USING PRIOR INFORMATION ON THE INTRACLASS CORRELATION COEFFICIENT TO ANALYZE DATA FROM UNREPLICATED AND UNDER-REPLICATED EXPERIMENTS

by

JAMIS J PERRETT

B.S., Brigham Young University, 1998
M.S., Brigham Young University, 1999

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the

requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2004

ABSTRACT

Many studies are performed on units that cannot be replicated due to cost or other restrictions. There often is an abundance of subsampling to estimate the within unit component of variance, but what is needed for statistical tests is an estimate of the between unit component of variance. There is evidence to suggest that the ratio of the between component of variance to the total variance will remain relatively constant over a range of studies of similar types. Moreover, in many cases this intraclass correlation, which is the ratio of the between unit variance to the total variance, will be relatively small, often 0.1 or less. Such situations exist in education, agriculture, and medicine to name a few.

The present study discusses how to use such prior information on the intraclass correlation coefficient (ICC) to obtain inferences about differences among treatments in the face of no replication. Several strategies that use the ICC are recommended for different situations and various designs. Their properties are investigated. Work is extended to under-replicated experiments. The work has a Bayesian flavor but avoids the full Bayesian analysis, which has computational complexities and the potential for lack of acceptance among many applied researchers. This study compares the prior information ICC methods with traditional methods. Situations are suggested in which prior information ICC methods are preferable to traditional methods and those in which the traditional methods are preferable.

USING PRIOR INFORMATION ON THE INTRACLASS CORRELATION COEFFICIENT TO ANALYZE DATA FROM UNREPLICATED AND UNDER-REPLICATED EXPERIMENTS

by

JAMIS J PERRETT

B.S., Brigham Young University, 1998
M.S., Brigham Young University, 1999

A DISSERTATION

submitted in partial fulfillment of the

requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics

College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2004

Approved by _____
Major Professor: James Higgins

ABSTRACT

Many studies are performed on units that cannot be replicated due to cost or other restrictions. There often is an abundance of subsampling to estimate the within unit component of variance, but what is needed for statistical tests is an estimate of the between unit component of variance. There is evidence to suggest that the ratio of the between component of variance to the total variance will remain relatively constant over a range of studies of similar types. Moreover, in many cases this intraclass correlation, which is the ratio of the between unit variance to the total variance, will be relatively small, often 0.1 or less. Such situations exist in education, agriculture, and medicine to name a few.

The present study discusses how to use such prior information on the intraclass correlation coefficient (ICC) to obtain inferences about differences among treatments in the face of no replication. Several strategies that use the ICC are recommended for different situations and various designs. Their properties are investigated. Work is extended to under-replicated experiments. The work has a Bayesian flavor but avoids the full Bayesian analysis, which has computational complexities and the potential for lack of acceptance among many applied researchers. This study compares the prior information ICC methods with traditional methods. Situations are suggested in which prior information ICC methods are preferable to traditional methods and those in which the traditional methods are preferable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1: OVERVIEW

Example 1

A researcher, wishing to compare two different teaching methods, teaches two classes: one with method 1 and the second with method 2. The grade of each student in the two classes is recorded with the purpose of comparing the average grade for the students taught by method 1 with the average grade for the students taught by method 2. The within class variation is the variability from student to student. The between class variation is due to such factors as time of day, difference in classroom setting, etc. We would expect the variation from class to class to be small relative to the within class variation, regardless of whether the students are being taught mathematics, creative writing, etc. The majority of the total variability will be explained by the difference in performance of the students within a class, and that variability should be similar from one academic subject to another. Thus, the ICC should be consistent in studies of similar types, and it will tend to be small (In an education example such as this, it would not be unusual for the ICC to be less than 0.1).

Example 2

A researcher would like to compare the effect of two different soil preparation methods on the yield of corn. The researcher has two different plots of land available for the study. Because the methods involve burning the fields, replication is prohibitive. The researcher can take multiple measurements on each plot, but there is no replication. However, we might assume the observations within subplots are taken far enough apart so that subsample errors are spatially independent. It is possible that previous research has been conducted on these same fields. Thus, the researcher could possibly have an idea of the ICC for the current study.

Example 3

A researcher would like to compare the effect of three different fertilizers and four irrigation systems on the yield of corn. The researcher has four different plots of land available for the study. Each plot of land has a different irrigation system. Each plot of land is divided into several subplots, and to each subplot, one of three types of fertilizer is randomly applied. There is no replication on the whole plot. From similar previous studies on similar fields, the researcher should have an idea of the ratio of the whole plot to the subplot variances.

In many cases a researcher may have a good idea of the value of the ICC from past studies. In this study, we propose methods that utilize information on the ICC of past studies to conduct hypothesis tests for current studies. Our methods are confined to situations in which reliable information is known on the value of the ICC, and in which that value is small, generally less than 0.5. The nature of the ICC in hypothesis testing allows for an increased power when the ICC is small. However, this does not hold in cases where the ICC is large.

Unreplicated Experiments

An unreplicated experiment is an experiment in which a treatment of interest is applied to only one unit. An under-replicated experiment is an experiment in which a treatment of interest is applied to a limited number of units (generally less than five units per treatment). Some experiments logistically cannot be replicated. Circumstances that might prevent replication are cost in time or money or both, scarcity of experimental units, destructive experimentation, among other things. Some researchers just don't have an extra plot of land they can experiment on. We consider what can be done in such cases.

We use the context of Example 1 to define the terms (using words like class and student) to simplify the explanation, with the understanding that these procedures can be applied to a variety of situations and disciplines.

Let y_{ijk} be the measurement taken on the kth student in the jth class given the ith treatment, μ_i is the fixed effect of treatment i, δ_{ij} is the random effect of the class j given treatment i, and ε_{ijk} is the random effect of student k in class j given treatment i, i = 1, 2, …,t; j = 1, 2, …, b_i; k

= 1, 2, …, n_{ij}. The total number of treatment levels is t. The total number of classes in treatment level i is b_i. The total number of students in class j of treatment level i is n_{ij}. Let $\delta_{ij} \sim n\left(0, \sigma_\delta^2\right)$, where σ_δ^2 represents the between class variability; let $\varepsilon_{ijk} \sim n\left(0, \sigma_\varepsilon^2\right)$, where σ_ε^2 represents the between student within class variability. It is assumed that δ_{ij} and ε_{ijk} are independent. We write out a model for the experiment as follows:

$$y_{ijk} = \mu_i + \delta_{ij} + \varepsilon_{ijk} \qquad (1.1.1)$$

This type of model is a single factor completely randomized design (CRD) with subsampling, where classes are the experimental units for each treatment level, and the students within each class are the subsamples, or observational units. A researcher who uses the students as the experimental units ignores the variability that can exist between different classes receiving the same treatment. Such an assumption is to claim that $\sigma_\delta^2 = 0$. If the researcher correctly uses classes as the experimental unit, there is only one unit per treatment level and zero error degrees of freedom available for testing the difference between these treatment means (Barcikowski, 1981 and Blair, 1986).

The Intraclass Correlation Coefficient and the Independence Assumption

The intraclass correlation coefficient (ICC) is defined as the correlation between y_{ijk} and $y_{ijk'}$ (two subsample units within one experimental unit). In this study, ρ refers to the true value of the ICC and ρ_0 refers to a best guess value, chosen by the researcher, to substitute into formulas in place of the ICC in the analysis. The ICC for model (*1.1.1*) can be obtained using the following formula:

$$\rho = \frac{\operatorname{cov}\left(y_{ijk}, y_{ijk'}\right)}{\sqrt{\operatorname{var}\left(y_{ijk}\right)\operatorname{var}\left(y_{ijk'}\right)}} = \frac{\sigma_\delta^2}{\sqrt{\left(\sigma_\delta^2 + \sigma_\varepsilon^2\right)\left(\sigma_\delta^2 + \sigma_\varepsilon^2\right)}} = \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma_\varepsilon^2} \qquad (1.1.2)$$

Thus, if $\sigma_\delta^2 = 0$, the ICC is also zero. The result is independent subsamples assuming normality of error terms.

To test the difference between two treatment means we test $H_0 : \mu_1 = \mu_2$ versus $H_0 : \mu_1 \neq \mu_2$, the following formulas apply: Let $n_{i\cdot} = \sum_{j=1}^{b_i} n_{ij}$ and $\langle n_{i\cdot} \rangle^2 = \sum_{j=1}^{b_i} n_{ij}^2$.

Then

$$\operatorname{var}(\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}) = \sigma_\delta^2 \left[\left(\frac{1}{n_{i\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 + \left(\frac{1}{n_{i'\cdot}} \right)^2 \langle n_{i'\cdot} \rangle^2 \right] + \sigma_\varepsilon^2 \left[\frac{1}{n_{i\cdot}} + \frac{1}{n_{i'\cdot}} \right] \tag{1.1.3}$$

$$= \sigma_\varepsilon^2 \left\{ \left(\frac{\rho}{1-\rho} \right) \left[\left(\frac{1}{n_{i\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 + \left(\frac{1}{n_{i'\cdot}} \right)^2 \langle n_{i'\cdot} \rangle^2 \right] + \left[\frac{1}{n_{i\cdot}} + \frac{1}{n_{i'\cdot}} \right] \right\} \tag{1.1.4}$$

since $\sigma_\delta^2 = \sigma_\varepsilon^2 \frac{\rho}{1-\rho}$. Let

$$Z = \frac{\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}}{\sqrt{\sigma_\varepsilon^2 \left\{ \left(\frac{\rho}{1-\rho} \right) \left[\left(\frac{1}{n_{i\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 + \left(\frac{1}{n_{i'\cdot}} \right)^2 \langle n_{i'\cdot} \rangle^2 \right] + \left[\frac{1}{n_{i\cdot}} + \frac{1}{n_{i'\cdot}} \right] \right\}}} \tag{1.1.5}$$

where Z has a standard normal distribution.

If ρ is incorrectly assumed to be zero, we have

$$Z = \frac{\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}}{\sqrt{\sigma_\varepsilon^2 \left(\frac{1}{n_{i\cdot}} + \frac{1}{n_{i'\cdot}} \right)}} \tag{1.1.6}$$

A test statistic for a hypothesis based on the incorrect assumption that observations are independent will be too large, consequently inflating the associated Type 1 error.

A Plug-in Value of ρ

In practice, σ_ε^2 may be estimated from the pooled variance of observational units within the experimental unit. With a lot of subsamples, σ_ε^2 can be estimated quite accurately. However, in an unreplicated experiment there is no way to estimate σ_δ^2 and consequently ρ. We assert that it is possible to learn something about ρ from a prior experiment. In an education context, for instance, ρ is typically found in the range of 0.1 or less (see *College Course Grades,* pg. 25 of this study). This information enables us to do tests in the unreplicated case by "plugging in" a reasonable value ρ_0 in place of ρ. By having an accurate estimate of σ_ε^2 and a reasonable value of ρ_0, σ_δ^2 can be approximated. In the under-replicated case, there may be two or three replications of the experiment yielding two or three experimental units per treatment. In the case of under-replicated experiments, using a plug-in value may yield more powerful tests than can be obtained by standard mixed model analysis. However, as the number of replications increase, the potential gain in using a plug-in value for ρ may diminish, and if the plug-in value ρ_0 is substantially in error, the plug-in method may yield a worse (less powerful and biased) test. The following chapters discuss the effects of different values of ρ_0, how accurate the choice of ρ_0 needs to be to maintain acceptable probabilities of Type 1 and 2 errors, and different strategies used to obtain useful values of ρ_0.

In Chapter 2 we study tests for differences among means in the case of no replication. First we look at the two-treatment completely randomized designs in which we study the effect that a plug-in value for the ICC has on the significance level and power of the test. We then propose several strategies that implement the plug-in method and a weighted p-value method for hypothesis testing. We give a description, demonstrate implementation, and examine properties of each proposed strategy. Then we look at the t-treatment completely randomized designs and demonstrate how multiple comparison procedures can be implemented.

In Chapter 3 we study the case of under-replication. First we look at the two-treatment completely randomized designs in which we study the effect of replication on significance levels and power for hypothesis tests. Then we look at t-treatment completely randomized

design case. Of special interest is the diminishing advantage of using a hypothesized ICC as the number of replications increases.

In Chapter 4 we look at the case of the split-plot with CRD whole plot design structure and a completely random assignment of subplot treatments to multiple subplot units. We give a description of the methods, examine the properties, and give an example of implementation.

In Chapter 5 we incorporate the strategies presented in this study into an analysis of a set of real data, demonstrating all the steps for a proper analysis.

In Chapter 6 we summarize the results, discuss future extensions to the case of randomized complete block design structure and suggest a Bayesian approach to choosing a plug-in value for the ICC. We discuss how this approach compares with the strategies mentioned in previous chapters, and then discuss how the plug-in method will be introduced to scientific literature.

1.2: LITERATURE REVIEW

A traditional method of comparing treatment means in an educational context is to first compute classroom means then perform t-tests using the classroom means as the response variables on the experimental units (the classroom). Such an analysis has zero error degrees of freedom in the no replication case and few error degrees of freedom in the under-replication case. In this context, Barcikowski (1981) studied the level of power for tests for means that use the class average as the unit of analysis. He analyzed sample size requirements for different levels of power. Blair, Higgins, Topping, and Mortimer (1983) show that the Type 1 error rate is grossly inflated when treatments are assigned to class but analysis is performed on students. One example demonstrated a Type 1 error rate of approximately 0.50 when treating observational units as the unit of analysis as opposed to a Type 1 error rate of approximately 0.05 when treating experimental units as the unit of analysis (inflation is 10-fold!).

Blair and Higgins (1986) showed that analysis of treatment means when the population intraclass correlation coefficient is known enhances the power of the test by allowing observational unit degrees of freedom to be used in place of experimental unit degrees of freedom. The power is always greater when the intraclass correlation coefficient is known – appreciably higher when the sample size is small.

Most of the recent literature regarding the intraclass correlation coefficient focuses on constructing confidence intervals of the ICC, and ways of using the ICC as a measure of reliability.

Groggel, Wackerly, and Rao (1988) looked at a rank-based method for obtaining point and interval estimates of a scale version of the intraclass correlation coefficient in a one-way random effects model. This rank-based method is applicable to a broad class of situations, easy to implement, and fairly accurate. It performs well when compared with other procedures. However, this method does not perform well for minimal replication.

Burch and Harris (1999) compared an estimator associated with the likelihood function derived from a pivotal quantity to estimators using both subjective and objective priors. The estimators all performed well with respect to Bayes risk. However, the authors recommend the estimator obtained from the pivotal approach for several reasons. Among these are that none of the other methods performs uniformly better and that the pivotal method has a simple closed form, is widely applicable, and is easy to incorporate prior information. However, this method does not perform well for minimal replication.

Bansal (2000) derived Bayesian estimators of the intraclass correlation coefficient for improper priors. His paper consists mainly of derivations of estimators and their asymptotic distributions. Bond and Higgins (2001) compared method of moments with Bayes estimators of the intraclass correlation coefficient for different priors. They found the Bayes estimate performed better than the method of moments estimators (smaller MSE). In addition, favorable results were found even for minimal replication (m=2).

CHAPTER 2: USING A KNOWN INTRACLASS CORRELATION COEFFICIENT TO TEST THE DIFFERENCE BETWEEN TREATMENT MEANS (NO REPLICATION)

2.1: THE TWO-TREATMENT CASE

In the two-treatment case we obtain a test statistic as if ρ is known. Then we substitute, or "plug-in," a presumed value ρ_0 for the unknown ρ, where the value ρ_0 is based on prior information. Strategies for choosing ρ_0 are discussed later. For now we consider properties of the test statistic using the plug-in value. This section examines the two-treatment case with no replication. The model for this case is represented by model 1.1.1 with the subscript j suppressed.

Test Statistic: T-Test

A test statistic for testing H_0: $\mu_1 = \mu_2$ vs. H_A: $\mu_1 \neq \mu_2$ can be computed from our data as follows: Let ρ denote the true ICC. Let ρ_0 denote the value the researcher chooses for the ICC. Let $\mu_{01} - \mu_{02}$ represent the hypothesized difference of the mean for treatment 1 and the mean for treatment 2 respectively. The test statistic is

$$T = \frac{(\bar{y}_{1\cdot} - \bar{y}_{2\cdot}) - (\mu_{01} - \mu_{02})}{\sqrt{\hat{\sigma}_\varepsilon^2 \left\{ 2\left(\frac{\rho_0}{1-\rho_0} \right) + \left[\frac{n_1 + n_2}{n_1 n_2} \right] \right\}}} \qquad (2.1.1)$$

where $\hat{\sigma}_\varepsilon^2$ represents a pooled estimate of the within class variance. Let

$$s_i^2 = \frac{\sum_{k=1}^{n_i} (y_{ik} - \bar{y}_{i\cdot})^2}{n_i - 1}, \text{ i=1, 2} \qquad (2.1.2)$$

be the variance of the measurements under treatment *i*. Then

$$\hat{\sigma}_\varepsilon^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \qquad (2.1.3)$$

We will denote the nominal level of the test α_0 and the true level of significance α. Depending on assumptions, α_0 may or may not equal α.

Properties of the test statistic

For simplicity, we will assume the number of subsamples per class is the same for all classes, i.e. $n = n_1 = n_2$. For a more general formula, see Appendix B. The sample size is one class per treatment. Let $\mu_1 - \mu_2$ be the true value of the difference between the treatment means to be compared in our hypothesis. Let $\upsilon = 2(n-1)$, the degrees of freedom for our test. Let $t_{0.05,\upsilon}$ denote the upper tail 0.05 value of the t-distribution with υ degrees of freedom.

The probability of rejecting H_0 for an upper-tail test at $\alpha_0 = 0.05$ can be determined using the following steps:

$$P\left[\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_{01} - \mu_{02})}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n}\right)}} > t_{0.05,\upsilon}\right] \qquad (2.1.4)$$

$$= P\left[\frac{(\bar{y}_1-\bar{y}_2)-(\mu_{01}-\mu_{02})+(\mu_1-\mu_2)-(\mu_1-\mu_2)}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)}}\cdot\frac{\sqrt{\sigma_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}{\sqrt{\sigma_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}} > t_{0.05,\upsilon}\right]$$

$$= P\left[\frac{\frac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{\sqrt{2\sigma_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}+\frac{(\mu_1-\mu_2)-(\mu_{01}-\mu_{02})}{\sqrt{2\sigma_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}}{\frac{\sqrt{\hat{\sigma}_\varepsilon^2}}{\sqrt{\sigma_\varepsilon^2}}} > \left(t_{0.05,\upsilon}\right)\frac{\sqrt{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}}{\sqrt{\frac{\rho}{1-\rho}+\frac{1}{n}}}\right]$$

If we define $Z=\dfrac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{\sqrt{2\sigma_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}$ and $\lambda=\dfrac{(\mu_1-\mu_2)-(\mu_{01}-\mu_{02})}{\sqrt{2\sigma_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}$ it follows that Equation 2.1.4 can be expressed as

$$P\left[\frac{(\bar{y}_1-\bar{y}_2)-(\mu_{01}-\mu_{02})}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)}} > t_{0.05,\upsilon}\right] = P\left[\frac{Z+\lambda}{\sqrt{U/\upsilon}} > \left(t_{0.05,\upsilon}\right)\frac{\sqrt{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}}{\sqrt{\frac{\rho}{1-\rho}+\frac{1}{n}}}\right] \qquad (2.1.5)$$

where Z~n(0,1), λ is a constant (our non-centrality parameter), and $U\sim\chi_\upsilon^2$. To the left of the inequality is a random variable with a non-central t-distribution with non-centrality parameter λ and degrees of freedom υ.

Evaluation of Proposed Methods

There are two methods of evaluating the strategies discussed in this paper.

Method 1: Fixed ρ

The different plug-in methods involve choosing a value ρ_0. We evaluate each method by determining significance levels and power curves for tests using each method, for different values of ρ_0 and ρ. In particular, a method will be determined useful if it can maintain a Type-1 error level close to the nominal level (α_0) while providing power to detect differences in treatment means for a "reasonable" range of ρ_0 near ρ.

Method 2: Random ρ

If we do not assume there is a fixed value ρ, but rather that ρ is a random variable, we might choose as a plug-in value ρ_0 the expected value of ρ given its distribution. Under this assumption, we look at the significance level as being the proportion of times an experiment incorrectly rejects the null hypothesis as test data are generated using randomly selected values of ρ from its distribution. We denote this as the average significance level given the prior distribution of ρ. The average power can be computed similarly.

The Effect of Different Values of the ICC on Significance Level and Power of the Test (Evaluation Method 1)

We can use Equation 2.1.5 to evaluate the probability of rejection for tests of the hypothesis that the two means are equal. These probabilities depend on the values for ρ, ρ_0, n, and the non-centrality parameter, λ. In order to measure the deviation between the two means, we define a standardized difference as $StDiff = \frac{\mu_1 - \mu_2}{\sigma_\varepsilon}$. Probabilities will depend on *StDiff* through λ. Using Equation 2.1.5 probabilities were generated using the following values as indices:

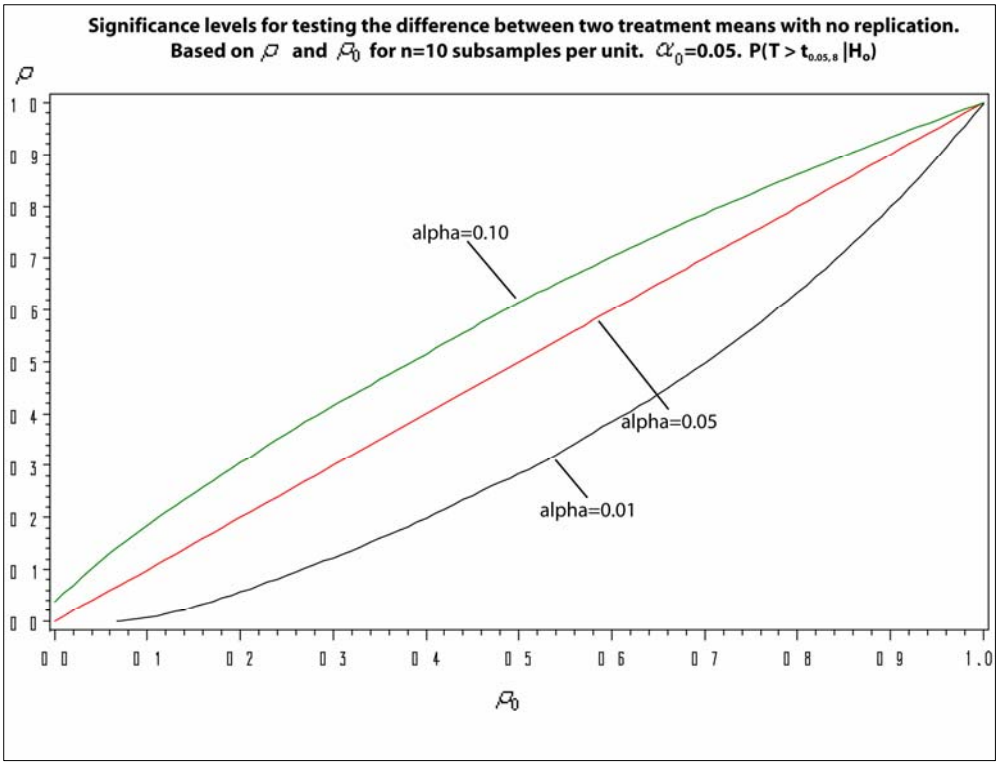
StDiff = 0.5 through 2.5 in increments of 0.5
n = {10, 30, 50, 100, ∞}
ρ = 0 through 0.99 in increments of 0.01
ρ_0 = 0 through 0.99 in increments of 0.01

$\alpha_0 = 0.05$

Table 1 in Appendix A includes a subset of the probabilities, in particular probabilities for ρ = 0 through 0.5 in increments of 0.05, and ρ_0 = 0 through 0.5 in increments of 0.05. Figures 2.1.1 through 2.1.4 include n = 10 and ∞. This range of ρ is for the situation in which the between class variability is less than or equal to the within class variability. Values of ρ larger than this will result in tests generally having low power.

