A Method for Analyzing Unreplicated Agricultural Experiments

Jamis J. Perrett* and James J. Higgins

ABSTRACT

Many studies are conducted in which replication of units is prohibitive. Traditional methods of hypothesis testing do not allow for analysis of unreplicated experiments. An abundance of subsampling allows for accurate estimation of within-treatment variance, but does not constitute experimental error. The intraclass correlation coefficient (ICC) is a ratio involving both the within-treatment variance and the between-treatment variance. We propose a method for analyzing unreplicated experiments that exploits the relationship among the ICC, within-treatment variance, and between-treatment variance by placing a reasonable upper bound on the ICC (from prior research) and using subsampling to carry out classical tests of significance that have conservative levels of significance. The methodology has wide applicability for analyzing unreplicated experiments and may be implemented in SAS (Cary, NC) using the MIXED procedure. As a demonstration of the methodology, the authors used data from an unpublished study in which a researcher tested the effectiveness of four different treatments [(i) control; (ii) sample-dependent release of the predacious phytoseiid mite (*Phytoseiulus persimilis* Athias-Henriot, PP); (iii) scheduled release of PP; (iv) Floramite {2-(4-methoxy-[1,1-biphenyl]-3-yl)-1-methylethyl ester; UniRoyal Chemical Company, Inc., Middlesbury, CT} pesticide application] controlling two-spotted spider mites (hereafter refered to as just *mites*) in commercial greenhouses. Four greenhouses were used for the study. Within each greenhouse, eight potted ivy geranium [*Pelargonium peltatum* (L.) L'Hér. ex Ait., 'Summer-Rose Red'] plants were inoculated with mites. One of the four treatments was applied in each of the four greenhouses. At the end of 1 wk, the number of mites was counted on each potted plant in each greenhouse. Using the proposed methodology, one-factor ANOVA was performed with follow-up tests identifying significant differences among the treatment means.

An unreplicated experiment is one in which a treatment of interest is applied to only one experimental unit. Some experiments logistically cannot be replicated. Circumstances that might prevent replication include cost in time or money or both, scarcity of experimental units, and destructive experimentation. For instance, some researchers just do not have an extra plot of land for experimentation. In the example presented in this paper, the researcher had to rely on the generosity of an owner of commercial greenhouses to conduct the study, and limitations meant there could be only one greenhouse per treatment. However, this experiment also had a condition common to many unreplicated experiments; namely, multiple observational units (potted plants) within each experimental unit (greenhouse). Other experiments of this type include: feeding treatments that are applied on a pen basis with weight gains that are measured on individual animals within the pens and irrigation treatments that are applied to large areas with yields that are measured on multiple plots within the areas.

ANALYSIS USING A KNOWN INTRACLASS CORRELATION COEFFICIENT

A researcher who uses variation among potted plants as the experimental error ignores the variability that can exist between different greenhouses receiving the same treatment. Such an assumption is to claim that $\rho = 0$, which can lead to a large inflation in the Type 1 error rate even if ρ is small (Barcikowski, 1981; Blair and Higgins, 1986). If the researcher correctly uses greenhouse as the experimental unit, there is only one experimental unit per treatment level and zero error degrees of freedom available for testing the differences among treatments using the conventional ANOVA. However, if the value of the ICC is known, then it is possible to analyze unreplicated data by plugging the known ICC into test statistic equations and basing error degrees of freedom on the amount of within-treatment replication (subsampling). Theoretical details of the methodology are discussed in Graybill (1976, p. 207–212); and analysis may be performed using SAS (sample code to be shown later).

As an illustration, consider the two-treatment case in which we test $H_0{:}\mu_1 = \mu_2$ vs. $H_0{:}\mu_1 \neq \mu_2$. The variance of the difference between the two sample means is given by

$$\begin{aligned}\text{var}\,(\bar{y}_{1\cdot} - \bar{y}_{2\cdot}) &= 2\sigma_\delta^2 + \sigma_\varepsilon^2\left[\frac{1}{n_1} + \frac{1}{n_{2\cdot}}\right] \\ &= \sigma_\varepsilon^2\left\{2\left(\frac{\rho}{1-\rho}\right) + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\} \quad [1]\end{aligned}$$

using the substitution $\sigma_\delta^2 = \sigma_\varepsilon^2\rho/(1-\rho)$. Thus, the variance of the differences between means can be written in terms of σ_ε^2 and ρ, excluding the need for an estimate of σ_δ^2.

