in comparing the three types of pairings (UU, UR, and RR) to see if there is an ordering based on aggressiveness (e.g., $\mu_{UU} < \mu_{UR} < \mu_{RR}$). A linear mixed model is proposed to account for the fact that multiple assays involve ants from the same colony (i.e., to account for between colony variation in aggressiveness). However, the incorporation of random colony factors is not feasible for this study because of the UU and RR assays. As a result, we perform this mixed model analysis by specifying the covariance matrix (i.e., using the LINEAR covariance structure). A simulation study is used to assess the Type I and Type II errors of this mixed model approach relative to the standard one-way ANOVA. We conclude with an analysis of the real data set.

**KEY WORDS:** Aggression assay, Paired competition design, Mixed models, Linear covariance structure

## 1. Introduction

Odorous house ants (*Tapinoma sessile*) are a cosmopolitan North American ant species and a major urban pest (Buczkowski 2010). Throughout its range, *T. sessile* is known to contain a variety of social structures. Colonies in rural habitats, such as forests, tend to be small, containing only a single queen and a few hundred workers. In contrast, colonies in urban habitats can contain hundreds of queens and thousands of workers spread across multiple nesting sites connected by worker trails (Buczkowski 2010, Menke et al. 2010).

While similar social structures have been studied in several other ant species, the mechanisms driving this change in behavior is unknown for *T. sessile*. In other species with varying social structures, queen number and colony size have be associated with the level of aggression ants exhibit towards ants from foreign colonies of the same species (Crosland 1990, Stuart 1991, Fournier et al. 2016). This study was designed to investigate whether colonies of *T. sessile* from different habitats, and thereby different social structures, exhibit different levels of aggression against ants from foreign colonies of the same species.

These levels of aggression were assessed via a series of aggression assays (Roultson et al. 2003). The one we focus on is an assay where 50 worker ants from each of two colonies are placed in a petri dish and allowed to interact/fight. The response is the total number of dead ants after 24 hours, a proxy for the degree of aggressiveness between the two colonies. Only the total number of dead ants is observable because the ants are the same species and the ants from at least one colony were not uniquely marked.

This assay is typically performed many times using ants from different pairs of colonies. There are three types of pairings possible: rural/rural (RR), urban/urban (UU) and urban/rural (UR). It is predicted that, due to the greater number of queens present in urban colonies, urban colonies would be less aggressive than rural colonies towards odorous house ants from foreign colonies. In terms of the average number of dead ants per pairing type, the anticipated result is $\mu_{UU} < \mu_{UR} < \mu_{RR}$.

Prior to describing the real study, we first describe a general approach to analyze this type of experiment. This is followed by a simulation study to assess the power of this approach and compare it to the naïve one-way ANOVA. We then conclude with the analysis of the *T. sessile* study.

## 2. Analytic Approach

Consider a study in which there are *N* colonies of each habitat type and all $\binom{2N}{2}$ pairings of colonies are assayed. This results in two sets of $\binom{N}{2}$ pairings of colonies of the same type (RR and UU) and $N^2$ pairings of different types (UR). We will assume that these assays are run in a randomized order although various blocking schemes can be implemented.

Consider indexing the colonies such that the first $N$ represent the rural colonies and the last $N$ represent the urban colonies. Because worker ants from the same colony are likely to behave similarly (i.e., be more or less aggressive), we propose the following mixed model to account for this potential colony-to-colony variation

$$Y_{ij} = \mu + T_{[ij]} + C_i + C_j + \varepsilon_{ij} \tag{1}$$

where

$1 \le i < j \le 2N$

$T_{[ij]}$ is the pairing type effect for colony $i$ versus colony $j$

$C_k$ is the colony $k$ effect with $C_k \sim N(0, \sigma_c^2)$

$\varepsilon_{ij}$ represents experimental variation with $\varepsilon_{ij} \sim N(0, \sigma^2)$

Unfortunately, this model cannot be fit by specifying random colony effects because of the UU and RR pairings. As a result, we need to consider the marginal mixed model

$$Y_{ij} = \mu + T_{[ij]} + \delta_{ij} \tag{2}$$

where

$T_{[ij]}$ is the pairing type effect for colony $i$ versus colony $j$

$\delta_{ij}$ represents the combined variation with $\delta \sim N(0, \Sigma)$

### 2.1 Constructing the covariance matrix $\Sigma$

Based on our mixed model in (1),

$$\text{Cov}(Y_{ij}, Y_{kl}) = \begin{cases} \sigma^2 + 2\sigma_c^2 & \text{if } i = k \text{ and } j = l \\ \sigma_c^2 & \text{if } i = k \text{ or } j = l \\ 0 & \text{otherwise} \end{cases}$$