The table of probabilities in Appendix A gives an idea of the effect of changes in *StDiff*, n, ρ, and ρ_0 on the significance level and power of a test of hypothesis comparing two treatment means. The following four plots also show the effect. The data used for these plots were created by evaluating Equation 2.1.5. In the first two plots $StDiff = 0$. In the next two plots $StDiff = 1.0$. For the plots in Figures 2.1.1 and 2.1.3 n=10. For the plots in Figures 2.1.2 and 2.1.4 we let $n \to \infty$ resulting in $\lim_{n\to\infty} \frac{1}{n} = 0$ being used in Equation 2.1.5. The nominal significance level $\alpha_0 = 0.05$ is used for all four graphs. Also for all four graphs, neither the power nor the significance level is defined at ρ =1 or ρ_0 =1.

Figure 2.1.1 shows the two-tail significance levels for the test with only 10 students per class. The lines represent the two-tail probabilities of rejecting our null hypothesis. If $\rho = \rho_0$, then α=0.05. If the plug-in value ρ_0 is less than the true value ρ ($\rho \geq \rho_0$), the level of significance is inflated (α≥0.05). If the plug-in value ρ_0 is greater than the true value ρ ($\rho \leq \rho_0$), the level of significance is smaller than α=0.05. So, for the test to have an actual level of significance no more than the nominal level, the researcher must select ρ_0 such that it is greater than or equal to ρ.



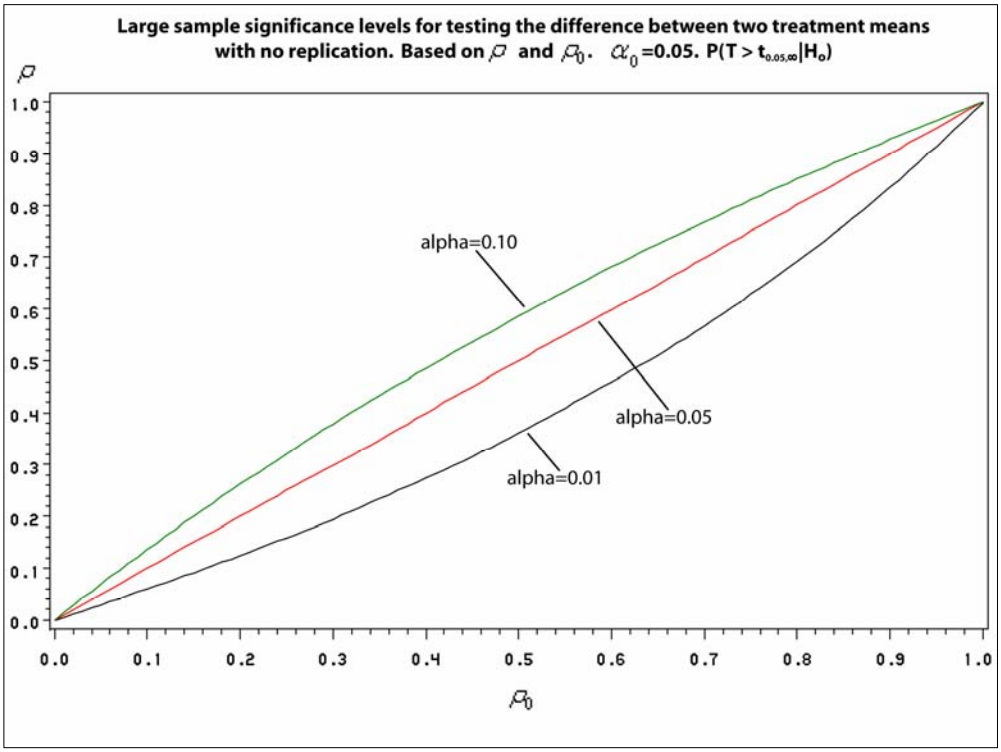


Figure 2.1.2. Large sample significance levels for infinite number of students per class.

Figure 2.1.2 shows significance levels for the test with a theoretically infinite number of students per class. This second situation is less forgiving of ρ being misspecified, but not by much. For example, if a value of ρ_0 =0.2 is used when ρ =0.4, alpha will be 0.2253; whereas, in the n=10 case, alpha would be 0.1728. Figures 2.1.1 and 2.1.2 are very similar indicating little difference between significance values based on 10 subsamples per unit and those based on a theoretically infinite number of subsamples per unit.

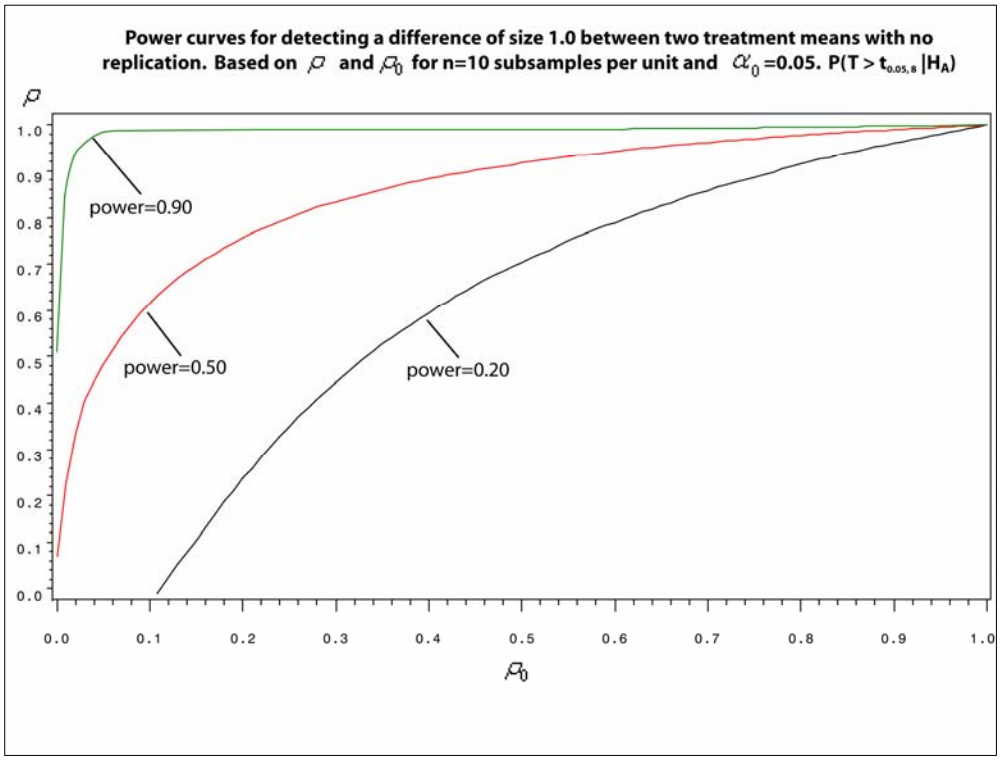


Figure 2.1.3. Small sample power curves for 10 students per class.

Figure 2.1.3 illustrates the probability of correctly detecting a difference of size *StDiff*=1.0 between treatment means for various values of ρ and ρ_0. The smaller the value of ρ_0 is in relation to ρ, the greater is the power of the test, but this is at the price of potentially inflating the probability of a type 1 error. Figure 2.1.3 depicts the situation with 10 subsamples per class. The three lines indicate ρ and ρ_0 combinations that produce a power of 0.20, 0.50, and 0.90.

The next figure, Figure 2.1.4 depicts the case with a theoretically infinite number of students per class. This power plot is overlayed with the corresponding plot of significance levels, Figure 2.1.2. The red, green, and blue areas denote power in the ranges ≤0.2, 0.2 to 0.5, and 0.5 to 0.9 respectively. Similar to the plots of significance levels, there is not a large difference in the plots of the power levels when going from 10 students per class to an infinite number of students per class.

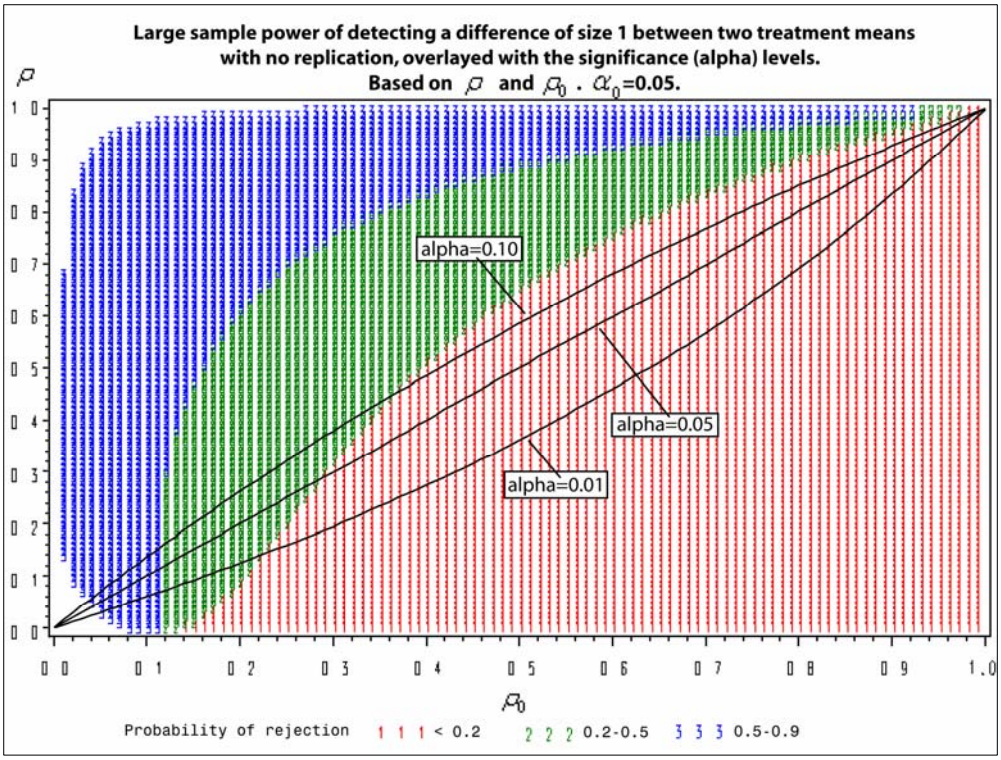


Figure 2.1.4. Overlay of large sample power and significance levels for infinite students per class.

To summarize, if $\rho_0 \le \rho$ power increases, but the significance level will be larger than α_0. If $\rho_0 \ge \rho$ the significance level will be lower than α_0, but the power of the test will be deflated. The effect that ρ and ρ_0 have on the level of significance and power shown in the tables and figures will provide a basis for developing strategies for testing for differences of means in unreplicated and under-replicated experiments.

It is also useful to examine the maximum power attainable using a plug-in value for ρ in hypothesis testing. The following figure demonstrates the maximum attainable power under the most ideal conditions: namely $\rho_0 = \rho$ and $n = \infty$.

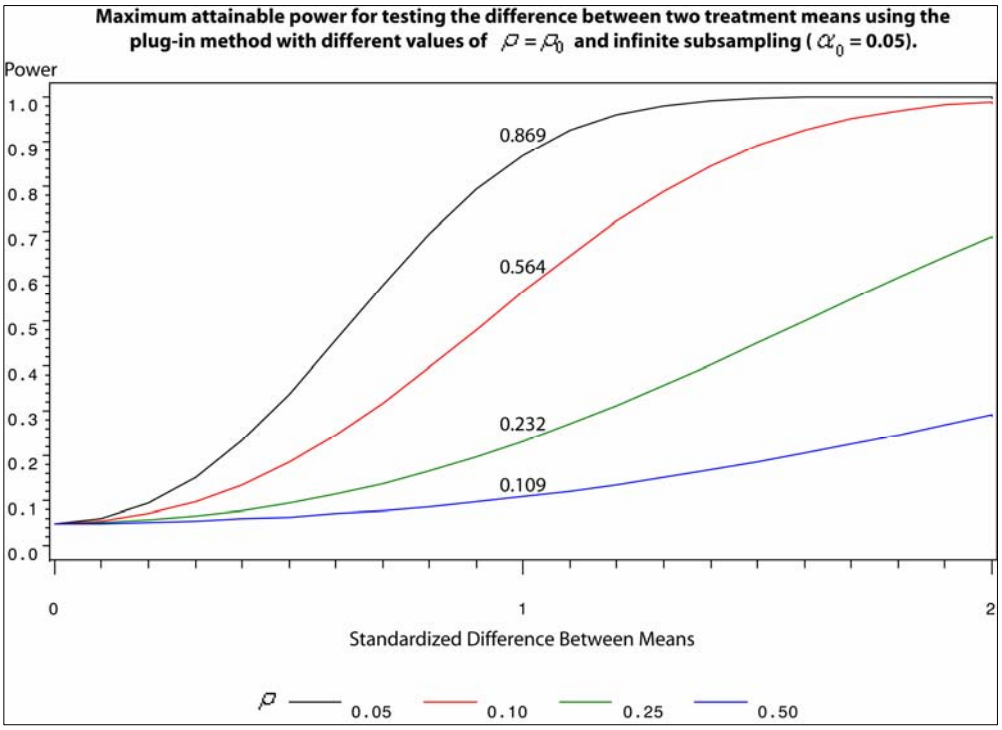


Figure 2.1.5. Maximum attainable power for infinite number of students per class.

As seen in Figure 2.1.5, if the true value ρ is 0.5, the power is low. For smaller values of ρ the power increases considerably. For instance, for a standardized difference of 1.0 and ρ =0.1, the power is 0.564. Thus, using a known ICC is only going to be effective for smaller values of ρ.

2.2: STRATEGIES FOR CHOOSING ρ_0

What the Researcher May Know About ρ

It is possible for the researcher to obtain information about ρ based on prior experiments of a similar nature, or from knowledge about the behavior of ρ for a current experiment.

Distributional Information

If much distributional information is available from prior studies of a nature similar to that of the current study, the researcher may be able to put a prior distribution or empirical distribution on ρ.

Point or Interval Information

The researcher may not have extensive distributional information about ρ, but may have an indication of the mean, or maximum value of ρ.

The Plug-in Value

In the case of no replication, zero error degrees of freedom exist for conducting a hypothesis test comparing means. So, the test cannot be performed using traditional methods. With this procedure a value, chosen by the researcher, is used as if it were the true value, ρ. This value, ρ_0, called a plug-in value, can be used in hypothesis testing and in producing confidence intervals of differences of treatment means. The strategies for choosing a value for ρ_0 given in this chapter are proposed to researchers who have an unreplicated experiment and have a reasonable idea of the actual value, ρ.

Weighted P-value

A second proposal for dealing with unreplicated experiments is to find several p-values based on a range of likely values of ρ and then find an average p-value, weighted by research-supplied weights for the values of ρ.

Plot of the Conditional P-value given ρ_0

For a given set of data, the p-value may be obtained for various assumed values of ρ_0. It is computed as the probability that the t-distribution with $n_1 + n_2 - 2$ degrees of freedom is more extreme than the observed value of the statistic defined by equation 2.1.1.

The following plots (2.2.1-2.2.3) of the p-value as a function of ρ_0 illustrate three possible situations a researcher may encounter when using a plug-in method. Depending on which of these situations the researcher encounters, the researcher may choose to reject the null hypothesis, fail to reject the null hypothesis, or simply report the p-values without making a decision to reject or fail to reject the null hypothesis.

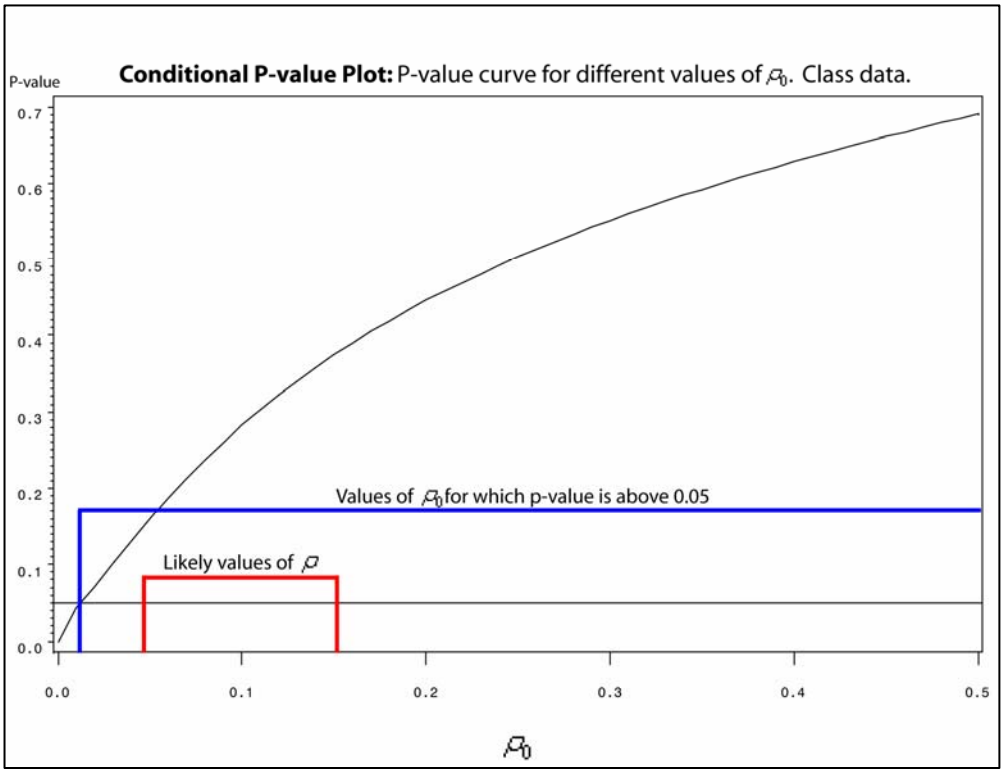


Figure 2.2.1. Conditional p-value plot for class data.

The *class data* (C.1) consists of final course grades for five classes of introductory statistics taught by five different instructors. See Appendix C for more information about the data. We

will use the data from classes one and two. We test $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$ at the $\alpha_0 = 0.05$ level of significance.

Figure 2.2.1 is the conditional p-value plot of the *class data* (C.1).

It can be seen that p-values are only significant (<0.05) if ρ_0 is less than 0.013. If the likely value of ρ is greater than 0.013, then the result of the test is to fail to reject H_0 at $\alpha_0 = 0.05$.

Figure 2.2.2 is the p-value plot of the *class data* (C.1) with a value of one added to treatment 2 scores.

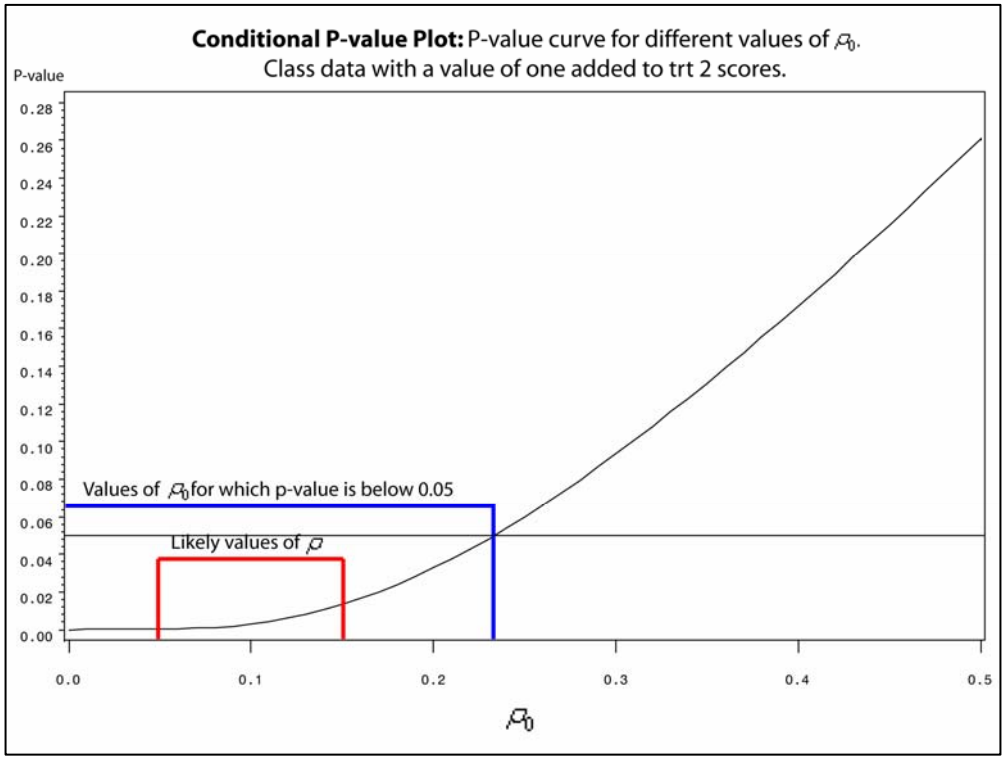


Figure 2.2.2. Conditional p-value plot for class data with a value of one added to treatment 2 scores.

If the researcher is certain ρ is less than 0.23, the result of the test is to reject H_0 at $\alpha_0 = 0.05$.

The next figure, Figure 2.2.3, demonstrates the situation in which some p-values are <0.05 for likely values of ρ and some are <0.05 for likely values of ρ. In this case a value of 0.5 was added to treatment 2 scores of the *class data* (C.1).

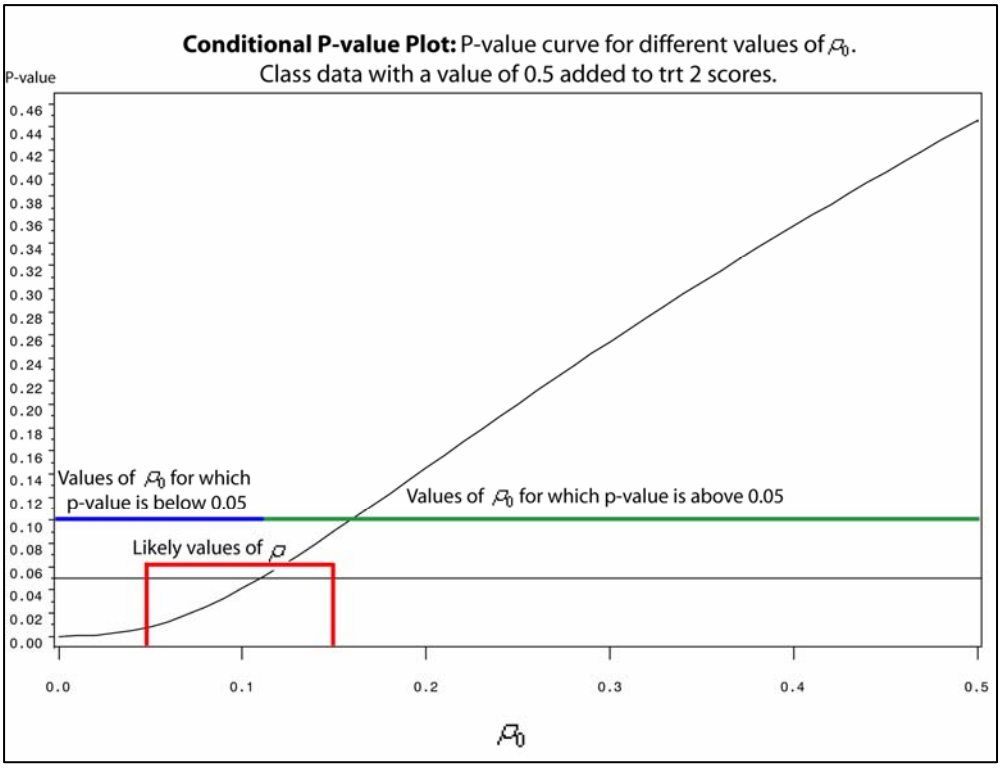


Figure 2.2.3. Conditional p-value plot for class data with a value of 0.5 added to treatment 2 scores.

There are some likely values of ρ for which the p-value is <0.05 and other likely values of ρ for which the p-value is >0.05. When this situation occurs, it is less obvious what the results of a test of hypothesis should be.

We introduce formal strategies that make use of prior information about ρ to test hypotheses about treatment means in a no-replication, two-treatment case.