This paper proposes the researcher determine a value for the ICC based on information about the current as well as similar prior replicated experiments, and plug that value into the equation for the test statistic in place of the unknown ICC.

MATERIALS AND METHODS

Study

In Spring 2004, a researcher grew eight cultivars of ivy geranium—Cajun Cranberry, Cajun White, Impulse Lilac Blue, Impulse Orange, Impulse Orange White, Shiva 2003, Summer-Rose Lilac, and Summer-Rose Red—in four commercial

J.J. Perrett, Program of Applied Statistics and Research Methods, Univ. of Northern Colorado, Greeley, CO 80639; J.J. Higgins, Dep. of Statistics, Kansas State Univ., Manhattan, KS 66506. Received 18 Apr. 2006. *Corresponding author (jamis.perrett@unco.edu).

Published in Crop Sci. 46:2482–2485 (2006).
Analysis of Unreplicated Experiments (Symposium)
doi:10.2135/cropsci2006.04.0255

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: ICC, intraclass correlation coefficient; PP, *Phytoseiulus persimilis.*

greenhouses in eastern Kansas. There were eight potted plants for each cultivar in each of the greenhouses. In March, the plants were inoculated with two-spotted spider mites. Two weeks after infestation, an initial count of mites was taken on the plants. The ICC was computed for each cultivar as the ratio of the variability among the greenhouses vs. the total variability among the counts for each potted plant for that cultivar. Those ICC values are recorded in Table 1. Subsequently, one of the following four treatments was applied to each of the four available greenhouses.

Treatment 1, control. No methods of pest control were performed.

Treatment 2, sample-determined release of biological control agent PP. Two weeks after plants were inoculated with mites, sampling was done on ivy geranium plants according to an existing sampling plan. Using the sampling plan, the number of mites per leaf was determined. The PP–mite ratio known to effectively control mites is 1:4. To determine the amount of PP required, the total number of mites in the area occupied by plants was divided by four—PP was released only once, 2 wk after plants were inoculated with mites.

Treatment 3, Scheduled release of biological control agent PP. Two weeks after plants were inoculated with mites, PP was released at a rate of 50 per square meter—a rate recommended by insectaries that sell natural enemies. Release of PP was done on a weekly basis for 4 wk.

Treatment 4. Application of a chemical insecticide. Two weeks after plants were inoculated with mites, a single chemical application was done (Floramite, UniRoyal Chemical Company, Inc., Middlesbury, CT). One week after the treatments were applied, counts were made of mites on the potted ivy geranium plants in each greenhouse. For simplicity, since the data from the study is only used as an illustration of the methodology, we will present only the analysis for Summer-Rose Red.

Model

Let y_{ij} be the measurement taken on the jth potted plant within the ith treatment (or greenhouse). Let μ_i denote the fixed effect of treatment i, δ_i the random effect of the greenhouse i, and ε_{ij} is the random effect of potted plant j given treatment i, $i = 1, 2, \ldots, t$; $j = 1, 2, \ldots, n_i$. Let $\delta_i \sim n(0,\sigma_\delta^2)$, where σ_δ^2 represents the between-greenhouse variability; let $\varepsilon_{ij} \sim n(0,\sigma_\varepsilon^2)$, where σ_ε^2 represents the between-potted-plant within-greenhouse variability. It is assumed that δ_i and ε_{ij} are independent. A model for the experiment is as follows:

$$y_{ij} = \mu_{ij} + \delta_i + \varepsilon_{ij} \quad [2]$$

This model is a single factor completely randomized design (CRD) with subsampling, where greenhouses are the experimental units with one greenhouse per treatment, and the potted plants within each greenhouse are the subsamples, or observational units.

Table 1. Estimated intraclass correlation coefficients for plant cultivars, 2 wk after being inoculated with two-spotted spider mites, representing the ratio of the variability among the greenhouses vs. the total variability among the counts for each potted plant for that cultivar. Counts were taken before application of any experimental treatments.

Cultivar	Estimated ρ†
Cajun Cranberry	0.07
Cajun White	0.22
Impulse Lilac Blue	0.00
Impulse Orange	0.16
Impulse Orange White	0.17
Shiva 2003	0.05
Summer-Rose Lilac	0.06
Summer-Rose Red	0.30

† ρ, population intraclass correlation coefficient.

The ICC is defined as the correlation between y_{ij} and y_{ij}' (two subsample units within one experimental unit). The ICC for Eq. [2] is as follows:

$$\rho = \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma_\varepsilon^2} \quad [3]$$

Thus, if $\sigma_\delta^2 = 0$, the result is independent observations assuming normality of error terms—a questionable, if not inaccurate presumption in many cases.