Thus, we can express the covariance matrix as a linear combination of two matrices. Specifically,

$$\Sigma = \sigma_c^2 M + \sigma^2 I \tag{3}$$

To aid in the general construction of $M$, we order the assays in a combinatorial fashion: (1, 2), (1, 3),…,(1, 2N), (2, 3), (2, 4),…, (2N-1, 2N). This results in recursive patterns of 0's, 1's, and 2's that we coded in R in order to construct $M$ for any size $N$. Figure 1 uses colors to identify these patterns for $N=4$.

In practice, an experiment may not have an equal number of colonies per habitat. The model in (2) does not change should there be missing pairs of colonies (by design or not) or multiple assays of the certain pairings. What does change is the covariance matrix $\Sigma$.

| | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 23 | 24 | 25 | 26 | 27 | 28 | 34 | 35 | 36 | 37 | 38 | 45 | 46 | 47 | 48 | 56 | 57 | 58 | 67 | 68 | 78 |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 12 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 16 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 17 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| 18 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| 23 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 26 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 27 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| 28 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| 34 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 35 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 36 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 2 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 37 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| 38 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| 45 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 46 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| 47 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 48 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 0 | 0 | 1 | 0 | 1 | 1 |
| 56 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 1 | 1 | 1 | 1 | 0 |
| 57 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 2 | 1 | 1 | 0 | 1 |
| 58 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 0 | 1 | 1 |
| 67 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 2 | 1 | 1 |
| 68 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 2 | 1 |
| 78 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 2 |

Figure 1: Matrix M for N=4.  The colors are used to highlight recursive patterns in the matrix of coefficients for $\sigma_c^2$. The first row and column indicate the two colonies in the pairing.

Our R code is designed to handle the more general experiment.  For example, for studies that involve a different number of colonies, say $N_1$ and $N_2$ of rural and urban, respectively, we construct an *M* matrix using $N = \max(N_1, N_2)$ and then delete those rows and columns that involve colonies that do not exist.  Likewise, if there are repeat assays of certain pairings, we simply replicate the row and column associated with that assay.

### 2.2 Fitting the marginal mixed model in SAS

To our knowledge, only SAS provides the option to fit a marginal mixed model with a linear structure covariance matrix.  Thus, we imported a .csv file containing the matrix *M* coefficients (see Appendix) and used the following SAS code to fit the model.

```
proc mixed;
class trt;
model y = trt/ ddfm=kr;
parms (1) (2);
repeated / subject=intercept type=lin(2) ldata=covmat;
```

## 3. Simulation Study

To better understand the importance of accounting for colony-to-colony variation, we performed a simulation study to compare our approach with a naive one-way ANOVA. We consider a complete experiment involving $N = 10$ colonies per type (i.e., 190 assays) under different degrees of colony variation and differences in means relative to the experimental variation $\sigma$. In particular, we considered the following scenarios

$$\sigma_c = k_1 \sigma$$
$$\mu_{RR} - \mu_{UR} = 0.5 k_2 \sigma$$
$$\mu_{UU} - \mu_{UR} = -0.5 k_2 \sigma$$

where $k_1$ and $k_2$ ranged between 0 and 2. For each combination of $k_1$ and $k_2$ we simulated 500 experiments using (1) and analyzed each using our approach and the standard one-way ANOVA.

For the power calculations, we also considered the SAS trick using Proc MIXED (Stroup, 1999). The little twist here is that our SAS code needs a denominator degrees of freedom adjustment in order function properly and the "ddfm=" option does not work in this trick. To get around this, we used a property of a complete paired competition experiment. Specifically, regardless of the values of $\sigma$ and $\sigma_c$, the Kenward-Roger adjustment always provides the same denominator degrees of freedom for the $F$ test. Thus, we plugged in these degrees of freedom without using "ddfm=" and the results agreed with our simulations.