Strategy 1: Maximum Rho

Description

The first strategy is called the Maximum Rho procedure. This procedure simply involves choosing ρ_0 to have the maximum value the researcher believes reasonable for ρ. That value, ρ_0, is then incorporated into the test statistic. The test rejects the null hypothesis if the p-value is less than α_0.

Properties

The Maximum Rho procedure assures the true significance level, α, is less than or equal to the nominal value α_0, with equality when $\rho_0 = \rho$. The closer ρ_0 is to ρ, the greater the power of the test.

Implementation

To implement the Maximum Rho procedure, the researcher simply uses the maximum value of ρ in place of ρ_0 in the plug in method.

College Course Grades

The value ρ was estimated from a variety of courses offered at Kansas State University using the SAS®* MIXED procedure. The components of variance consist of variability of scores due to section σ_δ^2 and the variability of scores due to students within sections σ_ε^2. Fourteen different courses were selected (CHM 111, 210, 230; CIS 101; ENGL 100, 125; MATH 010, 100; MUSIC 250, 255; PSYCH 110, 202, 350; SPAN 161) each with multiple sections, covering both Fall and Spring semesters over the years 2001-2003, for a total of 43 course-semester combinations. These values are listed as C.2 of Appendix C.

* SAS® is the registered trademark of SAS Institute Inc., Cary, NC.

All 43 estimated ICC values were at or below 0.33. The majority, 95%, were at or below 0.2 – 90% were below 0.15. Only one value was at 0.33. That value is for an honors English course (ENGL 125). Other courses include undergraduate courses in chemistry, English, music, CIS, math, psychology, and Spanish. Based on these values, it would be reasonable to use ρ_0=0.15 for the Maximum Rho strategy in an education-type study.

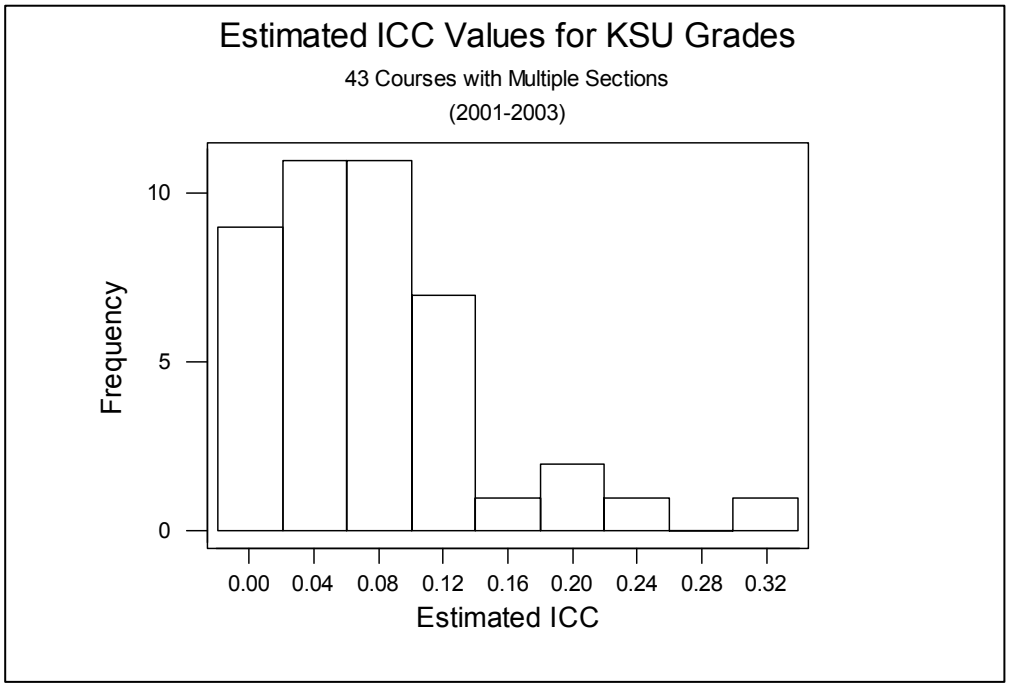


Figure 2.2.4. Estimated ICC values for KSU grades data.

The *class data* (C.1) consists of final course grades for five classes of introductory statistics taught by five different instructors. See Appendix C for more information about the data. We will use the data from classes one and two to demonstrate the implementation of the plug-in method using the Maximum Rho strategy.

Let $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$. Let $\alpha_0 = 0.05$. Let ρ_0=0.15.

The following SAS®code can be used to implement the plug-in method.

```
/** STEP 1 **/
%let p0=.15;  /** p0 = Plug-in ICC **/
%let gi=1;  /** gi = # of classes per treatment **/
%let ti=2;  /** ti = # of treatments **/
```

```
/** STEP 2 **/
proc iml;
RATIO=((&p0/(1-&p0))*I(&gi*&ti));
create gratio from RATIO;
append from ratio;
quit;
data gratio;set gratio;row=(_N_);run;

/** STEP 3 **/
proc mixed data=classdata ratio;
class class trt;
model y=trt/ddfm=kr;
random class(trt)/ gdata=gratio Ratios;
run;quit;
```

The first step (STEP 1) creates global variables and sets their values so that these values can be used in procedures and data steps throughout the program. The MIXED procedure allows known ratios of variance components to be put in the model. The second step (STEP 2) creates a matrix and subsequently a data set containing the ratios that will be used for the matrix that represents the variance of the random effects. The third step (STEP 3) analyzes the data using the appropriate model. The GDATA= and RATIOS options indicate the data set GRATIO will be used in place of the traditional estimates of the matrix representing the variance of the random effects, known as the G matrix. The DDFM=KR option ensures the test is using the appropriate degrees of freedom, adjusting for the known ratios.

The following are the results:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	1	68	0.80	0.3745

The p-value, 0.3745, is larger than $\alpha_0 = 0.05$. So the result of the test is to fail to reject the null hypothesis in favor of the alternative (at $\alpha_0 = 0.05$). The conclusion is that the difference between the mean grades for the two classes is not significant. The probability of a Type 1 error in this case is *at most* 0.05. If, in fact, ρ is 0.15, the power for detecting a

difference of one grade point is approx. 0.3373. However, if ρ is less than 0.15, the power will be less.

Strategy 2: Acceptable Interval (Best- Worst-Case Scenario)

Description

The Acceptable Interval strategy involves setting ρ_0 equal to the maximum and minimum value ρ is likely to have. Next, the researcher analyzes the data at each of those ρ_0 values. Finally, the researcher reports both p-values. If both p-values are less than α_0, the result is to reject the null hypothesis in favor of the alternative. If both the p-values are greater than or equal to α_0, the result is to fail to reject the null hypothesis. If one p-value is less than α_0 and the other is greater than α_0, the researcher will simply report both p-values, without rejecting or failing to reject the null hypothesis.

To understand this strategy, it is helpful to recall plots 2.2.1-2.2.3. With both plots 2.2.1 and 2.2.2, it is clear the researcher should reject or fail to reject the null hypothesis. However, in plot 2.2.3 it is not clear what the result should be, and so the researcher simply reports the p-values and lets the reader decide what conclusions, if any, are appropriate.

Properties

If the results of the test agree for both analyses, the researcher can choose to reject or fail to reject the null hypothesis. In either case, $\alpha \leq \alpha_0$ as long as ρ is not greater than the greater of the two ρ_0's.

Implementation

Again we use the *class data* (C.1), consisting of final course grades for five classes of introductory statistics taught by five different instructors. We will use the data from classes one and two to demonstrate the implementation of this strategy. The researcher believes ρ is at most 0.15 and at least 0.05. The SAS® code for this strategy will be analogous to the SAS® code presented for the first strategy, because both use plug-in estimators of the ICC.

Let $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$. Let $\alpha_0 = 0.05$. Let ρ_0=0.15.

The following are the results:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	1	68	0.80	0.3745

Next, let ρ_0=0.05.

The following are the results:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	1	68	2.02	0.1600

Because both p-values 0.3745 and 0.1600 fail to reject the null hypothesis at $\alpha_0 = 0.05$, the researcher concludes the difference in the treatment means is not significant.

Strategy 3: Adjusted Degrees of Freedom Using Mean Rho (ADFMR)

Description

This procedure involves choosing ρ_0 to be the expected value of ρ assuming that ρ is a random variable. The researcher puts a prior distribution on ρ and uses the expected value of that prior distribution as the plug-in value ρ_0.

Computing the Error Degrees of Freedom

We consider two ways of computing error degrees of freedom when treating ρ as a random variable.

Full d.f.: The first method uses the equation $d.f. = 2(n-1)$. This is the full error degrees of freedom for the test assuming ρ is known.

Reduced d.f.: The other method uses fewer error degrees of freedom than the full $2(n-1)$. One way to reduce the error degrees of freedom is to use a Satterthwaite-type adjustment. The Satterthwaite-type adjustment is of the form

$$d.f. = \frac{2[Var(\bar{y}_1 - \bar{y}_2)]^2}{Var[\hat{Var}(\bar{y}_1 - \bar{y}_2)]}. \qquad (2.2.1)$$

Let

$$W = \frac{\rho}{1-\rho} + \frac{1}{n}.$$

Then

$$Var(\bar{y}_1 - \bar{y}_2) = 2\sigma_\varepsilon^2 W$$

and

$$\hat{Var}[Var(\bar{y}_1 - \bar{y}_2)] = \hat{Var}[2\sigma_\varepsilon^2 W \mid \rho] = Var[E(2\sigma_\varepsilon^2 W \mid \rho)] + E[Var(2\sigma_\varepsilon^2 W \mid \rho)]$$

$$= 4\sigma_\varepsilon^4 Var(W) + \frac{4\sigma_\varepsilon^4}{(n-1)} E(W^2) = \frac{4\sigma_\varepsilon^4}{(n-1)} \{n \cdot Var(W) + [E(W)]^2\}.$$

So, Equation 2.2.1 becomes

$$d.f. = \frac{2(n-1)[E(W)]^2}{n \cdot Var(W) + [E(W)]^2}. \qquad (2.2.2)$$

As $Var(W) \to 0$, the error degrees of freedom using the Satterthwaite-type adjustment approaches $2(n-1)$, the full d.f., indicating that the more sure and accurate the choice of ρ_0, the more appropriate the $2(n-1)$ error degrees of freedom. Also, as $n \to \infty$, the error degrees of freedom using the Satterthwaite-type adjustment approaches

$$d.f.(Satterthwaite-type) = 2\left\{\frac{\left[E\left(\frac{\rho}{1-\rho}\right)\right]^2}{Var\left(\frac{\rho}{1-\rho}\right)}\right\}.$$

Note: Because the computation of the Satterthwaite-type d.f. involves the variance of W, the researcher must have a nonsingular distribution for ρ in order to compute the Satterthwaite-type degrees of freedom.

Properties

The significance level and power for tests using the ADFMR strategy will differ according to the value of the error degrees of freedom. On average, tests using the full $2(n-1)$ error d.f. appear to have a significance level slightly above the nominal level, but will have much greater power to detect significant differences. This is investigated in more detail later (pages 44-50). In the unreplicated case, the Satterthwaite adjustment drastically reduces the error degrees of freedom, almost guaranteeing the significance level will stay below its nominal level. However, the test using a Satterthwaite adjustment will have very little power to detect significant differences. Other methods of reducing the error degrees of freedom may more appropriately reduce the error degrees of freedom to a point in which the significance level is below the nominal level, while maintaining an acceptable level of power. Many proposed methods exist for adjusting error degrees of freedom in addition to a Satterthwaite-type adjustment. This study does not consider other methods. Further, it is the author's view that finding more accurate methods of choosing ρ_0 has a more positive effect than adjusting degrees of freedom because resulting tests maintain a greater level of power than do tests using adjusting degrees of freedom while closely maintaining a specified level of significance, especially in the no replication case.

Implementation

To implement the ADFMR procedure using a prior distribution, either empirical or theoretical, for ρ, the researcher first finds $\rho_0 = E(\rho)$. This value is then used in the plug-n

method shown previously. If the researcher chooses to use the Satterthwaite adjustment for the error degrees of freedom, such an adjustment must also be computed before computing the p-value of the test. Note: the SAS® MIXED procedure uses 2(n-1) denominator degrees of freedom in this situation using the DDFM=KR or DDFM=SATT option.

Example 1

Suppose the researcher wishes to test the equality of two treatment means from the class data. Further, suppose the researcher wishes to use the distribution of the *KSU Grades data* (Appendix C.2) estimated ICC values as the prior empirical distribution of ρ. The distribution has a mean of 0.077710. So, $\rho_0 = 0.077710$. Next, the researcher computes $W = \frac{\rho}{1-\rho} + \frac{1}{n}$ for each value of ρ in the prior empirical distribution of ρ and then computes the mean and variance of that distribution. In this case, $E(W) = 0.119641$ and $Var(W) = 0.00894563$. Consequently, d.f.=2.97282.

Let $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$. Let $\alpha_0 = 0.05$. Let ρ_0=0.077710. The following SAS® code can be used to implement the plug-in method.

```
/** STEP 1 **/
%let p0=0.077710;  /** p0 = Plug-in ICC **/
%let gi=1;  /** gi = # of classes per treatment **/
%let ti=2;  /** ti = # of treatments **/

/** STEP 2 **/
proc iml;
RATIO=((&p0/(1-&p0))*I(&gi*&ti));
create gratio from RATIO;
append from ratio;
quit;
data gratio;set gratio;row=(_N_);run;

/** STEP 3 **/
proc mixed data=classdata ratio;
class class trt;
model y=trt/ddf=2.97287;
random class(trt)/ gdata=gratio Ratios;
run;quit;
```

The following are the results:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	1	2.97	1.45	0.3152

The result is to fail to reject the null hypothesis. On average the significance level will be close to $\alpha_0 = 0.05$ and, using Equation 2.1.5, the probability of detecting a difference of $StDiff = 1$ will be close to 0.3108. Using $d.f. = 2(n-1)$, the result would be: Error d.f. = 68 P-value = 0.2323. On average the significance level will be close to $\alpha_0 = 0.05$ and, using Equation 2.1.5, the probability of detecting a difference of $StDiff = 1$ will be close to 0.5459.

Example 2

Now suppose the researcher wishes to use Beta(5,30) as the prior distribution of ρ. The following graph depicts the Beta(5,30) distribution.

Figure 2.2.5 is similar in appearance to the empirical distribution of the *KSU Grades data* (Appendix C.2). The probability is mostly between 0.05 and 0.3 with a peak around 0.13. $\rho_0 = E(\rho) = 0.14286$.

Using the prior distribution, $E(W) \approx 0.2007924$ and $Var(W) \approx 0.0071633$ making Satterthwaite-type error degrees of freedom 9.4202781. The results of the analysis are

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	1	9.42	0.84	0.3824

The result is to fail to reject the null hypothesis. On average the significance level will be close to $\alpha_0 = 0.05$ and, using Equation 2.1.5, the probability of detecting a difference of $StDiff = 1$ will be close to 0.3016.

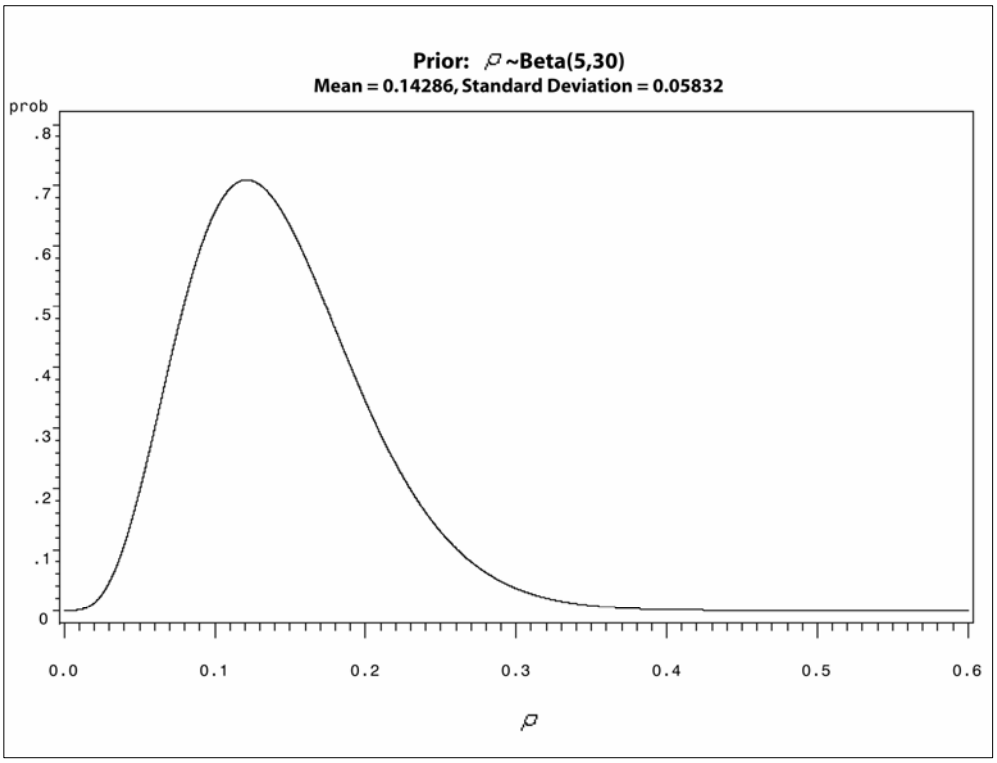


Figure 2.2.5. Prior distribution of ρ, Beta(5,30).

Strategy 4: Weighted P-Value

Description

The weighted p-value strategy is to take a weighted average of the p-values for different values of ρ_0 according to prior knowledge of the behavior of the possible values of ρ. The researcher then rejects the null hypothesis at level alpha if the resulting average weighted p-value is smaller than α_0. For simplicity of computation we take the weighted average of no more than five p-values, where weights are based on prior belief as to where ρ will fall.

Let ρ_{0i} be one of three to five plug-in values of the ICC (i=1 to 3 or 5) used in testing a hypothesis. Let w_i be an individual weight such that $w_i \geq 0$ and $\sum_i w_i = 1$. Let p_i be the p-value of a hypothesis test using ρ_{0i} as the plug-in value. Then $\sum_i w_i p_i$ is the *weighted p-value.*

To compare the possible effects of weights on the p-values we propose the candidate weight functions T1, T2, T3, T4, T5 shown in Figures 2.2.6 through 2.2.10. For instance, with T1, p-values are calculated at ρ_{01}=0.10, ρ_{02}=0.15, and ρ_{03}=0.20. The resulting p-values are then multiplied by weights w_1=0.25, w_2=0.50, and w_3=0.25 respectively. The resulting three values were then summed to give the weighted p-value. Average weighted p-values based on T2 through T5 are found similarly.

Properties

Now we look at how this weighted p-value method performs. A simulation study was performed using the five weight functions T1-T5. One thousand data sets were generated for each of the following combinations:

- n=100 students per class
- t=2 treatment levels
- b=1 class per treatment
- Standardized difference (*StDiff*) =0, 1.0.
- ρ =0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35

Two-sided t-tests were conducted on each of the 1,000 data sets, using the plug-in values of ρ_{0i} associated with the weight functions T1-T5, and p-values were recorded. Weighted p-values were then computed according to the weight functions T1-T5 represented in Figures 2.2.6-2.2.10. The proportion of the tests that rejected the null hypothesis at α_0=0.05, based on the weighted p-values was recorded for each combination of ρ and *StDiff*, which is the simulated p-value of the test.

The Figures 2.2.11-2.2.14 illustrate these proportions for each weighted scheme across each value of ρ. These graphs illustrate the comparison of weighting p-values for the five chosen weight functions. Figures 2.2.13 and 2.2.14 also contain proportions for tests using the plug-in method with all the weight at a single point to show how the weighted p-values compare with the plug-in method.

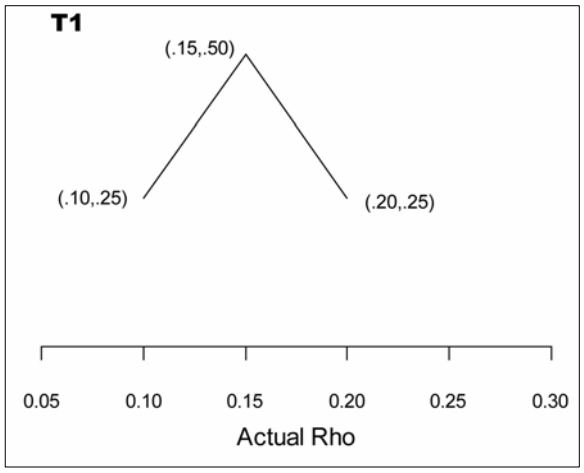


Figure 2.2.6. Weight function T1.

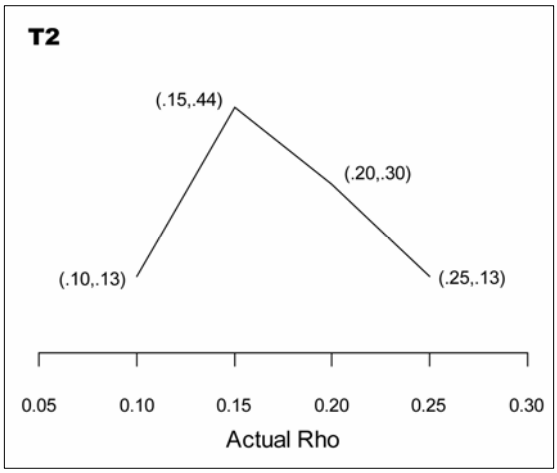


Figure 2.2.7. Weight function T2.

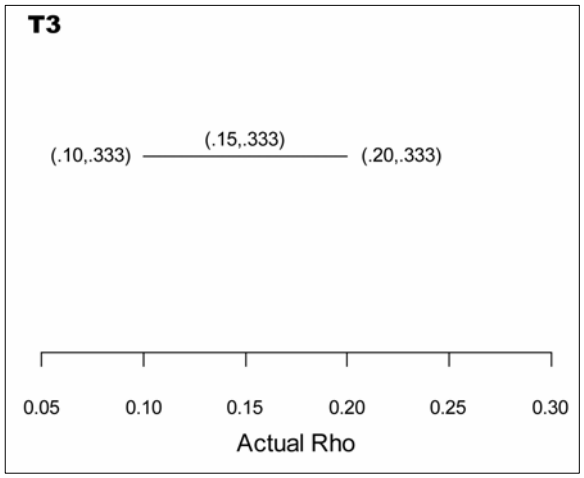


Figure 2.2.8. Weight function T3.

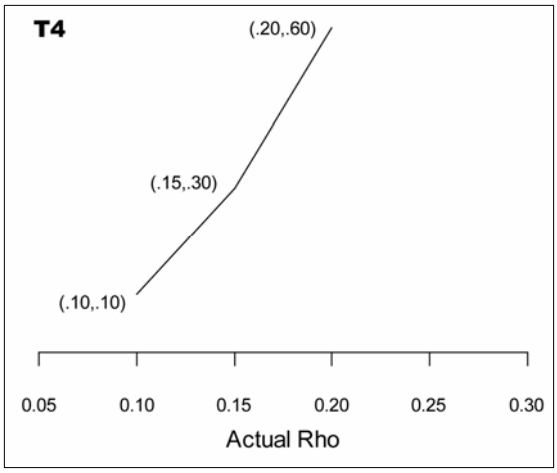


Figure 2.2.9. Weight function T4.

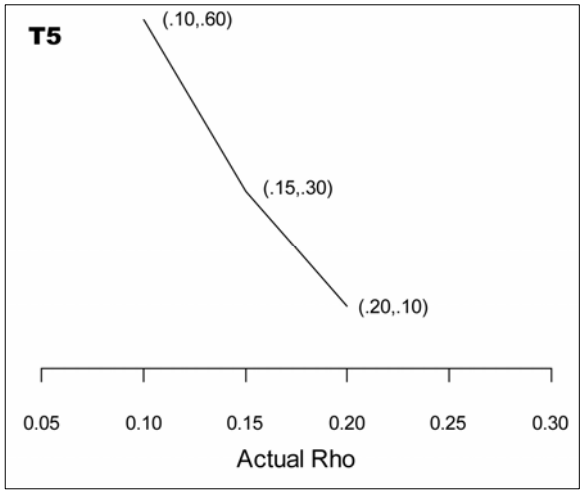


Figure 2.2.10. Weight function T5.