Bounding the Intraclass Correlation Coefficient

It is reasonable, in many cases, to place an upper bound on the ICC by considering the relative size of the between-unit variability to the within-unit variability, and by looking at past research or current data. In the example considered here, the greenhouses were similar, so it is reasonable to assume that the component of the variance due to greenhouses is relatively small. On the other hand, the component of variance due to differences among insect counts on plants within a greenhouse tends to be relatively large due to the unpredictability of insect behavior. Thus, it seems reasonable to place a bound on the ICC that is <0.5 and possibly a lot smaller than this. Data in Table 1 support the use of $\rho \leq 0.3$ as a reasonable upper bound for our example. We will denote the upper bound as ρ_{max} and use $\rho_{max} = 0.30$ in our analysis. Prior research on ICC values from previous, replicated studies might also have aided in the choice of ρ_{max} for this study.

Testing Strategy

Let ρ_0 denote a value of ρ the researcher assumes to be reasonable based on prior experience, and let ρ_0 denote the conditional P value given $\rho = \rho_0$. Let $\hat{\sigma}_\varepsilon^2$ be the estimate of σ_ε^2 based on the pooled within-treatments sample variances. Let $\mu_1 - \mu_2$ represent the hypothesized difference of the treatment means. The test statistic for the two-treatment hypotheses H_0:$\mu_1 = \mu_2$ vs. H_0:$\mu_1 \neq \mu_2$ with $n_1 + n_2 - 2$ degrees of freedom is as follows:

$$T = \frac{(\bar{y}_{1\cdot} - \bar{y}_{2\cdot}) - (\mu_1 - \mu_2)}{\sqrt{\hat{\sigma}_\varepsilon^2 \left\{ 2\left(\frac{\rho_0}{1-\rho_0} \right) + \left(\frac{n_1 + n_2}{n_1 n_2} \right) \right\}}} \quad [4]$$

The P value for such a test is denoted P_0. Perrett (2004) investigated various strategies for using the conditional P values in carrying out statistical tests. We recommend that tests based on P values, as well as confidence intervals and multiple comparisons be performed using $\rho_0 = \rho_{max}$. So, in the analysis of the greenhouse data, we will use $\rho_0 = \rho_{max} = 0.30$. The greater the value of ρ_{max} over the actual value of ρ, the less power the test will have. So, it is best that ρ_{max} be determined in a way that it exceeds ρ by as little as possible.

An informative way to display and interpret the results is to plot the P values vs. ρ_0, conditional on the value of ρ_0. Such a plot is nondecreasing and indicates at which value of ρ_0 the test results change from rejecting the null hypothesis to failing to reject the null hypothesis, for a given level α. Figure 1 is the conditional P value plot for the overall F test for this example. The researcher can look to the graph to see if the test results are uniform across an appropriate range for ρ_0. If so, the researcher can be confident in the results.

We are in no way suggesting that problems with lack of replication magically disappear with this methodology. All we

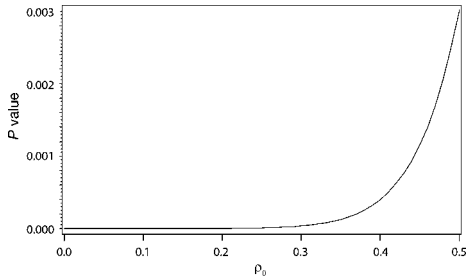

Fig. 1. Conditional *P* value plot: *P* value for testing equality of treatment means for different values of ρ_0 of the counts of two-spotted spider mites for each of four different treatment levels: (i) control, no mite-control treatment; (ii) sample, PP applied at 1:4 ratio according to sampling; (iii) schedule, PP was released at a rate of 50 per square meter; (iv) Floramite [UniRoyal Chemical Company, Inc., Middlesbury, CT), pesticide application] obtained 1 wk after the application of treatments designed to control the infestation of the mites in 'Summer-Rose Red' ivy geranium. Each of the four treatments was applied to all the potted plants in one of four commercial greenhouses.

are suggesting is a method for analysis when conventional analyses cannot be performed due to the lack of replication.