### 3.1 Results

Figure 2 displays the probability of a Type I error (false positive) for both approaches. Our approach does an adequate job controlling this rate as it is either at or slightly above 0.05 for all values of $k_1$. The one-way ANOVA, on the other hand, does a very poor job for even a relatively
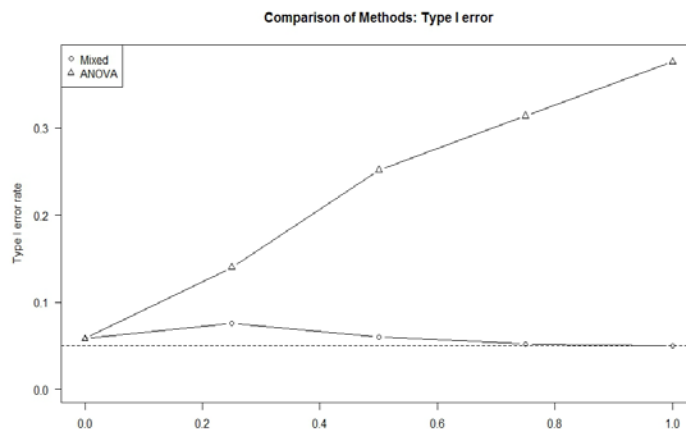


*Figure 2 : Comparison of Type I error rates. The mixed model approach controls the Type I error rate while the one-way ANOVA does a very poor job.*

low amount of colony variation.  Under the ANOVA model, the differences in the three pairing types are functions of the observed differences in average colony effects across the two types. The chance of a large observed difference in average colony effects increases with $\sigma_c$ as shown in the figure.

Figure 3 shows the power of our approach for different levels of colony variation.  For a relatively moderate colony standard deviation (0.4), the power is above 80% for a treatment difference of about 1.3 standard deviations.  For cases where the colony variance is similar to the experimental noise, the power is very low.  This could be increased with larger *N* or with repeated assays.
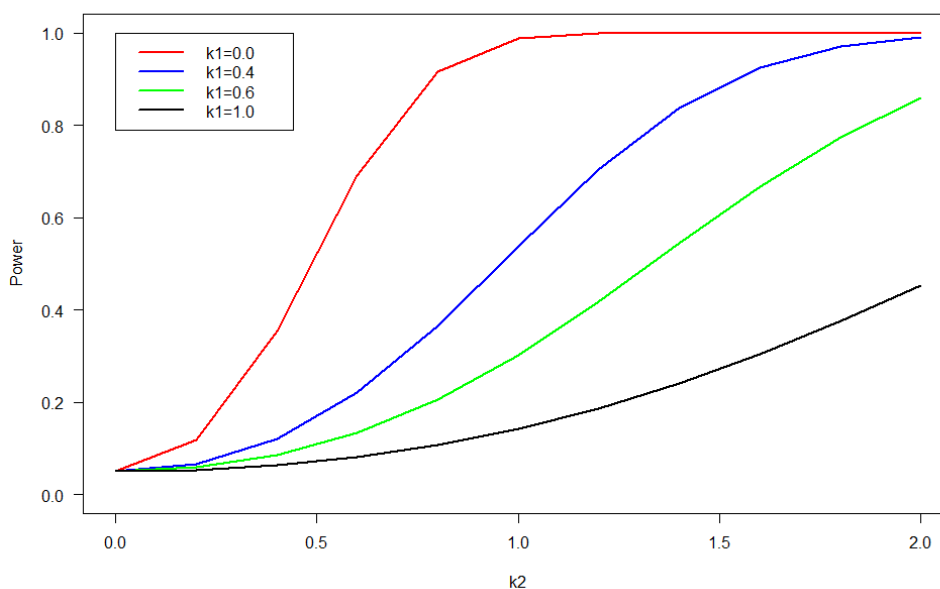


*Figure 3: Power curves for the mixed model.  For differences in means of one standard deviation, power ranges from almost 100% down to roughly 15% when $\sigma_c = \sigma$.*

## 4.  Real Study Analysis

For the real study, worker ants from six rural and seven urban colonies were assayed. Not all possible pairings were considered.  Furthermore, the pairings that were selected were replicated between three and five times for a total of 138 assays.

### 4.1 Results

The standard one-way ANOVA model resulted a highly significant *F* test (P=0.0003).  The estimated pairing type means were

$$\hat{\mu}_{UU} = 26.35, \hat{\mu}_{UR} = 24.72, \text{ and } \hat{\mu}_{RR} = 11.14.$$

For the mixed model, the *F* test was not significant (P=0.1745) and the estimated means were

$$\hat{\mu}_{UU} = 24.72, \hat{\mu}_{UR} = 19.45, \text{and } \hat{\mu}_{RR} = 7.82.$$

Both the log and square root transformations were considered but they did not alter the conclusions.