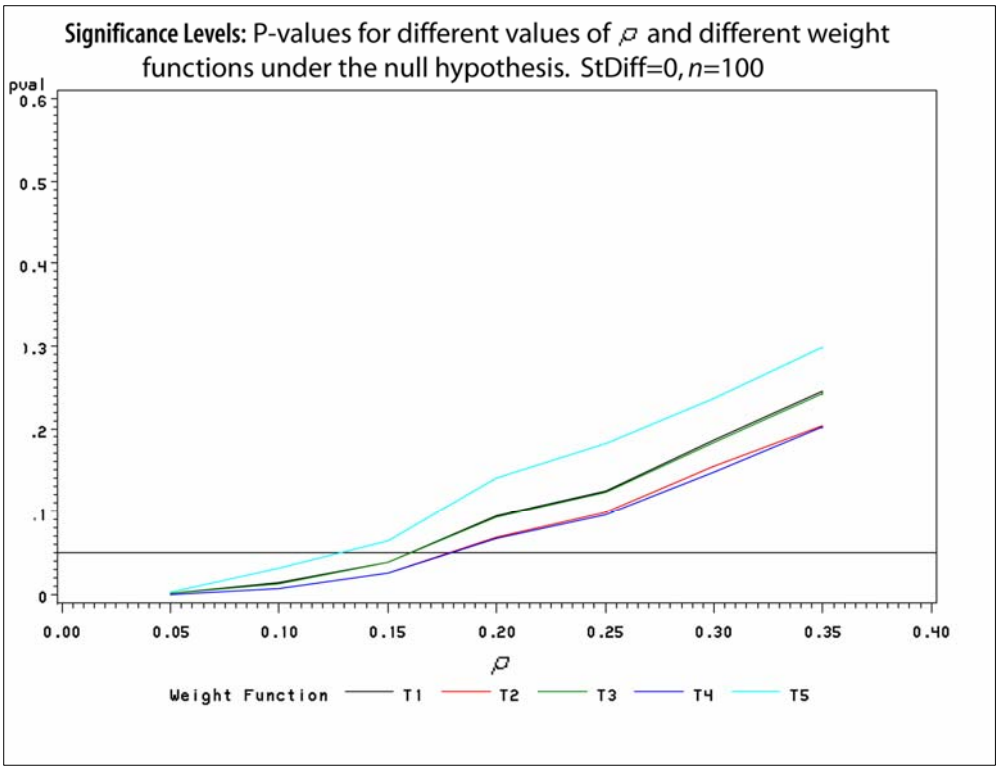


Figure 2.2.11. Significance levels for various weight functions.

The first graph, Figure 2.2.11, gives simulated Type 1 error rates for the different weight functions for this procedure. We see that T4 is the most conservative weight function. It is able to maintain the level of alpha <0.05 for values of ρ up to 0.17. The least conservative, T5, maintains the level of alpha <0.05 only for values of ρ up to 0.12.

The second graph, Figure 2.2.12, shows the power curves for the different weight functions. At ρ=0.17, the most conservative weight function, T4, has a power of about 0.33. At ρ=0.12, the least conservative weight function, T5, has a power of about 0.43. It can be seen that T1 and T3 result in virtually identical p-values, as do T2 and T4.

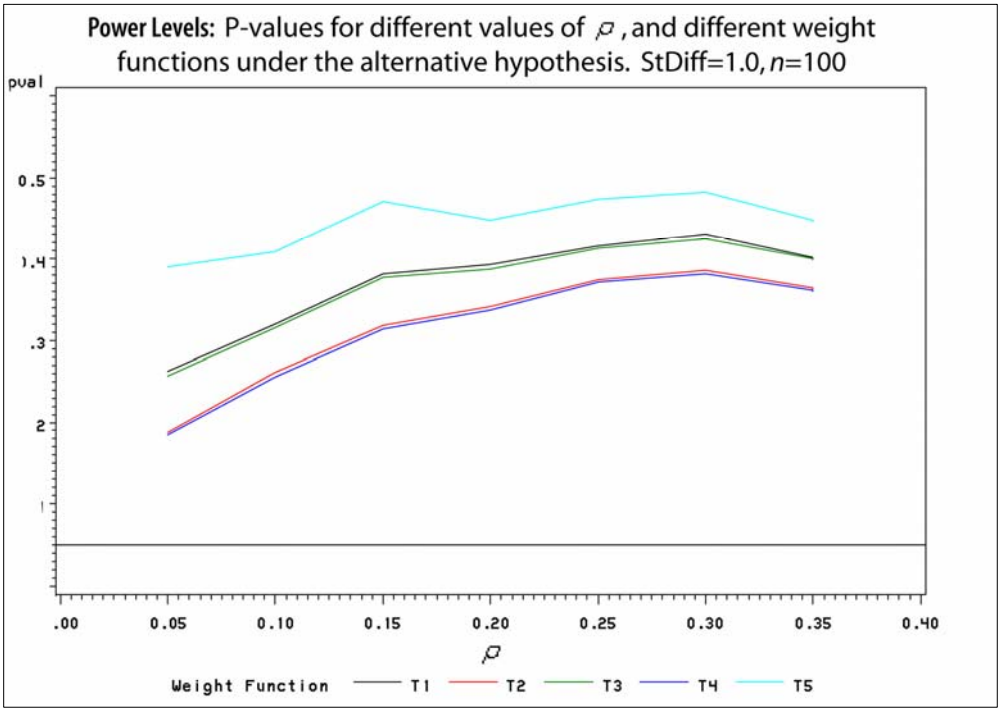


Figure 2.2.12. Power levels for various weight functions.

The next graph, Figure 2.2.13, allows comparison of the weighted p-value method and the plug-in method. Weighted p-values are graphed as well as p-values obtained using the plug-in method.

According to Figure 2.2.13, the most conservative approach is the plug-in method and using a very large value as ρ_0. Choosing ρ_0=0.25, using the plug-in method, maintains a significance level below 0.05 for values of ρ less than 0.25. The least conservative approach is the plug-in method, using a very small value for ρ_0. Choosing ρ_0=0.10, using the plug-in method, maintains a significance level below 0.05 for values of ρ less than 0.10. It can be noted that T1, T3, and ρ_0=0.15 all result in virtually the same p-values. That is due to the fact that the distribution of p-values as a function of ρ_0 is fairly linear over the small interval covered by

the weight function, so it is similar to using ρ_0 at the average weight. In T1 and T3, the average weight is 0.15.

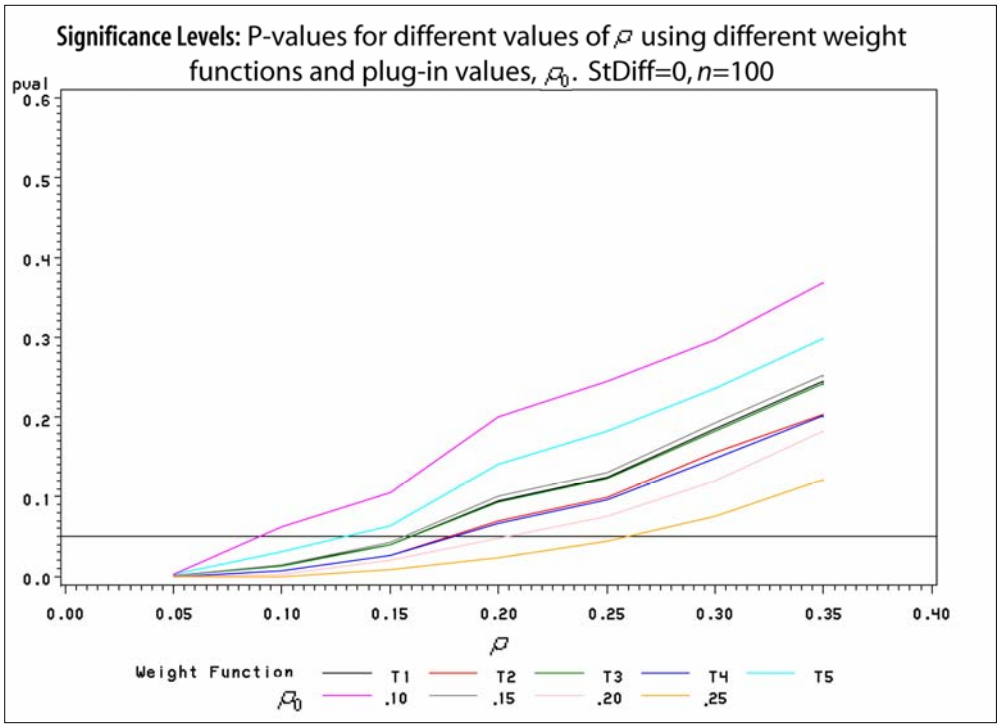


Figure 2.2.13. Significance levels for various weight functions and plug-in values.

Figure 2.2.14 shows the greater power levels are found using small plug-in values and weight functions whose average weight is small. The level of power consistently decreases as the plug-in value or weight function average increases.

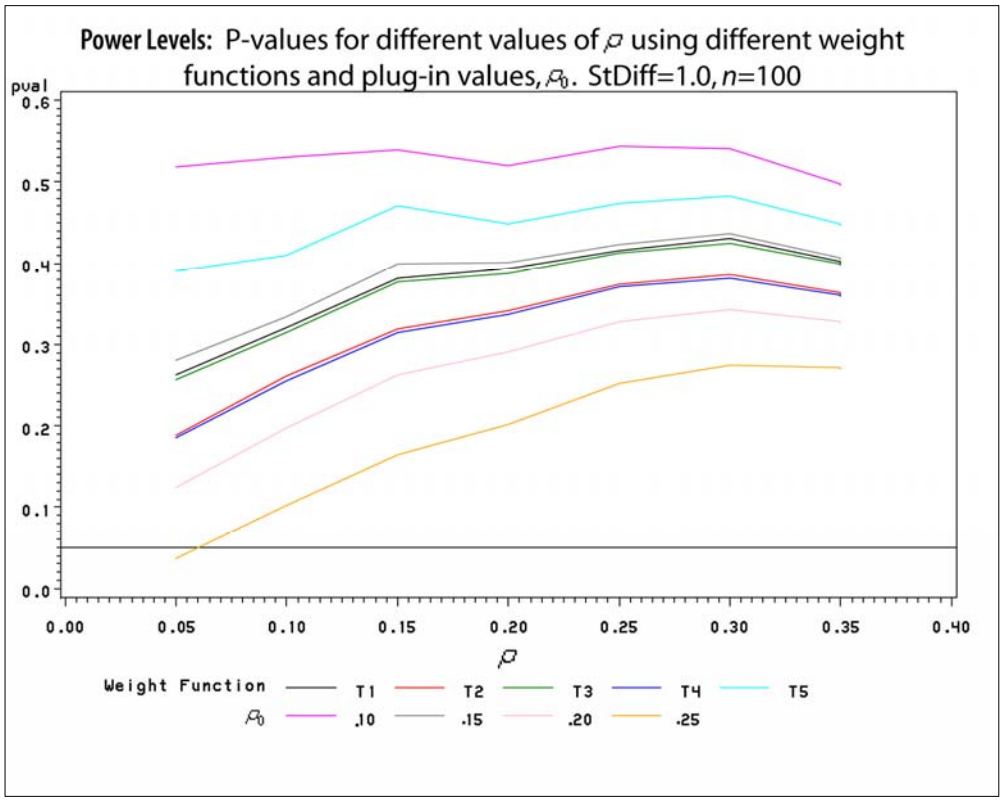


Figure 2.2.14. Power levels for various weight functions and plug-in values.

Implementation

To perform the analysis using the weighted p-value method, first determine at which values of ρ ($\rho_{01} \ldots \rho_{05}$, say) the p-values will be evaluated. Use the SAS® code presented for analyzing data using the plug-in method shown earlier to obtain a p-value using each of the chosen values ρ_{0i}. Multiply each of the p-values ($p_1 \ldots p_5$, say) by an associated weight ($w_1 \ldots w_5$, say). Sum the weighted p-values ($\sum_{i=1}^{5} w_i p_i$, say).

Random ρ

We now look at average significance level and power for the various strategies by randomly selecting ρ from among its probable values and determine the proportion of times our test correctly rejects or fails to reject the null hypothesis.

A simulation was conducted with the following specifications:

Iterations = 100,000
n = 5, 10, 100
Error d.f. = Satterthwaite-type and 2(n-1)
$\rho_0 = \overline{\rho}$ and 75th quantile
ρ ~Beta(1,10), Beta(5,30), Uniform(0, 0.5), Beta(10,4)

The following distributions show plausible prior information about ρ and the consequences on the average significance and power levels.

The first example represents the belief of the researcher that the smaller possible values of ρ are the most likely. This would be a situation of near independence among students in a class. With the Satterthwaite-type error degrees of freedom, the significance levels are kept well below the nominal level. Power values are also very low. The Satterthwaite-type analysis tends to be very conservative. With $\rho_0 = \overline{\rho}$ and 2(n-1) error degrees of freedom, the average significance level is very close to the nominal level, and power is much greater than in the Satterthwaite-type analysis. Equating ρ_0 to the 75th quantile simulates the situation of the Max Rho method where the value used for ρ_0 is reasonably close to the maximum. This choice of ρ_0 was able to maintain a significance level below the nominal level, but had greater power than the Satterthwaite-type approach.

The second example represents the belief of the researcher that the values of ρ are most likely between 0.05 and 0.25, with more weight on the lesser values. The results are similar to those of the first example.

The third example demonstrates why the researcher must have strong prior information about ρ in order to produce a useful analysis. With this non-informative prior distribution, the impression is that the researcher knows nothing about possible values of ρ, except that it is less than 0.5, and so gives equal probability to all possible values of ρ. The power of the tests is unacceptably low and thus may put any subsequent analysis into question.

Example 1: ρ ~Beta(1,10)

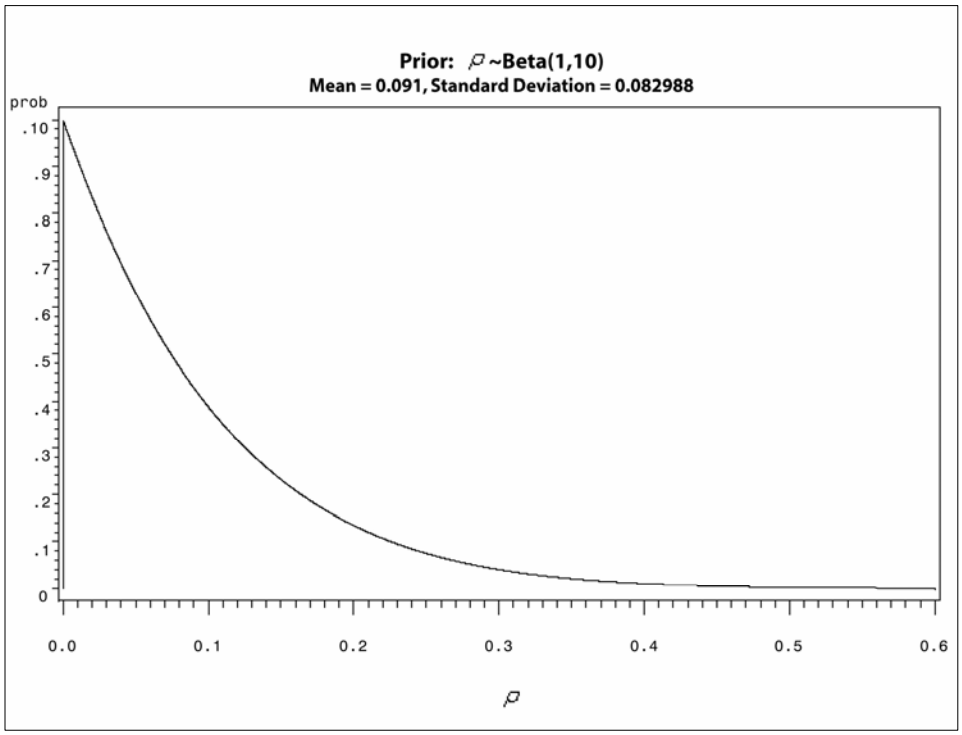


Figure 2.2.15. Prior distribution of ρ, Beta(1,10).

Table 2.2.1. Simulated significance levels and power levels using ρ ~Beta(1,10) prior distribution; n = 5, 10, 100; error d.f. = Satterthwaite-type and 2(n-1); $\rho_0 = \overline{\rho}$ and 75th quantile.

$\rho_0 = \overline{\rho}$	n=5	n=10	n=100
Satterthwaite d.f. =	4.451	4.034	1.864
Simulated Alpha =	0.029	0.018	0.002
Simulated Power(diff=1) =	0.134	0.134	0.012
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.050	0.053	0.061
Simulated Power(diff=1) =	0.200	0.299	0.533
$\rho_0 = 75^{th}$ *Quantile*	n=5	n=10	n=100
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.040	0.039	0.041
Simulated Power(diff=1) =	0.172	0.243	0.383

Example 2: ρ ~Beta(5,30)

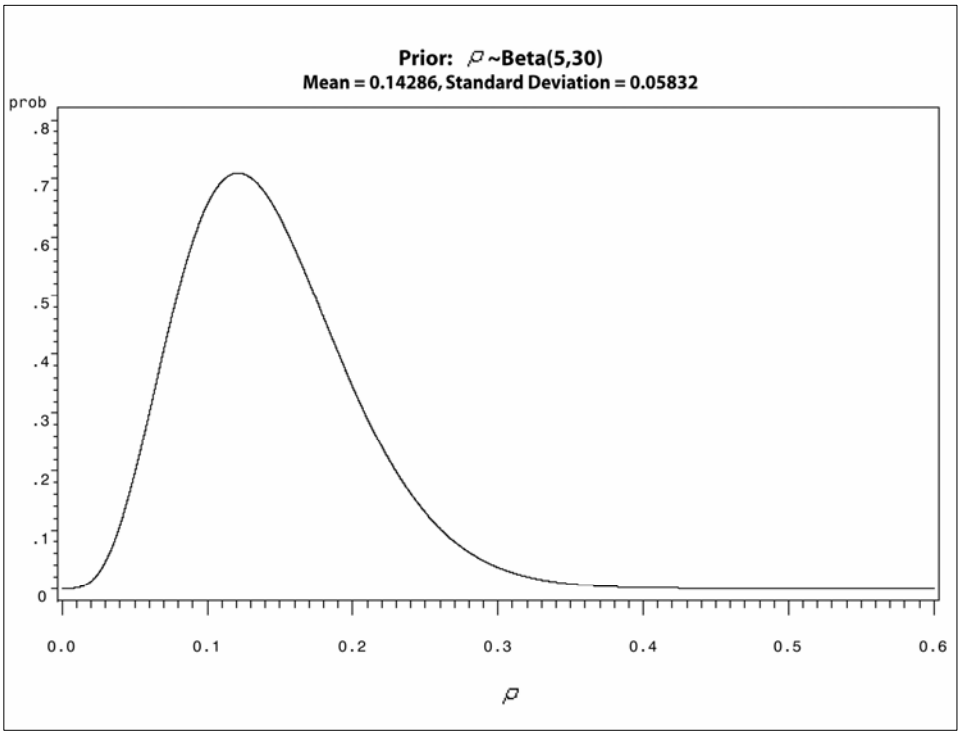


Figure 2.2.16. Prior distribution of ρ, Beta(5,30).

Table 2.2.2. Simulated significance levels and power levels using ρ ~Beta(5,30) prior distribution; n = 5, 10, 100; error d.f. = Satterthwaite-type and 2(n-1); $\rho_0 = \overline{\rho}$ and 75th quantile.

$\rho_0 = \overline{\rho}$	n=5	n=10	n=100
Satterthwaite d.f. =	6.348	9.124	8.724
Simulated Alpha =	0.043	0.039	0.030
Simulated Power(diff=1) =	0.155	0.205	0.255
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.050	0.051	0.054
Simulated Power(diff=1) =	0.175	0.246	0.367
$\rho_0 = 75^{th}$ *Quantile*	n=5	n=10	n=100
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.041	0.038	0.034
Simulated Power(diff=1) =	0.151	0.202	0.280

Example 3: ρ ~Uniform(0,0.5) – non-informative

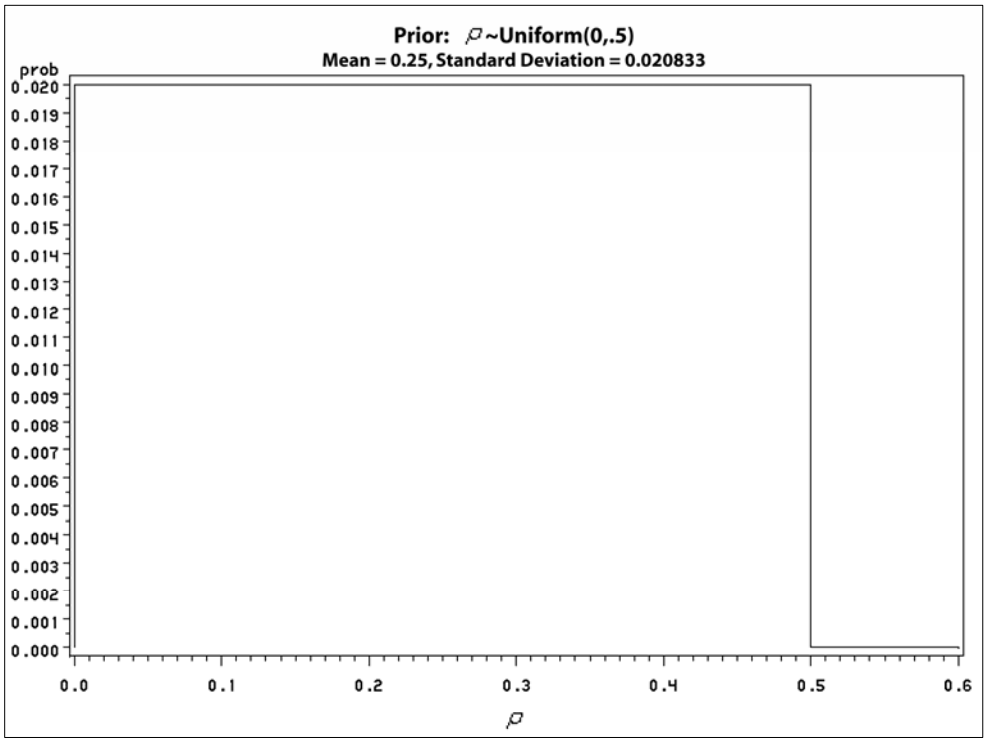


Figure 2.2.17. Prior distribution of ρ, Uniform(0,0.5).

Table 2.2.3. Simulated significance levels and power levels using ρ ~Uniform(0,0.5) prior distribution; n = 5, 10, 100; error d.f. = Satterthwaite-type and 2(n-1); $\rho_0 = \overline{\rho}$ and 75th quantile.