RESULTS

Analysis of Greenhouse Data

Figure 2 contains box plots representing the greenhouse data for the four different treatments. It is apparent from the plot a convincing difference exists between the count of mites in the greenhouse treated with Treatment 1 and those treated with Treatments 2 and 4. However, there is question about the difference between the mite counts under Treatment 3 vs. the others. Specifically, Treatments 2 and 4 appear to reduce significantly the count of mites. Treatment 3 is questionable. Both Bartlett's and Levene's tests of homogeneity

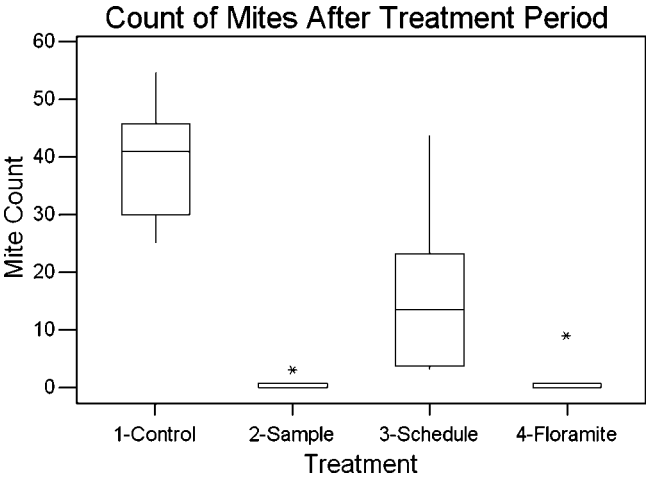

Fig. 2. Side-by-side box plots of the counts of two-spotted spider mites for each of four different treatment levels [control, no mite-control treatment; sample, PP applied at 1:4 ratio according to sampling; schedule, PP was released at a rate of 50 per square meter; Floramite (UniRoyal Chemical Company, Inc., Middlesbury, CT), miticide application] obtained 1 wk after the application of treatments designed to control the infestation of the mites in 'Summer-Rose Red' ivy geranium. Each of the four treatments was applied to all the potted plants in one of four commercial greenhouses.

of variance indicate unequal variances [$U(3) = 33.10$, $P = 0.000$; $F(3,24) = 4.38$, $P = 0.012$]. This leads to a square-root transformation on the response variable, mite count. After a square-root transformation was performed, neither Bartlett's nor Levene's test of homogeneity of variance indicated significant heterogeneity of variances ($U(3) = 6.73$, $P = 0.081$; $F(3,24) = 1.63$, $P = 0.205$). The following analyses use the transformed data rather than the original counts. The SAS code for the analyses are included in the appendix.

Figure 1 shows the conditional *P* value plot of the test H_0:$\mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. H_A: at least two means differ. The *P* value is <0.05 for every value of $\rho_0 < 0.5$. So, using our method for any value of $\rho_0 < 0.5$ leads to rejection of the null hypothesis.

Using $\rho_0 = \rho_{max} = 0.30$, the overall *F* test is significant [$F(3,28) = 11.95$, $P = 0.000$], indicating a difference in at least two of the treatment means. Treatment means were also compared using $\rho_0 = \rho_{max} = 0.30$. The results are listed in Table 2. The differences that are considered significant at $\alpha = 0.05$ using four multiple comparison methods include 1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 3 vs. 4.

DISCUSSION

Final Comments Regarding the Analysis

It is obvious from Fig. 2 that a difference exists between the means for the control and some of the treatment levels. However, without replication, standard methods do not allow for appropriate tests of significance. Using a plug-in value for ρ allows a valid analysis to be conducted. The conditional *P* value plot for the overall *F* test also provides clear support to the claim of difference in treatment means.

With the pairwise comparisons, again, Fig. 2 shows a clear drop in the number of mites on the potted plants treated with all three treatments vs. the control, with extreme drops with Treatments 2 and 4. There is a questionable difference between Treatments 2 and 4 vs. 3

Table 2. Differences of least squares means, unadjusted-, and multiple-testing-adjusted *P* values of the counts of two-spotted spider mites for each of four different treatment levels obtained 1 wk after the application of treatments designed to control the infestation of the mites in 'Summer-Rose Red' ivy geranium. Each of the four treatments was applied to all the potted plants in one of four commercial greenhouses.