For the mixed model, the estimated colony variance was 108.8, which is slightly larger than the estimated experimental error 100.22. Assuming there are no differences across pairings, this combination of variances suggests the probability of a false positive using ANOVA near 40%. On the other hand, if there is a difference in groups of roughly twice the standard deviation ($\approx 20$), this combination of variances suggests the power is very low.

Perhaps more interesting than the difference in significance across is the fact that the estimated means are in the exact opposite direction of what was anticipated. To investigate this a bit further, Model (1) was also fit using WINBUGS. By taking a Bayesian approach, we were able to estimate the posterior probability that the means were in the anticipated ordering. This turned out to be 4.4%, suggesting aggression was not the likely driver for smaller single queen colonies. Further studies are being considered to investigate other mechanisms.

## 5. *Summary*

Paired competition assays are frequently utilized in the study of social insects. Our mixed model approach provides a general framework to analyze these studies when there is a repeated measures component to the study (i.e., multiple measurements are obtained for the same experimental unit). Our simulation study demonstrated the importance of accounting for this possible correlation. When ignored (i.e., analyzing the data using a one-way ANOVA), the false positive rate can approach 40%.

The power of a complete study depends heavily on the relative amount of colony to colony variability. When this is as large as the experimental variability, as it was with the real data study, either a very large number of colonies is needed or numerous replicates are required. Further investigations into the optimal designs for these types of studies is needed.

Although not the focus of this paper, we did consider a Bayesian analysis of the proposed mixed model. This approach not only allows us to estimate the probability of particular mean orderings but it also allows us to estimate the colony effects and check whether the assumption of these effects being Normally distributed is reasonable.

For the real data set, the estimated colony effects of the rural ants were at the extremes of the distribution. As a result, we did consider a more general model in which the colony variability differed across the two colony types. The approach is similar with the covariance matrix now a sum of three rather than two matrices. Although the general conclusions did not change, we are investigating the power of this model and comparing it to the model in which the colony variances are assumed equal.

## *References*

Buczkowski, G. (2010). Extreme life history plasticity and the evolution of invasive characteristics in a native ant. *Biological Invasions*, 12, 3343-3349.

Crosland, M.W.J. (1990). The influence of the queen, colony size, and worker ovarian development on nestmate recognition in the ant *Rhytidoponera confusa*. *Animal Behaviour*, 39(3), 413-425.

Fournier, D., de Biseau, J.-C., De Laet, S., Lenoir, A., Passera, L., and Aron, S. (2016). Social structure and genetic distance mediate nestmate recognition and aggressiveness in the facultative polygynous ant *Pheidole pallidula*. *PLoS ONE*, 11(5), e0156440.

Menke, S.B., Booth, W., Dunn, R.R., Schal, C., Vargo, E.L., and Silverman, J. (2010). Is it easy to be urban? Convergent success in urban habitats among lineages of a widespread native ant. *PLoS ONE*, 5(2): e9194.

Roulston, T., Buczkowski, G., and Silverman, J. (2003) Nestmate discrimination in ants: Effectof bioassay on aggressive behavior. *Insectes Sociaux*, 50(2), 151-159.

Stroup, W. W. (1999). Mixed model procedures to assess power, precision, and sample size in the design of experiments. In *Proceedings-Biopharmaceutical Section American Statistical Association* (pp. 15-24). American Statistical Association; 1997.

Stuart, R. J. (1991). Nestmate recognition in leptothoracine ants: testing for effects of queen number, colony size and species of intruder. *Animal Behaviour*, 42(2), 277-284.

## *Appendix*

Here is the code we used to read in the M matrix and prepare it for use in Proc MIXED. We used the numeric labeling of 4000, 30, and 0 to represent the three covariance values $2\sigma_c^2 + \sigma^2, \sigma_c^2$, and 0, respectively.

```
data covmat(drop=i c1-c190);
infile 'U:\Data\covmatrix.csv';
array columns{190} c1-c190;
array columns1{190} col1-col190;
input columns{*};
parm = 1;
row = _n_;
do i=1 to 190;
 columns1{i}=0;
 if columns{i}=4000 then columns1{i}=2;
 if columns{i}=30 then columns1{i}=1;
end;
output;
parm = 2;
do i=1 to 190;
 columns1{i}=0;
 if columns{i}=4000 then columns1{i}=1;
end;
output;
run;

proc sort data=covmat;
 by parm row;
run;
```