$\rho_0 = \overline{\rho}$	n=5	n=10	n=100
Satterthwaite d.f. =	3.744	4.186	3.888
Simulated Alpha =	0.026	0.021	0.015
Simulated Power(diff=1) =	0.063	0.063	0.051
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.053	0.056	0.062
Simulated Power(diff=1) =	0.123	0.149	0.171
$\rho_0 = 75^{th}$ *Quantile*	n=5	n=10	n=100
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.032	0.029	0.028
Simulated Power(diff=1) =	0.077	0.084	0.087

Example 4: ρ *~Beta(10,4)*

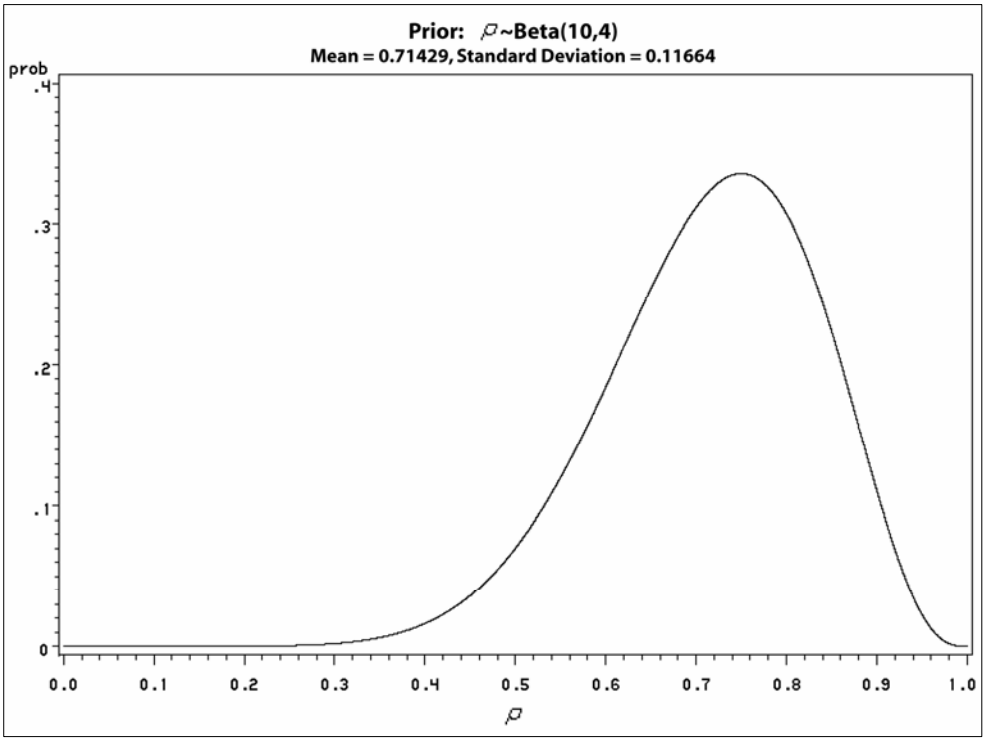


Figure 2.2.18. Prior distribution of ρ, Beta(10,4).

Table 2.2.4. Simulated significance levels and power levels using ρ ~Beta(10,4) prior distribution; n = 5, 10, 100; error d.f. = Satterthwaite-type and 2(n-1); $\rho_0 = \overline{\rho}$ and 75th quantile.

$\rho_0 = \overline{\rho}$	n=5	n=10	n=100
Satterthwaite d.f. =	2.055	2.526	3.018
Simulated Alpha =	0.005	0.006	0.008
Simulated Power(diff=1) =	0.006	0.009	0.009
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.052	0.053	0.054
Simulated Power(diff=1) =	0.062	0.065	0.068
$\rho_0 = 75^{th}$ *Quantile*	n=5	n=10	n=100
2(n-1) d.f. =	8	18	198
Simulated Alpha =	0.040	0.040	0.039
Simulated Power(diff=1) =	0.047	0.049	0.049

The fourth example demonstrates why the methods explained in this research only apply when the value of ρ is small, usually less than 0.5. As the values of ρ approach 1.0, the denominator of the test statistic in Equation 2.1.1 approaches ∞, which causes the test statistic to approach zero. The power of the tests is again unacceptably low and again puts into question any subsequent analysis.

From this second method of evaluation we arrive at the following recommendations for the researcher when distributional information is known about ρ:

1. Only use the plug-in method if strong information is known about possible values of ρ and those values are below 0.5.

2. If the researcher must maintain a significance level below the nominal level, using $\rho_0 = \overline{\rho}$ with a Satterthwaite-type adjustment would be a conservative safe choice.

3. If the researcher desires to maintain a significance level close to the nominal level and wishes to maximize power, using $\rho_0 = \overline{\rho}$ with *2(n-1)* degrees of freedom is a good choice.

The Distribution of T with Random ρ

The test statistic for testing the difference between two treatment means in a two-treatment no-replication case, $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$, was previously given (*2.1.4*) as

$$T = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_{01} - \mu_{02})}{\sqrt{2\hat{\sigma}_\varepsilon^2 \left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n} \right)}}$$

Knowing the distribution of the test statistic T is important because it allows us compute, mathematically, the significance level and power of tests.

The test statistic *T* can be rewritten as

$$T = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}} \cdot \frac{\sqrt{\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}{\sqrt{\left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)}} \qquad (2.2.3)$$

Under the null hypothesis with ρ random, we can define this as a random variable

$$T = T_0 Q \qquad (2.2.4)$$

where

$$T_0 = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}} \sim t_{0.05,2(n-1)} \qquad (2.2.5)$$

with mean=0 and variance=$\frac{n-1}{n-2}$ and

$$Q = \frac{\sqrt{\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}{\sqrt{\left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)}} \qquad (2.2.6)$$

Note that the distribution of T_0 is independent of ρ and hence independent of Q. Although the density function of T may be messy and difficult to compute, the first two moments are found quite easily.

$$E(T) = E(T_0 Q) = E[E(T_0 Q \mid \rho)] = E[QE(T_0 \mid \rho)] = E[Q(0)] = 0 \qquad (2.2.7)$$

$$Var(T) = Var(T_0 Q) = Var[E(T_0 Q \mid \rho)] + E[Var(T_0 Q \mid \rho)]$$

$$= Var[0] + E[Var(T_0 Q \mid \rho)]$$

$$= E\left[Var(T_0 Q \mid \rho)\right] = E\left[Q^2 Var(T_0 \mid \rho)\right] = E\left[Q^2\left(\frac{n-1}{n-2}\right)\right]$$

$$= \left(\frac{n-1}{n-2}\right)E\left[Q^2\right] = \left(\frac{n-1}{n-2}\right)\left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right)E\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right) \qquad (2.2.8)$$

If ρ_0 is chosen such that $\frac{\rho_0}{1-\rho_0} = E\left(\frac{\rho}{1-\rho}\right)$ then $Var(T) = \frac{n-1}{n-2} = Var(T_0)$, consequently T has the first two moments of $T_0 \sim t_{0.05,2(n-1)}$. This fact lends credence to the use of the full $2(n-1)$ error degrees of freedom when testing hypotheses using the plug-in method. It also allows us to use $T_0 \sim t_{0.05,2(n-1)}$ to approximate the behavior (significance level and power of tests) of T.

Another way we can look at probabilities associated with the random variable T is to compare them with those of an F-distributed random variable indexed by 1 and $2(n-1)$ degrees of freedom. Let $T_0 \sim t_{0.05,2(n-1)}$. Then $F_0 = (T_0)^2 \sim f_{0.05,1,2(n-1)}$. We consider the distribution of $F = T^2 = T_0^2 Q^2$ where ρ has a Beta distribution.

If $\rho \sim Beta(\alpha, \beta)$, the distribution function of $Y = Q^2$ is

$$f_Y(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} c\left[cy - \frac{1}{n}\right]^{\alpha-1}\left[1 + cy - \frac{1}{n}\right]^{-(\alpha+\beta)}, \quad \frac{1}{cn} < y < \infty \qquad (2.2.9)$$

where $c = \frac{\rho_0}{1-\rho_0} + \frac{1}{n}$.

We then expect $T^2 = (T_0 Q)^2 = F_0 Q^2$ to have a distribution close to $f_{0.05,1,2(n-1)}$. The mean and variance may be computed as follows:

$$E\left(T^2\right)=E\left(F_0Q^2\right)=E\left[E\left(F_0Q^2\mid\rho\right)\right]=\left(\frac{n-1}{n-2}\right)\left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right)E\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right) \qquad (2.2.10)$$

$$Var\left(T^2\right)=Var\left(F_0Q^2\right)=E\left[Var\left(F_0Q^2\mid\rho\right)\right]+Var\left[E\left(F_0Q^2\mid\rho\right)\right]=$$

$$\frac{(n-1)^2}{(n-2)^2(n-3)}\left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right)^2\left\{Var\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)3(n-2)+(2n-3)\left[E\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)\right]^2\right\} \qquad (2.2.11)$$

If $\rho\sim Beta(\alpha,\beta)$, the distribution function of $X=\frac{\rho}{1-\rho}$ is

$$f_X(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}\left(\frac{1}{1+x}\right)^{\alpha+\beta},0<x<\infty \qquad (2.2.12)$$

which is the equivalent to $\frac{\alpha}{\beta}F$, a constant times an F-distributed random variable with $\upsilon_1=2\alpha$ and $\upsilon_2=2\beta$. The mean and variance are

$$E\left(\frac{\rho}{1-\rho}\right)=\frac{\alpha}{\beta-1}$$

and

$$Var\left(\frac{\rho}{1-\rho}\right)=\frac{\alpha(\alpha+\beta-1)}{(\beta-1)^2(\beta-2)},$$

which leads to

$$E\left(T^2\right)=\left(\frac{n-1}{n-2}\right)\left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right)\left(\frac{\alpha}{\beta-1}+\frac{1}{n}\right) \tag{2.2.13}$$

$$Var\left(T^2\right)=$$

$$\frac{(n-1)^2}{(n-2)^2(n-3)}\left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right)^2\left\{\frac{\alpha(\alpha+\beta-1)}{(\beta-1)^2(\beta-2)}3(n-2)+(2n-3)\left(\frac{\alpha}{\beta-1}+\frac{1}{n}\right)^2\right\} \tag{2.2.14}$$

Also, if ρ_0 is chosen such that $\frac{\rho_0}{1-\rho_0}=E\left(\frac{\rho}{1-\rho}\right)$ then

$$E\left(T^2\right)=\left(\frac{n-1}{n-2}\right)=E\left(F_0\right)$$

and if $Var\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)$ is small then

$$Var\left(T^2\right)\approx\left(\frac{n-1}{n-2}\right)^2\left(\frac{2n-3}{n-3}\right)=Var\left(F_0\right).$$

Consequently T^2 has approximately the first two moments of $F_0 \sim f_{1,2(n-1)}$. This also supports the use of the full *2(n-1)* error degrees of freedom when testing hypotheses using the plug-in method and allows us to use $F_0 \sim f_{1,2(n-1)}$ to approximate the behavior (significance level and power of tests) of T^2.

The following table (2.2.5) shows the results of simulation of certain percentiles of the $F_0 \sim f_{1,2(n-1)}$ distribution and $T^2=\left(T_0Q\right)^2=F_0Q^2$. Exact percentiles from the F-distribution

were computed using Minitab (version 11), and T^2 percentiles were computed from a generation of 1,000,000 random variates of the T^2 distribution.

Table 2.2.5. Simulated percentiles of the T^2 distribution and actual percentiles of the $f_{1,2(n-1)}$ distribution.

			Percentiles					
Dist. of ρ	n	Random Variable	50th	75th	90th	95th	99th	P(T^2>95th)
Beta(1,10)	10	F	0.474	1.413	3.007	4.414	8.285	0.0717
		T^2	0.432	1.347	3.076	4.798	10.470	0.0500
	1,000,000	F	0.455	1.323	2.706	3.842	6.635	0.0713
		T^2	0.252	1.028	2.896	4.890	12.316	0.0500
Beta(5,30)	10	F	0.474	1.413	3.007	4.414	8.285	0.0536
		T^2	0.457	1.389	3.051	4.578	9.008	0.0500
	1,000,000	F	0.455	1.323	2.706	3.842	6.635	0.0576
		T^2	0.404	1.249	2.775	4.175	8.210	0.0500
Beta(10,4)	10	F	0.474	1.413	3.007	4.414	8.285	0.0980
		T^2	0.493	1.586	3.784	6.049	13.844	0.0500
	1,000,000	F	0.455	1.323	2.706	3.842	6.635	0.0859
		T^2	0.468	1.481	3.441	5.405	11.855	0.0500

Three prior distributions are represented in Table 2.2.5. The first gives most probability to values of ρ close to zero. The second gives most probability to values between 0.1 and 0.3. The third gives most probability to values of ρ above 0.5. For Beta(1,10) and Beta(5,30) each prior distribution, the percentiles of T^2 closely match those of $F \sim f_{1,2(n-1)}$ up to the 95th percentile. The percentiles don't match as well with Beta(10,4). The last column in the table contains the probability of the random variable T^2 is greater than the 95th percentile, indicating the probability of rejecting the null hypothesis using a critical value from the F-distribution and a T^2 test statistic. So, using the F-distribution to compute critical values will

slightly inflate the probability of a Type 1 error when using one of the suggested priors and a 0.05 nominal significance level.

2.3: THE T-TREATMENT CASE

It is reasonable to assume the researcher may want to work with more than two treatments. Again, we look at the Completely Randomized design with subsampling and no replication.

Description

Consider the model (*1.1.1*) defined in section 1:

$$y_{ijk} = \mu_i + \delta_{ij} + \varepsilon_{ijk}$$

This model reduces to $y_{ik} = \mu_i + \delta_i + \varepsilon_{ik}$ in the no replication case with subscript j supressed. This model can be written in matrix form. Let $N = \sum_{i=1}^{t} n_i$. Let Y be an $N \times 1$ observable vector of random variables; X be a $N \times t$ design matrix, that is a matrix of 0's and 1's denoting the design's treatment structure; β be a $t \times 1$ vector consisting of the treatment means μ_1, μ_2, …, μ_t; E be a $N \times 1$ vector of unknown random errors. Then we can write the generalized linear model as

$$Y = X\beta + E \qquad (2.3.1)$$

Let H be a $q \times t$ matrix and h be a $q \times 1$ vector, as defined in Graybill (1976) p. 184. The researcher is interested in testing hypotheses of the form

H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$. (*2.3.2*)

Let V be an $N \times N$ known positive definite matrix such that $Var(Y) = \sigma_\varepsilon^2 V$.

$$V = \begin{bmatrix} A_1 & 0 & \dots & \dots & 0 \\ 0 & A_2 & \dots & \dots & 0 \\ . & & . & & . \\ . & & & . & . \\ 0 & 0 & \dots & & A_t \end{bmatrix}, \; A_i = \frac{1}{1-\rho}\begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} \tag{2.3.3}$$

A_i is an $n_i \times n_i$ matrix.

Let $\hat{\beta} = \left(X'V^{-1}X\right)^{-1}X'V^{-1}Y$ be the best linear unbiased estimate of β. This is equal to $\hat{\beta} = [\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \dots, \bar{y}_{t\cdot}]$ in the no replication case. The estimate of σ_ε^2 is

$$\hat{\sigma}_\varepsilon^2 = \left(\frac{1}{N-t}\right)Y'\left[V^{-1} - V^{-1}X\left(X'V^{-1}X\right)^{-1}X'V^{-1}\right]Y \tag{2.3.4}$$

which in the no replication case reduces to

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{t}(n_i - 1)s_i^2}{\sum_{i=1}^{t}(n_i - 1)} \tag{2.3.5}$$

where

$$s_i^2 = \frac{\sum_{k=1}^{n_i}(y_{ik} - \bar{y}_{i\cdot})^2}{n_i - 1}. \tag{2.3.6}$$

Let V_0 denote the matrix V with $\rho = \rho_0$. The test statistic for testing H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$ is of the form

$$W = \frac{\left(H\hat{\beta} - h\right)'\left[H\left(X'V_0^{-1}X\right)^{-1}H'\right]^{-1}\left(H\hat{\beta} - h\right)}{q\hat{\sigma}_\varepsilon^2} \tag{2.3.7}$$

Reject H_0 at level α_0 when $W > f_{\alpha_0,q,t(n-1)}$, where $f_{\alpha_0,q,t(n-1)}$ is the $1-\alpha_0$ quantile of the F distribution with *q* and *t(n-1)* degrees of freedom.

In an analysis of variance situation, the researcher may be interested in performing F-tests to compare different combinations of treatments. The researcher may also wish to perform multiple comparison tests on different treatments. The methodology for using a plug-in value ρ_0 is similar in the t-treatment case to the methodology in the two-treatment case.

Properties

For simplicity, we consider the case where $n = n_i$ for all values of i. It can be shown that

$$\left(X'V_0^{-1}X\right)^{-1} = \left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)I_t = \left[\frac{1+(n-1)\rho_0}{(1-\rho_0)}\right]\times\left(\frac{1}{n}\right)I_t . \tag{2.3.8}$$

$$= d_0\left(X'X\right)^{-1}$$

where

$$d_0 = \frac{1+(n-1)\rho_0}{(1-\rho_0)} . \tag{2.3.9}$$

Let $\upsilon = t(n-1)$. The probability of rejection of the null hypothesis can be computed as follows:

$$P\left(W > f_{0.05,q,\upsilon}\right) = P\left(W^* > \left(\frac{d_0}{d}\right)f_{0.05,q,\upsilon}\right) = P\left(W^* > \frac{\left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)}{\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)} f_{0.05,q,\upsilon}\right) \tag{2.3.10}$$

where *W** has a non-central F distribution with *q* and υ degrees of freedom with non-centrality parameter

$$\lambda = \frac{1}{\sigma_\varepsilon^2}(H\beta - h)'\left[H\left(X'V^{-1}X\right)^{-1}H'\right]^{-1}(H\beta - h). \qquad (2.3.11)$$

Increasing ρ_0 decreases (*2.3.10*) for fixed ρ.

We now look at the effect of the number of treatments on the significance level. The following graphs (2.3.1 and 2.3.2) display significance levels and power curves for testing the equality of multiple treatment means in no replication settings. The error degrees of freedom are held constant. The number of students per class is determined such that the total error degrees of freedom (*t(n-1)*) is equal to 96. The true value ρ is set at 0.1, a likely value for many applications, and the values of ρ_0 range from 0.025 to 0.20 to demonstrate the consequence of discrepancy between the researchers choice, ρ_0, and the actual value, ρ.

As can be seen in Figure 2.3.1, in the no replication case, as the number of treatment levels increases, so does the significance level for underestimates of ρ (ρ_0=0.025, 0.05). Also, the choice of ρ_0 becomes more critical for a large number of treatment levels. However, it can also be seen that when overestimating ρ, the significance level remains nearly constant across all numbers of treatment levels.

Figure 2.3.2 indicates the power levels are fairly constant, but drop slightly as the number of treatments increases in the no replication case.

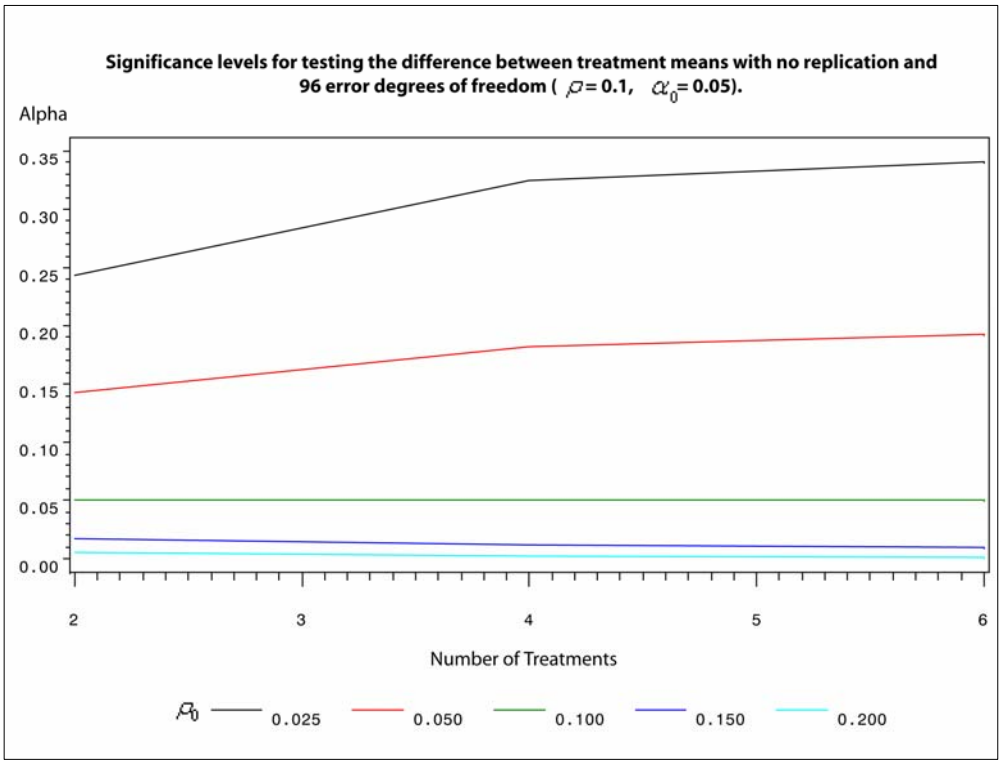


Figure 2.3.1. Significance levels for tests with t treatment levels and 96 error degrees of freedom.

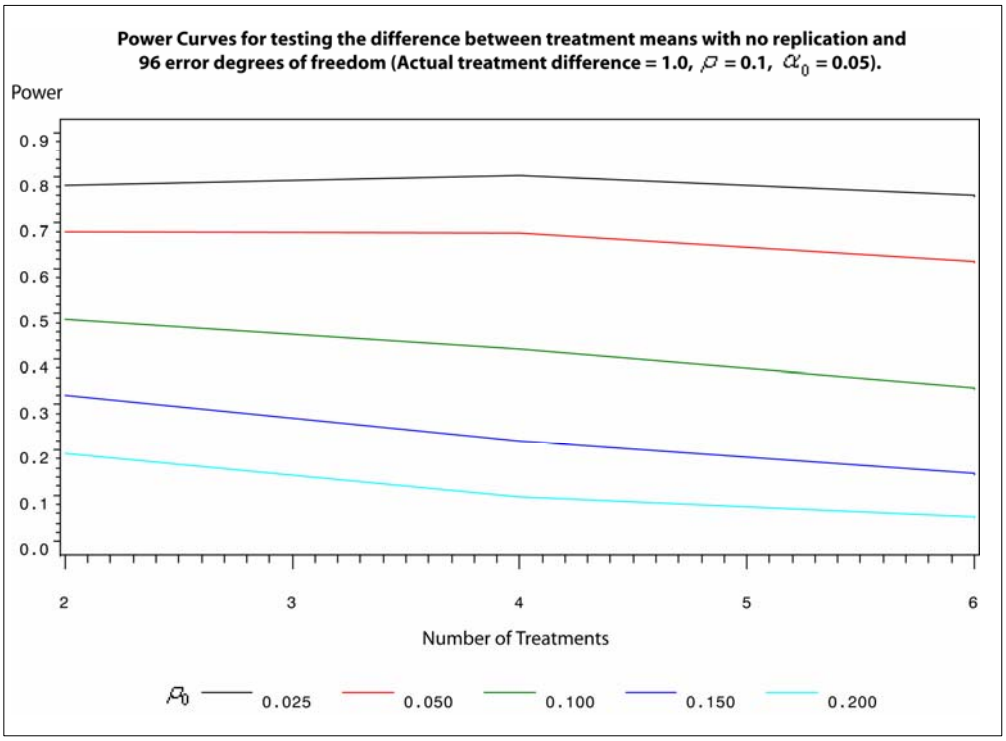


Figure 2.3.2. Power curves for tests with t treatment levels and 96 error degrees of freedom.