Effect†	df	*t*	Unadjusted *P* value	Bonferonni-adjusted *P* value	Tukey-adjusted *P* value	Simulate-adjusted *P* value
T1 vs. T2	28	5.857	***	***	***	***
T1 vs. T3	28	2.590	*	ns‡	ns	ns
T1 vs. T4	28	5.699	***	***	***	***
T2 vs. T3	28	−3.267	*	ns	*	*
T2 vs. T4	28	−0.159	ns	ns	ns	ns
T3 vs. T4	28	3.108	*	ns	ns	ns

* Significant at the 0.05 probability level.
*** Significant at the 0.001 probability level.
† T1, control, no mite-control treatment; T2, sample, *Phytoseiulus persimilis* applied at 1:4 ratio according to sampling; T3, schedule, *Phytoseiulus persimilis* was released at a rate of 50 per square meter; T4, Floramite (UniRoyal Chemical Company, Inc., Middlesbury, CT), miticide application.
‡ ns, not significant at the 0.05 probability level.

that is supported by the *P* values for some of the multiple comparison adjustment methods. However, there is some overlap in Fig. 2. On the basis of the analysis, one should recommend to the researcher that Treatments 2 and 4 controlled the mite infestation. Treatment 3 seems to reduce the mite infestation, just not as much or as clearly as the other two methods.

Insufficient data exists for an exact choice for the value of ρ. However, Table 1 makes it clear the value of ρ is probably below 0.5. This indicates a situation in which the variability among greenhouses is small compared with the variability among pots and the overall variability. This is a reasonable assumption supported by the data.

APPENDIX

SAS Code for the Analysis

The following SAS code was used to analyze the greenhouse data.

```
/** Define global variables for use in the analysis. **/
%LET p0 = .30;/** p0 = Plug-in ICC **/
%LET gi = 1;/** gi = # of classes (greenhouses) per
   treatment **/
%LET ti = 4;/** ti = # of treatments **/
/** Construct the Conditional P value Plot **/
%MACRO p_val_plot(p);
/** Create a matrix of ratios based on the plug-in value of
   rho. **/
PROC IML;
ratio = ((&p/(1-&p))*I(&gi*&ti));
CREATE gratio from ratio;
APPEND FROM ratio;
QUIT;
DATA gratio;SET gratio;row = (_N_);RUN;
/** Use analysis to generate P values for the plot. **/
PROC MIXED DATA = greenhouses RATIO;
CLASS greenhouse trt;
MODEL count = trt/DDFM = KR;
RANDOM greenhouse(trt)/GDATA = gratio RATIOS;
ODS OUTPUT TESTS3 = pvals;
RUN;QUIT;
DATA pvals;
SET pvals;
pval = probf;
rho = &p;
RUN;
%MEND;
DATA final;
SET pvals;
DELETE;
RUN;
%MACRO iterates;
%DO i = 1 %TO 50;
%p_val_plot(&i/100);
DATA final;
SET final pvals;
RUN;
%END;
%MEND;
%iterates;
DATA zero;
pval = 0;rho = 0;
RUN;
DATA final;SET final zero;RUN;
PROC SORT DATA = final;BY rho pval;RUN;
SYMBOL I = JOIN;
PROC GPLOT DATA = final;
PLOT pval*rho/vref = 0.05;
RUN;QUIT;
/** End of Conditional P value Plot **/
/** Create a matrix of ratios based on the plug-in value of
   rho. **/
PROC IML;
RATIO = ((&p/(1-&p))*I(&gi*&ti));
create gratio from RATIO;
APPEND FROM ratio;
QUIT;
DATA gratio;SET gratio;row = (_N_);RUN;
/** Perform the analysis on the data using the plug-in value
   for rho. **/
PROC MIXED DATA = greenhousedata ratio;
CLASS greenhouse trt;
MODEL y = trt/DDFM = KR;
RANDOM greenhouse(trt)/GDATA = gratio Ratios;
LSMEANS trt/PDIFF ADJUST = BON;
LSMEANS trt/PDIFF ADJUST = TUKEY;
LSMEANS trt/PDIFF ADJUST = SIMULATE (CVADJUST);
RUN;QUIT;
```


ACKNOWLEDGMENTS

The authors would like to thank Stephen Machado and Steve Petrie of Oregon State University for organizing the ASA-CSSA-SSSA conference section on unreplicated experiments and arranging for the publication of this and the other papers of that section. The authors would also like to thank George Opit of Kansas State University for the use of the greenhouse data.

REFERENCES

Barcikowski, R.S. 1981. Statistical power with group mean as the unit of analysis. J. Educat. Stat. 6:267–285.

Blair, R.C., and J.J. Higgins. 1986. Comment on statistical power with group mean as the unit of analysis. J. Educat. Stat. 2:161–169.

Graybill, F.A. 1976. Theory and application of the linear model. Wadsworth, Pacific Grove, CA.

Perrett, J.J. 2004. Using prior information on the intraclass correlation coefficient to analyze data from unreplicated and under-replicated experiments. Ph.D. diss. Kansas State Univ. Manhattan.