The Distribution of F when ρ is Random

The test statistic for testing H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$ was previously given in Equation 2.3.7 as

$$W = \frac{\left(H\hat{\beta} - h\right)'\left[H\left(X'V_0^{-1}X\right)^{-1}H'\right]^{-1}\left(H\hat{\beta} - h\right)}{q\hat{\sigma}_\varepsilon^2}$$

Knowing the distribution of the test statistic W is important because it allows us compute, mathematically, the significance level and power of tests. Under the null hypothesis we can define a random variable

$$W = W_0 Q^2 \qquad (2.3.12)$$

where

$$W_0 = \frac{\left(H\hat{\beta} - h\right)'\left[H\left(X'X\right)^{-1}H'\right]^{-1}\left(H\hat{\beta} - h\right)}{Y'\left[I - X\left(X'X\right)^{-1}X'\right]Y}\left(\frac{N-t}{q}\right) \sim f_{0.05,q,t(n-1)} \qquad (2.3.13)$$

and Q is defined by Equation 2.2.6, where ρ is a random variable. The distribution of Q is defined by Equation 2.2.9. The distribution of W_0 is independent of ρ and hence independent of Q. We then expect $W = W_0 Q^2$ to have a distribution close to $f_{0.05,q,t(n-1)}$. The mean and variance are computed as follows:

$$E(W) = E\left(W_0 Q^2\right) = E\left[E\left(W_0 Q^2 \mid \rho\right)\right] = \left[\frac{t(n-1)}{t(n-1)-2}\right]\left(\frac{1}{\frac{\rho_0}{1-\rho_0} + \frac{1}{n}}\right)E\left(\frac{\rho}{1-\rho} + \frac{1}{n}\right) \qquad (2.3.14)$$

$$Var(W) = Var(W_0 Q^2) = E[Var(W_0 Q^2 \mid \rho)] + Var[E(W_0 Q^2 \mid \rho)] =$$

$$\left[\frac{t(n-1)}{t(n-1)-2}\right]^2 \left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right)^2 \left\{\frac{q+t(n-1)-2}{q[t(n-1)-4]} 2E\left[\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)^2\right] + Var\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)\right\} \quad (2.3.15)$$

If $\rho \sim Beta(\alpha, \beta)$, the distribution function of $X = \frac{\rho}{1-\rho}$ is defined by Equation 2.2.12 with mean and variance

$$E\left(\frac{\rho}{1-\rho}\right) = \frac{\alpha}{\beta-1}$$

and

$$Var\left(\frac{\rho}{1-\rho}\right) = \frac{\alpha(\alpha+\beta-1)}{(\beta-1)^2(\beta-2)}.$$

This leads to

$$E(W) = \left[\frac{t(n-1)}{t(n-1)-2}\right] \left(\frac{1}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right) \left(\frac{\alpha}{\beta-1}+\frac{1}{n}\right) \quad (2.3.16)$$

$$Var(W) = \left[\frac{\frac{t(n-1)}{t(n-1)-2}}{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}\right]^2 \left\{\left(2\left[\frac{q+t(n-1)-2}{q[t(n-1)-4]}\right]+1\right)\left[\frac{\alpha(\alpha+\beta-1)}{(\beta-1)^2(\beta-2)}\right] + 2\left[\frac{\alpha}{\beta-1}+\frac{1}{n}\right]^2\right\}$$

$$(2.3.17)$$

Also, if ρ_0 is chosen such that $\frac{\rho_0}{1-\rho_0} = E\left(\frac{\rho}{1-\rho}\right)$ then

$$E(W) = \frac{t(n-1)}{t(n-1)-2} = E(W_0)$$

and if $Var\left(\frac{\rho}{1-\rho} + \frac{1}{n}\right)$ is small then

$$Var(W) \approx 2\left[\frac{t(n-1)}{t(n-1)-2}\right]^2\left[\frac{q+t(n-1)-2}{q[t(n-1)-4]}\right] = Var(W_0).$$

Consequently W has approximately the first two moments of $W_0 \sim f_{q,t(n-1)}$.

Table 2.3.1 lists the percentiles of W and W_0, where $\rho \sim Beta(5,30)$, for different values of t (number of treatments) and n (number of students per treatment).

Table 12.3.1. Simulated percentiles of the W distribution and actual percentiles of the $f_{q,t(n-1)}$ distribution.

			Percentiles					
t	n	Random Variable	50th	75th	90th	95th	99th	P(W>95th)
2	10	W_0	0.474	1.413	3.007	4.414	8.285	0.053
		W	0.458	1.392	3.043	4.571	9.016	0.050
	1,000,000	W_0	0.457	1.331	2.731	3.889	6.765	0.058
		W	0.404	1.251	2.766	4.161	8.142	0.050
4	10	W_0	0.819	1.494	2.416	3.160	5.092	0.061
		W	0.776	1.433	2.347	3.081	5.023	0.050
	1,000,000	W_0	0.791	1.380	2.112	2.650	3.882	0.076
		W	0.708	1.361	2.308	3.088	5.130	0.050
6	10	W_0	0.904	1.464	2.196	2.773	4.248	0.067
		W	0.851	1.383	2.077	2.617	3.984	0.050
	1,000,000	W_0	0.871	1.329	1.857	2.229	3.048	0.090
		W	0.788	1.354	2.123	2.738	4.300	0.050

The percentiles for *W* in Table 2.3.1 closely match those of $W_0 \sim f_{q,t(n-1)}$ up to the 95th percentile. It would not be unreasonable to use percentiles of the $f_{0.05,q,t(n-1)}$ distribution to find critical values for test statistics based on *W*.

T-Treatment Strategies for Choosing ρ_0

The same strategies (Maximum Rho, Weighted P-vlaue, etc.) used in the two-treatment case extend to the t-treatment case. One difference worthy of note is the formula for the Satterthwaite-type error degrees of freedom calculation in the t-treatment case.

In testing hypotheses of the form H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$, the Satterthwaite-type adjustment is of the form

$$d.f. = \frac{2\left[w\sigma_\varepsilon^2\right]^2}{\hat{Var}\left[w\sigma_\varepsilon^2\right]}. \qquad (2.3.18)$$

where

$$w = \frac{\rho}{1-\rho} + \frac{1}{n}$$

$$\hat{Var}\left[w\sigma_\varepsilon^2\right] = Var\left[E\left(w\sigma_\varepsilon^2 \mid \rho\right)\right] + E\left[Var\left(\sigma_\varepsilon^2 w \mid \rho\right)\right] \qquad (2.3.19)$$

$$= \sigma_\varepsilon^4 Var(w) + \frac{2\sigma_\varepsilon^4}{t(n-1)} E\left(w^2\right) = \frac{\sigma_\varepsilon^4}{t(n-1)}\left\{t(n-1)\cdot Var(w) + 2\left[E\left(w^2\right)\right]\right\}$$

$$= \frac{\sigma_\varepsilon^4}{t(n-1)}\left\{t(n-1)\cdot Var(w) + 2Var(w) + \left[E(w)\right]^2\right\}$$

So,

$$d.f. = \frac{2w^2\sigma_\varepsilon^4}{\frac{\sigma_\varepsilon^4}{t(n-1)}\left\{t(n-1)\cdot Var(w) + 2Var(w) + 2[E(w)]^2\right\}}$$

$$= \frac{2w^2 t(n-1)}{(tn-t+2)\cdot Var(w) + 2[E(w)]^2} \qquad (2.3.20)$$

and is estimated by

$$d.f. = \frac{2[E(w)]^2 t(n-1)}{(tn-t+2)\cdot Var(w) + 2[E(w)]^2}. \qquad (2.3.21)$$

As $Var(W) \to 0$, the error degrees of freedom using the Satterthwaite-type adjustment approaches $t(n-1)$, the full d.f. As $n \to \infty$, the error degrees of freedom using the Satterthwaite-type adjustment approaches

$$d.f.(Satterthwaite-type) = 2\left\{\frac{\left[E\left(\frac{\rho}{1-\rho}\right)\right]^2}{Var\left(\frac{\rho}{1-\rho}\right)}\right\}.$$

Multiple Comparison Tests

Let α be the probability of a type 1 error for testing the difference between treatment means. Suppose some number *g* is the number of comparisons to be made. If no adjustment is made, the error rate for all comparisons as a whole, known as the family-wise error rate (FWE) for a certain family (group) of comparisons, is at least α, and at most $g \times \alpha$. If the researcher is interested in maintaining an FWE of α, this can be done by making a multiple comparisons/multiple testing adjustment. See Westfall, Tobias, Rom, Wolfinger, & Hochberg, (1999) for more information on the FWE. We illustrate the use of the plug-in method in the

following multiple comparison procedures: Bonferroni, Fisher's Protected LSD, Tukey's, and SAS®'s Simulate Procedures. We describe the methods, examine properties, and demonstrate implementation, using examples in the case of no replication. These procedures apply to the under-replication case as well.

Bonferroni

The Bonferroni method of adjusting for multiple comparisons is a method of strictly controlling the FWE by keeping it $\leq \alpha_0$. This is done by concluding a pairwise difference is significant if its corresponding p-value, multiplied by the number of pairwise comparisons to be made, *g*, is $\leq \alpha_0$. No adjustment is made to the computation of the test statistic(s) when using the Bonferroni method.

Fisher's Protected LSD

Fisher's method is a two-step method. The first step is a composite hypothesis testing H_0: All of the treatment means are equal vs. H_A: At least one of the treatment means differs from at least one of the other treatment means. If that composite H_0 is not rejected, the test is finished and the difference between all treatment means is considered insignificant. If H_0 is rejected, it is followed up by a second step, which makes pairwise comparisons of each of the treatment means. Although criticized for not controlling the FWE in many cases (Westfall, Tobias, Rom, Wolfinger, & Hochberg, 1999, p. 20), Fisher's method is still widely used.

To conduct a test using Fisher's method, the composite null hypothesis is tested using an F-test as described at the beginning of this section. A plug-in value, ρ_0, is used in the V matrix in Equation 2.3.7. For the second step, a statistic, LSD, is computed from the variance of the difference in means as follows: $LSD = t_{\alpha_0/2, t(n-1)} \sqrt{\hat{\text{var}}\left(\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}\right)}$. If the absolute value of any of the pairwise differences exceeds *LSD*, the treatments are considered significantly different (Ostle & Malone, 1988, p. 317).

Tukey's Procedure

Tukey's method makes an FWE adjustment for all possible pairwise comparisons, whether or not the researcher is interested in making that many comparisons.

Tukey's method seeks to find the value of c_α such that

$$P\left(\max_{i,i'} \frac{|(\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'})|}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n}\right)}} \le c_\alpha \right) = 1 - \alpha \qquad (2.3.22)$$

If $Z_1, \ldots, Z_t$ are independent standard normal random variables, and V is a Chi-Square random variable with *df* degrees of freedom, then $Q_{t,df}^R = \max_{i,i'} \frac{|Z_i - Z_{i'}|}{\sqrt{V/df}}$ has the studentized range distribution indexed by *t* and *df* (Westfall et al., 1999, pp. 45-46), and $c_\alpha = q_{1-\alpha,t,t(n-1)}^R / \sqrt{2}$, where $q_{1-\alpha,t,t(n-1)}^R$ is the 1-α quantile of the studentized range distribution. Using the plug-in method, the test statistic for Tukey's method is

$$c_\alpha = \max_{i,i'} \frac{|(\bar{y}_i - \bar{y}_{i'})|}{\sqrt{2\hat{\sigma}_\varepsilon^2\left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n}\right)}}. \qquad (2.3.23)$$

The difference between the two treatment means is considered significant at the nominal level α_0 if the test statistic, $\sqrt{2}c_\alpha$, exceeds $q_{1-\alpha_0,t,t(n-1)}^R$, or equivalently if the p-vlaue, *p*, is smaller than α_0, where $p = P\left(\sqrt{2}c_\alpha > q_{1-\alpha_0,t,t(n-1)}^R\right)$.

Simulate Procedure

The Simulate procedure uses simulation-based methods to adjust the p-values such that a FWE error rate is maintained. Certain multiple comparison procedures, including Tukey's, are

based on a multivariate-t distribution with a correlation matrix R. In general, calculation of the multivariate-t quantile is intractable. However, in certain cases it is feasible to obtain approximations. In the case of Tukey's procedure, R has a certain symmetry that makes it tractable. In cases of comparing all pairwise differences with unequal sample sizes or differences between least squares means in many unbalanced designs there is not a structure of R that allows for the exact computations necessary for approximations. The Simulate procedure samples multivariate-t vectors from a distribution with the appropriate R and degrees of freedom parameters, and then obtains percentiles from the results. Sufficient samples are generated such that the exact multivariate-t percentile is accurate to within a specified degree of accuracy with a certain level of confidence. The Simulate procedure is performed in the SAS® MIXED procedure with the option ADJUST=SIMULATE in the LSMEANS statement (SAS Institute Inc., 1999, pp. 1546-1548).

Properties

The Bonferroni, Tukey, and Simulate procedures all maintain a maximum experiment-wise error rate under any complete or partial null hypothesis. The Fisher's Protected LSD does as well when there are three or fewer treatments. Otherwise, it simply maintains a comparison-wise error rate and not an experiment-wise error rate.

Implementation

The following code uses the *class data* (Appendix C.1) and implements multiple testing using the Bonferroni, Fisher's LSD, Tukey's, and Simulate methods in the SAS® MIXED procedure.

```
/** STEP 1 **/
%let p0=.1;  /** p0 = Plug-in ICC **/
%let gi=1;  /** gi = # of classes per treatment **/
%let ti=3;  /** ti = # of treatments **/

/** STEP 2 **/
proc iml;
RATIO=((&p0/(1-&p0))*I(&gi*&ti));
create gratio from RATIO;
append from ratio;
quit;
data gratio;set gratio;row=(_N_);run;
```

```
/** STEP 3 **/
proc mixed data=classdata ratio;
class class trt;
model y=trt/ddfm=kr;
random class(trt)/ gdata=gratio Ratios;
lsmeans trt/pdiff adjust=bon;
lsmeans trt/pdiff adjust=tukey;
lsmeans trt/pdiff adjust=simulate(cvadjust report);
run;quit;
```

The first step (STEP 1) creates global variables and sets their values so that these values can be used in procedures and data steps throughout the program. The MIXED procedure allows known ratios of variance components to be put in the model. The second step (STEP 2) creates a matrix and subsequently a data set containing the ratios that will be used for the matrix that represents the variance of the random effects. The third step (STEP 3) analyzes the data using the appropriate model. The GDATA= and RATIOS options indicate the data set GRATIO will be used in place of the traditional estimates of the matrix representing the variance of the random effects, known as the G matrix. The DDFM=KR option makes certain the test is using the correct degrees of freedom, adjusting them for the known ratios. The LSMEANS statements call for estimates of the least squares means, and tests of the pairwise differences. ADJUST=BON, ADJUST=TUKEY, and ADJUST=SIMULATE request the Bonferroni, Tukey, and Simulate multiple comparison adjustments, respectively. The CVADJUST option performs a more accurate simulation than without that option; however, it also takes more time to run the simulation. The REPORT option gives a report of the simulation results.

The following SAS® output results:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	2	102	0.58	0.5592

Figure 2.3.3. Results of type 3 test of fixed effects.

Least Squares Means

Effect	trt	Estimate	Standard Error	DF	t Value	Pr > |t|
trt	1	2.8286	0.3586	102	7.89	<.0001
trt	2	3.3714	0.3586	102	9.40	<.0001
trt	3	3.0857	0.3586	102	8.61	<.0001

Figure 2.3.4. Least squares means.

Differences of Least Squares Means

Effect	trt	_trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
trt	1	2	-0.5429	0.5071	102	-1.07	0.2869	Bonferroni	0.8607
trt	1	3	-0.2571	0.5071	102	-0.51	0.6132	Bonferroni	1.0000
trt	2	3	0.2857	0.5071	102	0.56	0.5744	Bonferroni	1.0000
trt	1	2	-0.5429	0.5071	102	-1.07	0.2869	Tukey	0.5345
trt	1	3	-0.2571	0.5071	102	-0.51	0.6132	Tukey	0.8681
trt	2	3	0.2857	0.5071	102	0.56	0.5744	Tukey	0.8398
trt	1	2	-0.5429	0.5071	102	-1.07	0.2869	Simulate	0.5345
trt	1	3	-0.2571	0.5071	102	-0.51	0.6132	Simulate	0.8681
trt	2	3	0.2857	0.5071	102	0.56	0.5744	Simulate	0.8398

Figure 2.3.5. Differences of least squares means.

Simulation Results

Method	95% Quantile	Exact Alpha
Simulated	2.377872	0.0501
Tukey	2.378431	0.0500
Bonferroni	2.434104	0.0436
Sidak	2.427577	0.0443
GT-2	2.426150	0.0445
Scheffe	2.484136	0.0385
T	1.983495	0.1215

Figure 2.3.6. Simulation results.

Bonferroni

In the third section of the output (2.3.5) entitled "Differences of Least Squares Means," the adjusted p-values next to the Bonferroni adjustments are compared with the nominal level of alpha. If smaller than α_0, the comparison is deemed significant. So, using the Bonferroni

adjustment, the comparison between treatment 1 and treatment 2 is not significant with a p-value of 0.8607 at a α_0 =0.05 level.

Fisher's Protected LSD

In the first section of the output (2.3.3) entitled "Type-3 Tests of Fixed Effects," the F-test is considered as the first step in the Fisher's Protected LSD procedure. The p-value is 0.5592 that is compared with α_0. Because it exceeds 0.05, the test is not significant at a α_0 =0.05 level. Had it been significant, the researcher would then use the unadjusted p-values in the third section of the output (2.3.5) to conduct the individual comparisons. The p-value used for comparing treatment 1 and treatment 2 would be 0.2869.

Tukey

In the third section of the output (2.3.5) entitled "Differences of Least Squares Means," the adjusted p-values next to the Tukey adjustments are compared with the nominal level of alpha. If smaller than α_0, the comparison is deemed significant. So, using the Tukey adjustment, the comparison between treatment 1 and treatment 2 is not significant with a p-value of 0.5345 at a α_0 =0.05 level.

Simulate

In the third section of the output (2.3.5) entitled "Differences of Least Squares Means," the adjusted p-values next to the Simulate adjustments are compared with the nominal level of alpha. If smaller than α_0, the comparison is deemed significant. So, using the Simulate adjustment, the comparison between treatment 1 and treatment 2 is not significant with a p-value of 0.5345 at a α_0 =0.05 level.

The fourth section of the output (2.3.6) entitled "Simulation Results" allows the researcher to see the actual level of alpha obtained by the test. It can be seen that alpha is clearly inflated when no multiple comparison adjustment is made (Method=T). The simulate method and Tukey's method yield identical p-values. This is not surprising because they both involve quantiles of a multivariate-t distribution. However, in this example there were equal sample

sizes and the design was balanced. When such is not the case, Simulate will be more accurate than Tukey's method.

Evaluation (Fixed ρ)

The value ρ=0.1 is chosen. This value is then used to generate a data set with equal population means and t=4, 6, or 8 treatments and n=10 or 100 students per class. The MIXED procedure is then used to analyze the data using the Simulate multiple comparison procedure to test the significance of pairwise differences. The values ρ_0=0.05, 0.10, 0.15, and 0.20 are used in the analysis. There are 1,000 iterations of the experiment at each level of ρ_0, t, and n. The following graphs display the results of the simulation.

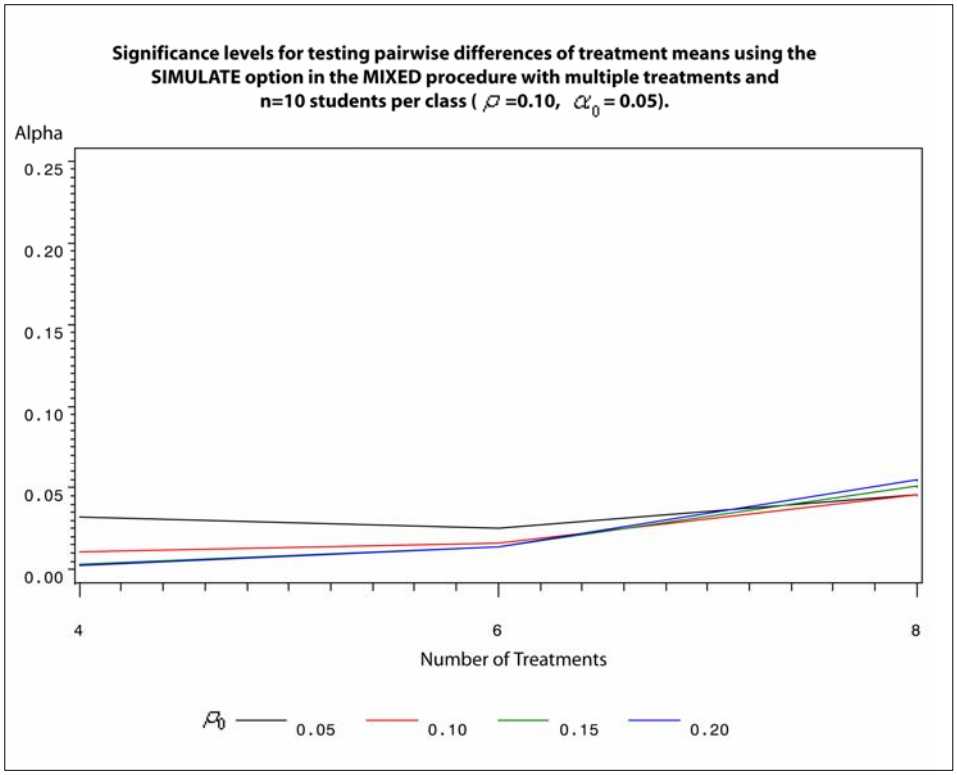


Figure 2.3.7. Significance levels using the SIMULATE option in the MIXED procedure with 10 students per class; ρ=0.10.

As the plots show in general, the greater the value of ρ_0 that is used in the analysis, the lower the significance level. However, there is some overlap. This is likely due to the relatively small

number of simulations (1000) used to create each point on the graph. For the values ρ_0 =0.10, 0.15, and 0.20, the significance level remains near or below the nominal level (0.05), increasing slightly for more treatments.

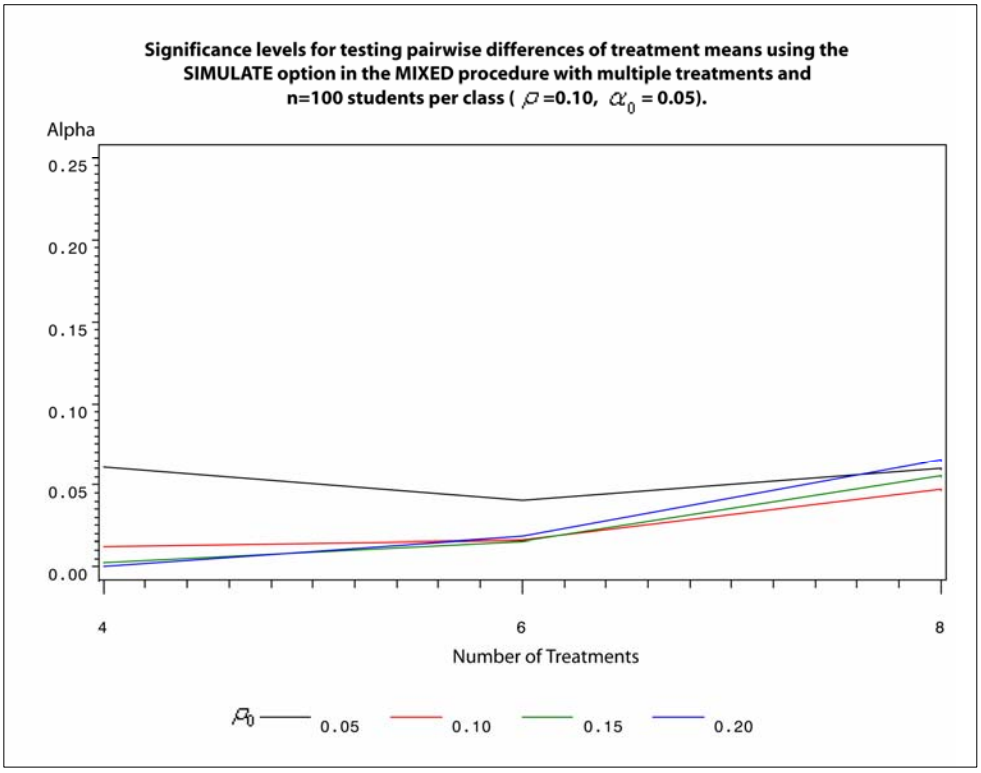


Figure 2.3.8 Significance levels using the SIMULATE option in the MIXED procedure with 100 students per class; ρ =0.10.

Evaluation (Random ρ)

The value ρ is chosen at random from a Beta(5,30) distribution. This value is then used to generate a data set with equal population means and t=4, 6, or 8 treatments and n=10 or 100 students per class. The MIXED procedure is then used to analyze the data using the Simulate multiple comparison procedure to test the significance of pairwise differences. The values ρ_0 =0.05, 0.10, 0.15, and 0.20 are used in the analysis. There are 1,000 iterations of the experiment at each level of ρ_0, t, and n. The following figures display the results of the simulation.

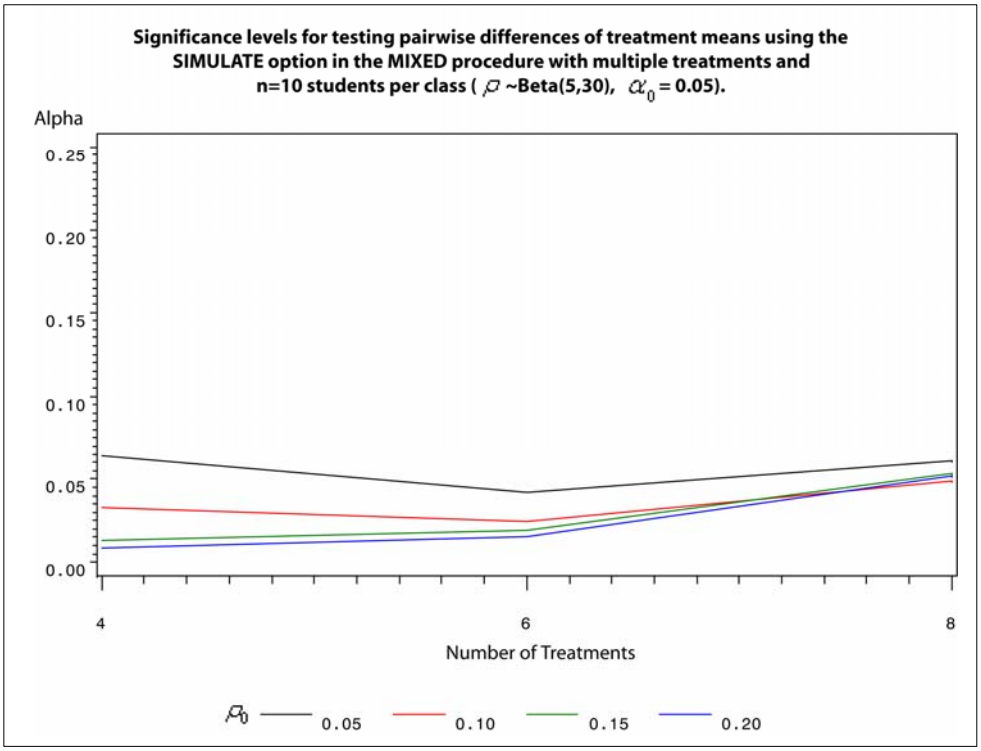


Figure 2.3.9. Significance levels using the SIMULATE option in the MIXED procedure with 10 students per class; ρ ~Beta(5,30).

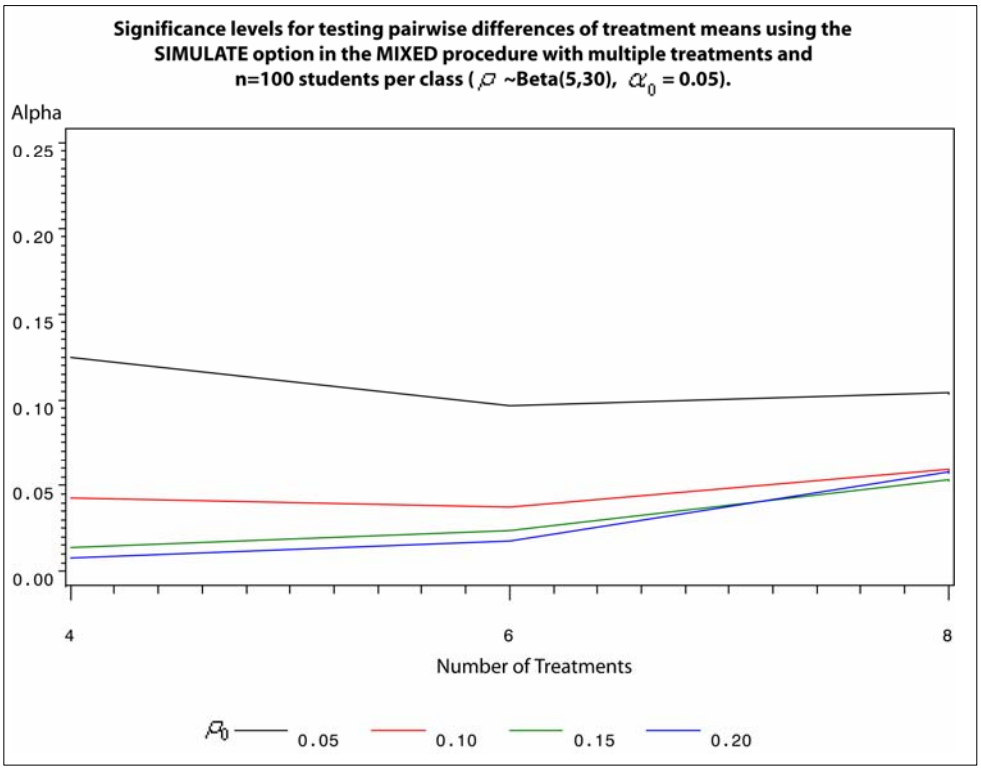


Figure 2.3.10. Significance levels using the SIMULATE option in the MIXED procedure with 100 students per class; ρ ~Beta(5,30).

As the figures show, the greater the value of ρ_0 that is used in the analysis, the lesser the significance level. The average of the Beta(5,30) distribution is 0.14286. For the values ρ_0 =0.10, 0.15, and 0.20, the significance level remains near or below the nominal level (0.05). Only when ρ_0 =0.05 is the significance level inflated.

CHAPTER 3: USING A KNOWN INTRACLASS CORRELATION COEFFICIENT TO TEST THE DIFFERENCES AMONG TREATMENT MEANS (UNDER-REPLICATION)

3.1: THE TWO-TREATMENT CASE

THE EFFECT OF REPLICATION ON THE ANALYSIS

This section looks at methods for dealing with under-replication. It is important to determine when (how often) it is better to use standard mixed model methods than to use a chosen value for ρ. That is the object of inquiry for this section.

Description

Recall the model (1.1.1) from section 1:

$$y_{ijk} = \mu_i + \delta_{ij} + \varepsilon_{ijk}$$

Likewise, recall (1.1.3)

$$\text{var}\left[\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}\right] = \sigma_\delta^2\left[\left(\frac{1}{n_{i\cdot}}\right)^2\left\langle n_{i\cdot}\right\rangle^2 + \left(\frac{1}{n_{i'\cdot}}\right)^2\left\langle n_{i\cdot}\right\rangle^2\right] + \sigma_\varepsilon^2\left[\frac{1}{n_{i\cdot}} + \frac{1}{n_{i'\cdot}}\right]$$

$$= \sigma_\varepsilon^2\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\left\langle n_{i\cdot}\right\rangle^2 + \left(\frac{1}{n_{i'\cdot}}\right)^2\left\langle n_{i\cdot}\right\rangle^2\right] + \left[\frac{1}{n_{i\cdot}} + \frac{1}{n_{i'\cdot}}\right]\right\},$$

since $\sigma_\delta^2 = \sigma_\varepsilon^2\frac{\rho}{1-\rho}$.

For the two-treatment case, this reduces to

$$\operatorname{var}\left(\bar{y}_{1\cdot\cdot}-\bar{y}_{2\cdot\cdot}\right)=\sigma_{\varepsilon}^{2}\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{1\cdot}}\right)^{2}\left\langle n_{1\cdot}\right\rangle^{2}+\left(\frac{1}{n_{2\cdot}}\right)^{2}\left\langle n_{2\cdot}\right\rangle^{2}\right]+\left[\frac{1}{n_{1\cdot}}+\frac{1}{n_{2\cdot}}\right]\right\} \qquad (3.1.1)$$

Next, we look at the effect replication (the number of classes per treatment) has on the significance level.

Let an estimate of σ_{ε}^{2} be defined as

$$\hat{\sigma}_{\varepsilon}^{2}=\frac{\sum_{i=1}^{2}\sum_{j=1}^{b}\left(n_{ij}-1\right)s_{ij}^{2}}{\sum_{i=1}^{2}\sum_{j=1}^{b}\left(n_{ij}-1\right)} \qquad (3.1.2)$$

where

$$s_{ij}^{2}=\frac{\sum_{k=1}^{n_{ij}}\left(y_{ijk}-\bar{y}_{ij\cdot}\right)^{2}}{n_{ij}-1}. \qquad (3.1.3)$$

The estimator in Equation 3.1.2 is not equivalent to the one in Equation 2.3.4. In the case of replication, Equation 2.3.4 depends on the unknown value ρ. Equation 3.1.2 will be used as an estimate for σ_{ε}^{2}. It is still an unbiased estimator, but does not depend on ρ. Another difference is that tests involving Equation 3.1.2 will have $tb(n-1)$ error degrees of freedom instead of $t(bn-1)$ error degrees of freedom from using Equation 2.3.4. In the two-treatment case, $t=2$.

If we consider the case when $n=n_{ij}$, $b=b_{i}$, the probability of rejection can be written as

$$P\left[\frac{(\bar{y}_1-\bar{y}_2)-(\mu_{01}-\mu_{02})}{\sqrt{\frac{2\hat{\sigma}_\varepsilon^2}{b}\left(\frac{\rho_0}{1-\rho_0}+\frac{1}{n}\right)}}>t_{0.05,2b(n-1)}\right]=P\left[T>\left(t_{0.05,2b(n-1)}\right)\frac{\sqrt{\frac{\rho_0}{1-\rho_0}+\frac{1}{n}}}{\sqrt{\frac{\rho}{1-\rho}+\frac{1}{n}}}\right] \qquad (3.1.4)$$

where T is a random variable with a non-central t-distribution with non-centrality parameter $\lambda=\frac{(\mu_1-\mu_2)-(\mu_{01}-\mu_{02})}{\sqrt{\frac{2\sigma_\varepsilon^2}{b}\left(\frac{\rho}{1-\rho}+\frac{1}{n}\right)}}$ which depends on the values of ρ, ρ_0, b, and n. As $n\to\infty$ and under the null hypothesis, (3.1.4) becomes

$$P\left[Z>\left(z_{0.05}\right)\frac{\sqrt{\frac{\rho_0}{1-\rho_0}}}{\sqrt{\frac{\rho}{1-\rho}}}\right] \qquad (3.1.5)$$

which only depends on ρ and ρ_0. As the error degrees of freedom increases, either by increasing n or b, the effect of b (the number of classes per treatment) on the probability of rejection using the plug-in method decreases. The following figures will illustrate this.

The following figures display significance levels for testing the equality of two treatment means in a replication setting. Ten and a theoretically infinite number of students per class are considered. The true value ρ is set at 0.1, a likely value for many applications, and the values of ρ_0 range from 0.025 to 0.20 to demonstrate the consequence of discrepancy between the researchers choice, ρ_0, and the actual value, ρ. Figure 3.1.1 is generated using Equation 3.1.4 and Figure 3.1.2 is generated using Equation 3.1.5.

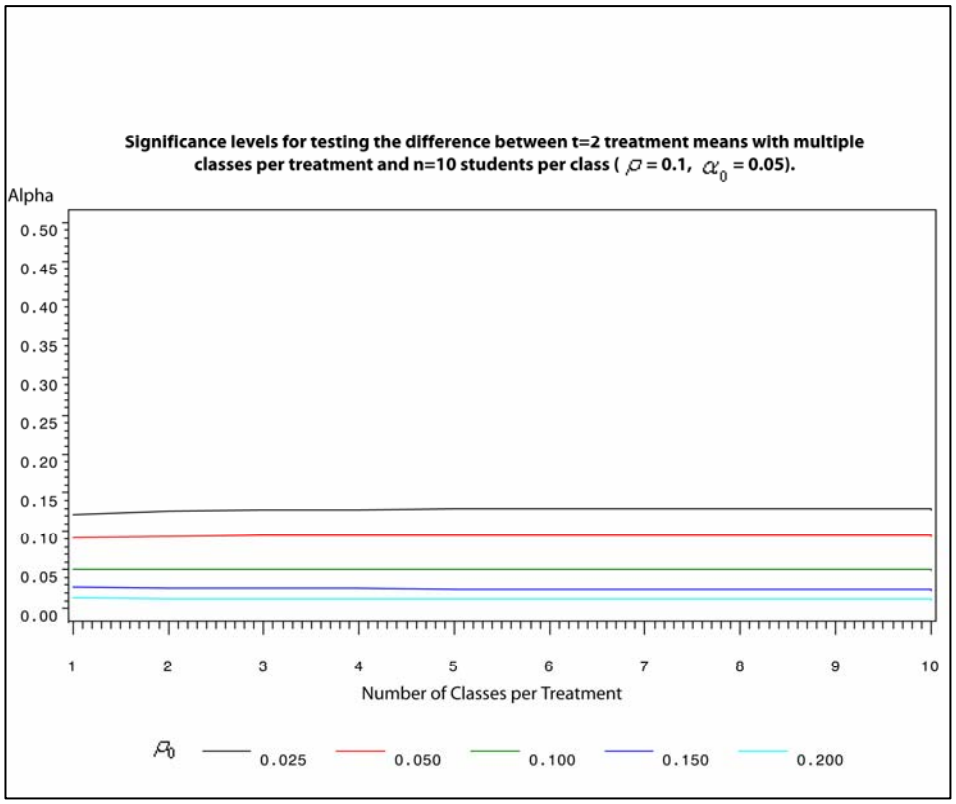


Figure 3.1.1. Small sample significance levels for different numbers of classes per treatment.

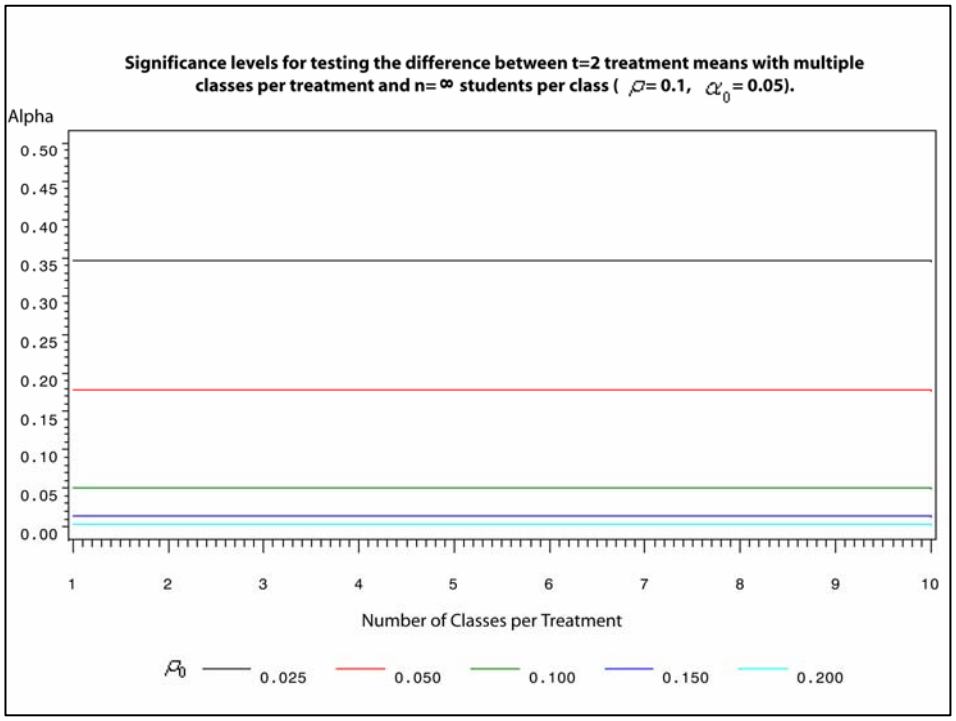


Figure 3.1.2. Large sample significance levels for different numbers of classes per treatment.

In the first graph (3.1.1), n=10 students per class. There is virtually no change in significance level as the number of classes per treatment changes from one to ten. This holds true for larger numbers of classes per treatment as well.

In the second graph (3.1.2), we let $n \to \infty$ resulting in $\lim_{n\to\infty} \frac{1}{n} = 0$ being used in Equation 3.1.4 to obtain Equation 3.1.5. There is no change in significance level as the number of classes per treatment changes from one to ten. Thus, the strategies for choosing ρ_0 developed in the no-replication case have essentially the same effect on type 1 error whether there is replication or not.

The associated graphs of the power curves in the n=10 and theoretically infinite students per class appear next (3.1.3 and 3.1.4). Appropriately, the power curves increase as the number of classes per treatment increases.

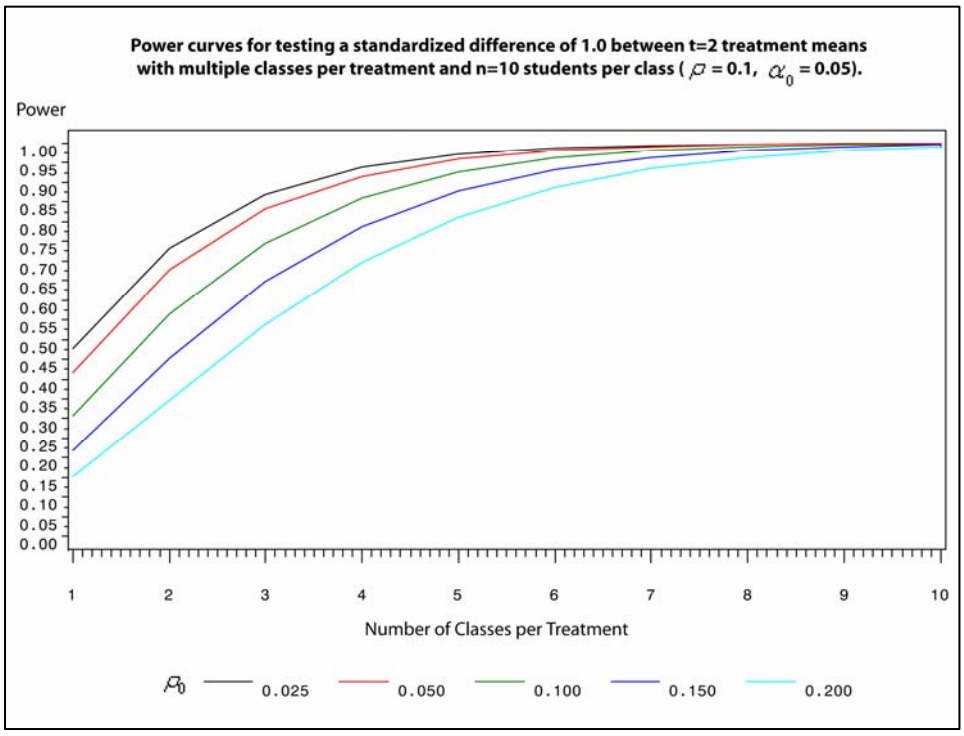


Figure 3.1.3. Small sample power curves for different numbers of classes per treatment.

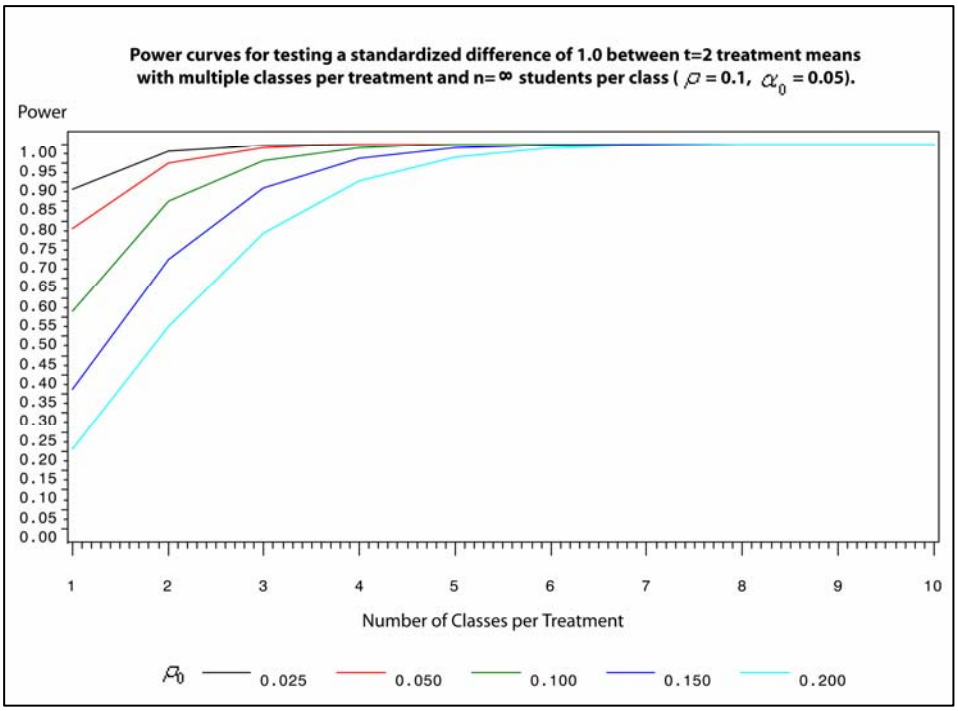


Figure 3.1.4. Large sample power curves for different numbers of classes per treatment.

Next, we compare two tests, again in the case when $n = n_{ij}$, $b = b_i$. The first is a traditional t-test of the difference between treatment means using class means as the response variable for the design. The second test is a t-test of the difference between treatment means using the plug-in method.

Traditional T-test on Class Means

Let $w_{ij} = \bar{y}_{ij\cdot}$, the mean of the students in class *j* receiving treatment *i*, μ_i is the fixed effect of treatment i, $\varepsilon_{ij}^* = \delta_{ij} + \bar{\varepsilon}_{ij\cdot}$ is the random effect of class j average given treatment i; i = 1, 2, j = 1, 2, …, b, k = 1, 2, …, n. We have $\varepsilon_{ij}^* \sim n\left(0, \sigma_{\varepsilon^*}^2\right)$ where $\sigma_{\varepsilon^*}^2 = \sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{n}$. We may write a model for the experiment as follows:

$$w_{ij} = \mu_i + \varepsilon_{ij}^* \qquad (3.1.6)$$

Let $s_i^2 = \dfrac{\sum_{j=1}^{b}\left(w_{ij} - \overline{w}_{i\cdot}\right)^2}{b-1}$ and let $s_p^2 = \dfrac{s_1^2 + s_2^2}{2}$. Then $\hat{\text{var}}\left(\overline{w}_{1\cdot} - \overline{w}_{2\cdot}\right) = \dfrac{2s_p^2}{b}$ with degrees of freedom $\upsilon = 2(b-1)$.

We test H_0: $\mu_1 = \mu_2$ vs. H_A: $\mu_1 \neq \mu_2$. Let $t = \dfrac{\overline{w}_{1\cdot} - \overline{w}_{2\cdot}}{\sqrt{\left[\hat{\text{var}}\left(\overline{w}_{1\cdot} - \overline{w}_{2\cdot}\right)\right]}}$ be our test statistic. The null hypothesis is rejected in favor of the alternative if $|t| > t_{\alpha_0/2,\upsilon}$, where $t_{\alpha_0/2,\upsilon}$ is the upper $\alpha_0/2$ quantile of the Student's t distribution with υ degrees of freedom.

Properties

The variance of the difference is

$$\text{var}\left[\overline{w}_{1\cdot} - \overline{w}_{2\cdot}\right] = \frac{2}{b}\left(\sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{n}\right) \qquad (3.1.7)$$

Let $\upsilon = 2(b-1)$, $\lambda = \dfrac{|\mu_1 - \mu_2|}{\sqrt{\dfrac{2}{b}\left(\sigma_\delta^2 + \dfrac{\sigma_\varepsilon^2}{n}\right)}}$, and let T be a random variable having a noncentral t-distribution with degrees of freedom υ and noncentrality parameter λ. The power of the traditional t-test can be found using the following formula:

$$Power = P\left[T < \left(-t_{0.975,\upsilon,-\lambda}\right)\right] + P\left[T > \left(t_{0.025,\upsilon,\lambda}\right)\right] \qquad (3.1.8)$$

T-test Using Plug-in Method

Using model (1.1.1) we can define our test statistic as

$$t^* = \frac{\left(\overline{y}_{1\cdot} - \overline{y}_{2\cdot}\right)}{\sqrt{\dfrac{2\hat{\sigma}_\varepsilon^2}{b}\left(\dfrac{\rho_0}{1-\rho_0} + \dfrac{1}{n}\right)}} \qquad (3.1.9)$$

where $\hat{\sigma}_\varepsilon^2$ represents a pooled estimate of the within class variance computed using Equation 3.1.2. The null hypothesis is rejected in favor of the alternative if $|t^*| > t_{\alpha_0/2,\upsilon^*}$, where $t_{\alpha_0/2,\upsilon^*}$ is the upper $\alpha_0/2$ quantile of the Student's t distribution with $\upsilon^* = 2b(n-1)$ degrees of freedom.

Properties

The variance of the difference is

$$\text{var}[\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}] = \frac{2\sigma_\varepsilon^2}{b}\left(\frac{\rho}{1-\rho} + \frac{1}{n}\right) \qquad (3.1.10)$$

Let $\lambda = \dfrac{|\mu_1 - \mu_2|}{\sqrt{\dfrac{2\sigma_\varepsilon^2}{b}\left(\dfrac{\rho}{1-\rho} + \dfrac{1}{n}\right)}} = \dfrac{|\mu_1 - \mu_2|}{\sqrt{\dfrac{2}{b}\left(\sigma_\delta^2 + \dfrac{\sigma_\varepsilon^2}{n}\right)}}$, the same as for the traditional t-test, and let T* be a random variable having a noncentral t-distribution with degrees of freedom υ^* and noncentrality parameter λ. The power of the test can be found using the following formula:

$$Power = P\left[T^* < \left(-t_{0.975,\upsilon^*,-\lambda}\right)\frac{\sqrt{\dfrac{\rho_0}{1-\rho_0} + \dfrac{1}{n}}}{\sqrt{\dfrac{\rho}{1-\rho} + \dfrac{1}{n}}}\right] + P\left[T^* > \left(t_{0.025,\upsilon^*,\lambda}\right)\frac{\sqrt{\dfrac{\rho_0}{1-\rho_0} + \dfrac{1}{n}}}{\sqrt{\dfrac{\rho}{1-\rho} + \dfrac{1}{n}}}\right] \qquad (3.1.11)$$

In the optimal case, $\rho_0 = \rho$, the power of the test (*3.1.11*) differs from that in (*3.1.8*) only by the degrees of freedom: υ for the traditional t-test on the class means and $\upsilon^* > \upsilon$ for the t-test using the plug-in method. As the sample size b increases, eventually a point will be reached in which the power based on analyzing the test using the plug-in method is not appreciably greater than the power based on analyzing the test using the traditional t-test.

The following figures demonstrate the difference between the power using equation 3.1.11 versus using equation 3.1.8 in the optimal case of $\rho_0 = \rho$.

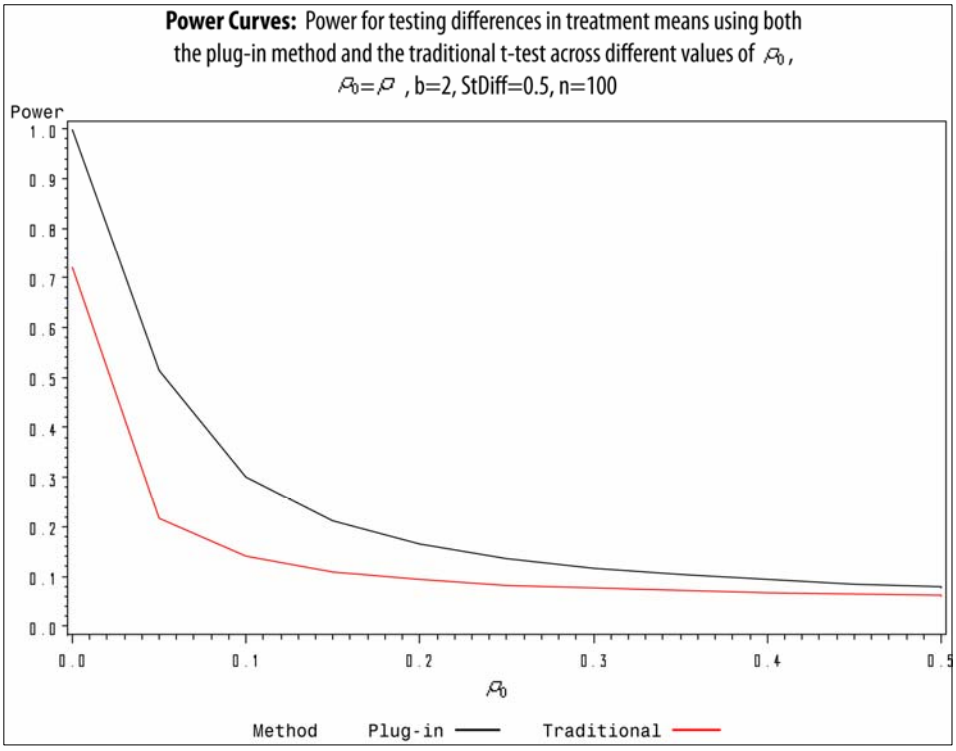


Figure 3.1.5. Power curves; plug-in vs. traditional method; 100 students per class; 2 classes per treatment.

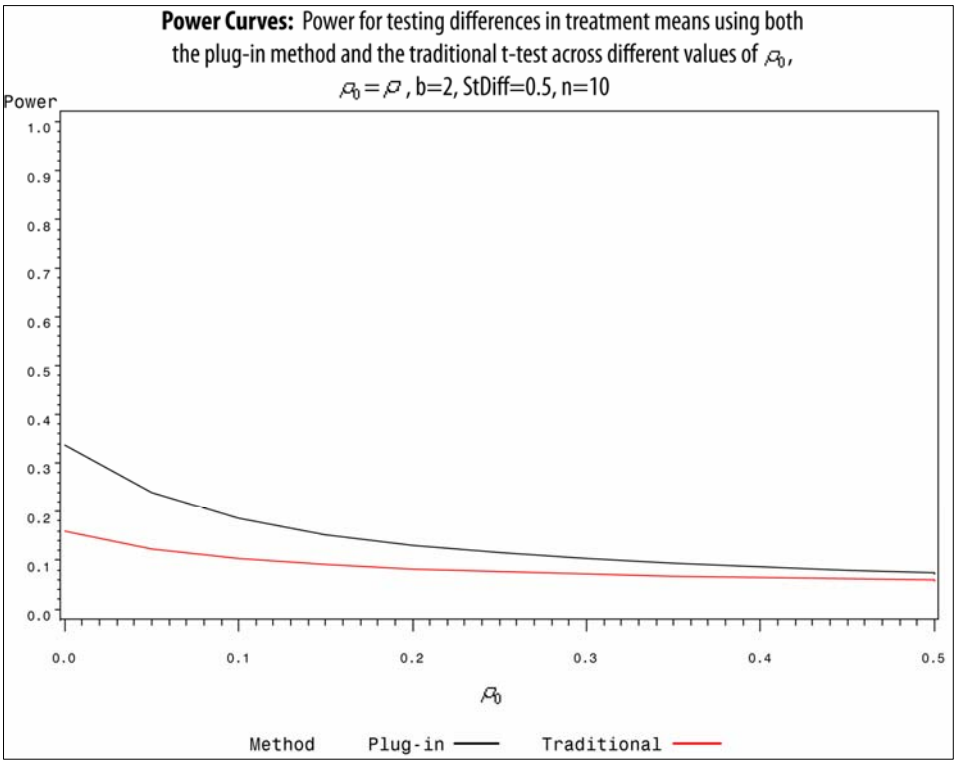


Figure 3.1.6. Power curves; plug-in vs. traditional method; 10 students per class; 2 classes per treatment.

Figure 3.1.5 represents the different power curves for b=2, two experimental units per treatment, 100 subsamples, and a standardized difference of 0.5. The black line represents the t-test using the plug-in method, and the red line represents a traditional t-test. If ρ is around 0.1 or less, there is a significant loss in power if the information about ρ is not used. In fact, at ρ =0.05, the difference in power is 0.295. For values of ρ >0.3 no appreciable difference exists.

Figure 3.1.6 uses 10 subsamples instead of 100. The decrease in the number of subsamples leads to a decrease in power in both cases and narrows the difference between the known ρ model and the traditional t-test. At ρ =0.05, the difference in power is 0.116. For values of ρ >0.1, no appreciable difference exists.

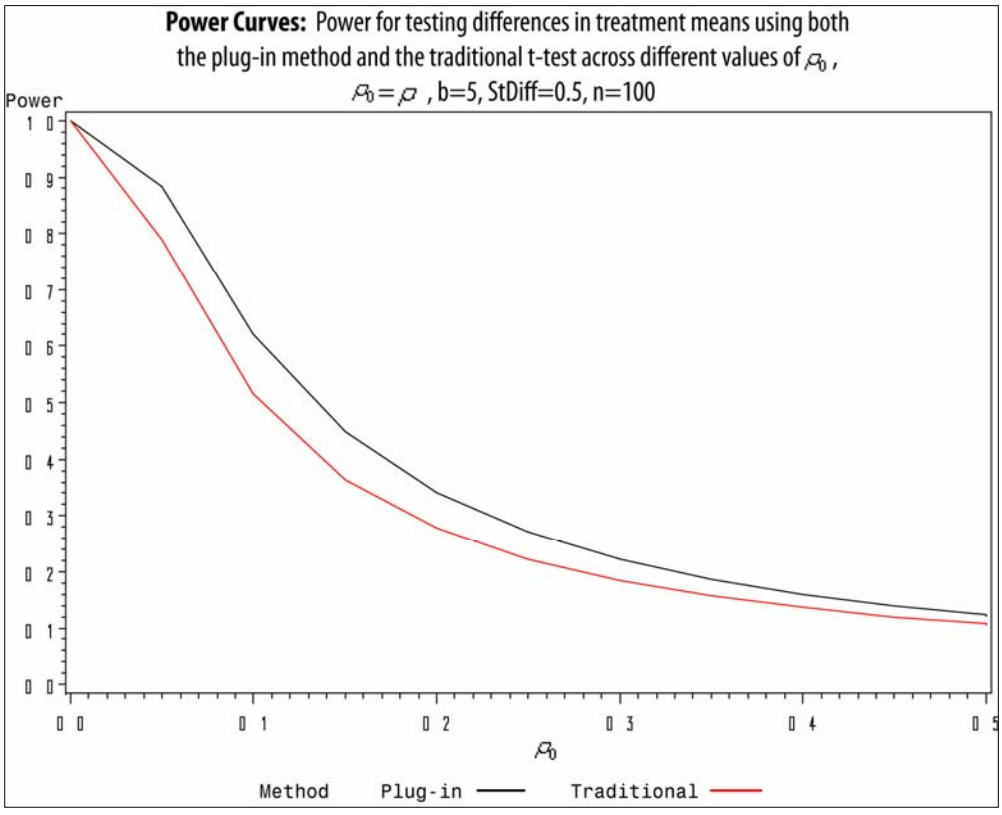


Figure 3.1.7. Power curves; plug-in vs. traditional method; 100 students per class; 5 classes per treatment.

Figure 3.1.7 uses 100 subsamples and now uses b=5, five experimental units per treatment. Although a difference in the power exists, it is at most about 0.1, so one might override the complexity of the plug-in method for the simplicity of the traditional t-test.

Another interesting comparison is that of the plug-in versus the traditional t-test if the choice of ρ_0 is off by ± 0.05. Figure 3.1.8 shows that with the choice of ρ_0 differing from ρ by $+0.05$, the plug-in method may even perform worse (less power) than a traditional t-test—with as few as two experimental units per treatment. So, the choice of ρ_0 is critical if power is to be improved using the plug-in method over the traditional method.

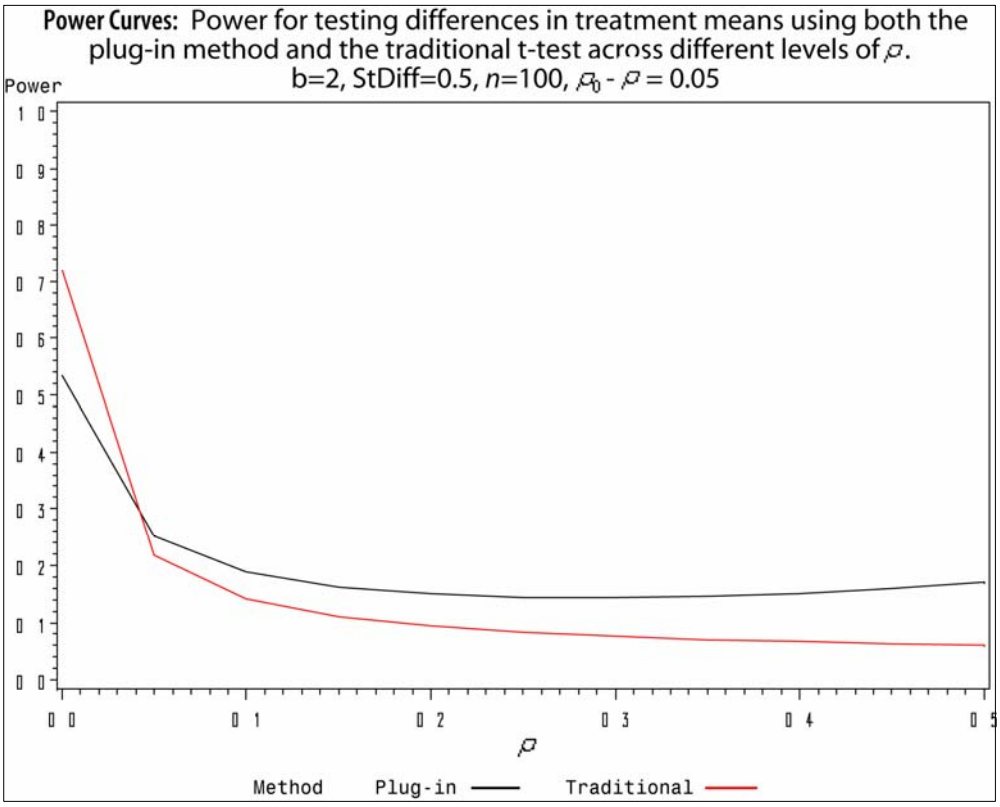


Figure 3.1.8. Power curves; plug-in vs. traditional method; 100 students per class; 2 classes per treatment; $\rho_0 - \rho = 0.50$.

It appears then that using the plug-in method in the two-treatment case, with ρ_0 close to ρ, is preferable with up to five experimental units per treatment, but beyond that would not yield significant improvements over the traditional t-test.

3.2: STRATEGIES FOR CHOOSING ρ_0

The strategies recommended for choosing ρ_0 in an unreplicated experiment extend to the case of the under-replicated experiment. The main difference between the unreplicated case and the under-replicated case is the computation of the error degrees of freedom. Whereas in the unreplicated case, the full error degrees of freedom used in testing hypotheses were *2(n-1)*, in the under-replicated case, the full error degrees of freedom are *2b(n-1)*, where b represents the number of classes per treatment.

As shown in Figures 3.1.1, the effect of the number of classes per treatment (b) on the significance level is minimal. No changes need be made to the Acceptable Interval procedure nor to the Weighted P-Value method for them to work in the under-replicated experiment case, other than the change in error degrees of freedom naturally associated with replication. The ADFMR strategy (Strategy 3) will change with respect to the Satterthwaite-type adjustment to the degrees of freedom. Recall, the Satterthwaite-type adjustment (*2.2.1*) is of the form $d.f. = \dfrac{2[Var(\bar{y}_1 - \bar{y}_2)]^2}{Var[\hat{Var}(\bar{y}_1 - \bar{y}_2)]}$.

In the two-treatment under-replication case, this becomes

$$d.f. = \frac{2b(n-1)[E(W)]^2}{(bn-b+1)Var(W) + [E(W)]^2} \qquad (3.2.1)$$

where

$$W = \frac{\rho}{1-\rho} + \frac{1}{n}.$$

As $Var(W) \to 0$, the error degrees of freedom using the Satterthwaite-type adjustment approach $2b(n-1)$, the full d.f, indicating that the more precise the choice of ρ_0, the more appropriate the $2b(n-1)$ error degrees of freedom. Also, as $n \to \infty$, the error degrees of freedom using the Satterthwaite-type adjustment approach

$$d.f.(Satterthwaite-type) = 2\left\{\frac{\left[E\left(\frac{\rho}{1-\rho}\right)\right]^2}{Var\left(\frac{\rho}{1-\rho}\right)}\right\}.$$

Random ρ

We now look at average significance level and power for the various strategies by randomly selecting ρ from a population of ρ's and determine the proportion of times our test correctly rejects or fails to reject the null hypothesis.

A simulation was conducted with the following specifications:

Iterations = 100,000
n = 10, 100
b = 2, 4, 6
Error d.f. = Satterthwaite-type and *2b(n-1)*
$\rho_0 = \bar{\rho}$ and 75th quantile
ρ ~Beta(1,10) and Beta(5,30)

The following distributions show plausible prior information about ρ and the consequences on the average significance and power levels.

Example 1: ρ ~Beta(1,10)

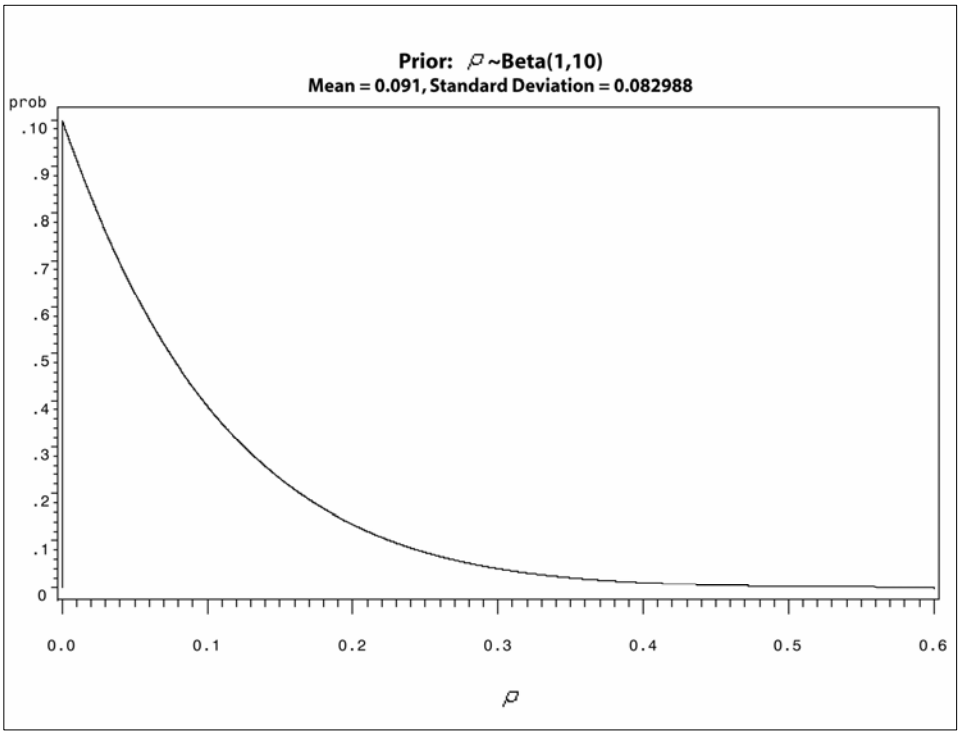


Figure 3.2.1. Prior distribution for ρ, Beta(1,10).

Table 3.2.1. Simulated significance levels and power levels using ρ ~Beta(1,10) prior distribution; n = 10, 100; b=2, 4, 6; error d.f. = Satterthwaite-type and 2(n-1); $\rho_0 = \overline{\rho}$ and 75th quantile.

	b=2		b=4		b=6	
$\rho_0 = \overline{\rho}$	n=10	n=100	n=10	n=100	n=10	n=100
Satterthwaite d.f. =	4.750	1.882	5.213	1.892	5.388	1.895
Simulated Alpha =	0.019	0.002	0.020	0.002	0.021	0.002
Simulated Power(diff=1) =	0.335	0.045	0.715	0.265	0.900	0.717
2b(n-1) d.f. =	36	396	72	792	108	1188
Simulated Alpha =	0.052	0.060	0.053	0.061	0.053	0.060
Simulated Power(diff=1) =	0.568	0.860	0.872	0.975	0.962	0.993
$\rho_0 = 75^{th}$ *Quantile*	n=10	n=100	n=10	n=100	n=10	n=100
2b(n-1) d.f. =	36	396	72	792	108	1188
Simulated Alpha =	0.039	0.040	0.039	0.041	0.040	0.039
Simulated Power(diff=1) =	0.498	0.786	0.832	0.963	0.948	0.990

Example 2: ρ ~Beta(5,30)

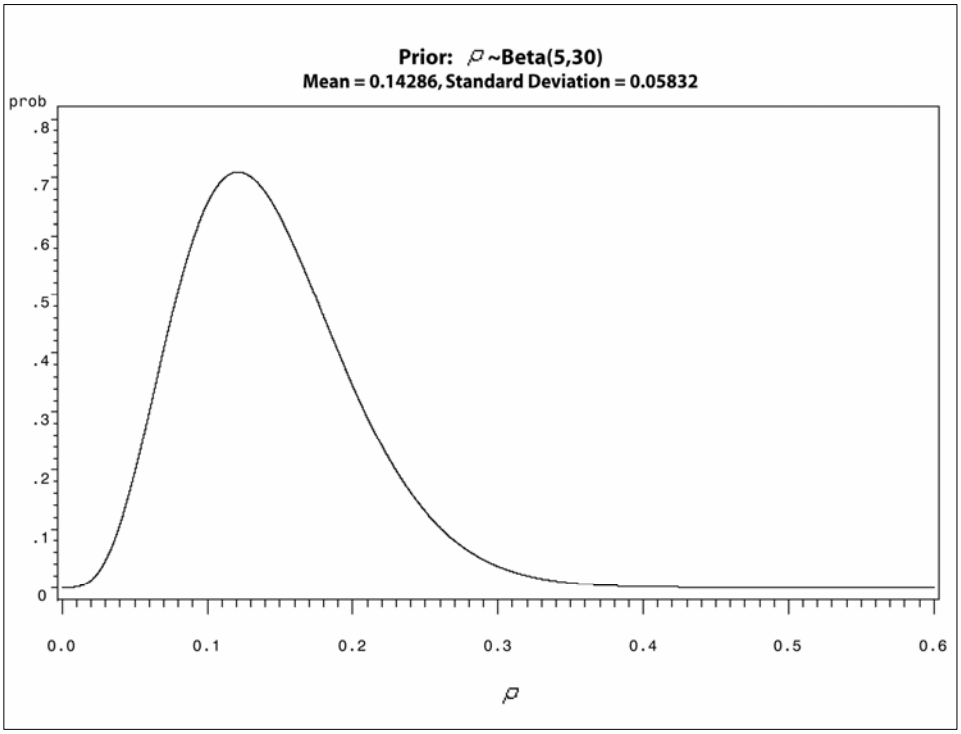


Figure 3.2.2. Prior distribution for ρ, Beta(5,30).

Table 3.2.2. Simulated significance levels and power levels using ρ ~Beta(5,30) prior distribution; n = 10, 100; b=2, 4, 6; error d.f. = Satterthwaite-type and 2(n-1); $\rho_0 = \overline{\rho}$ and 75th quantile.

	b=2		b=4		b=6	
$\rho_0 = \overline{\rho}$	n=10	n=100	n=10	n=100	n=10	n=100
Satterthwaite d.f. =	12.639	8.964	15.654	9.090	17.006	9.132
Simulated Alpha =	0.039	0.030	0.040	0.030	0.039	0.031
Simulated Power(diff=1) =	0.408	0.533	0.725	0.864	0.888	0.964
2b(n-1) d.f. =	36	396	72	792	108	1188
Simulated Alpha =	0.052	0.055	0.052	0.053	0.050	0.055
Simulated Power(diff=1) =	0.462	0.657	0.767	0.916	0.910	0.979
$\rho_0 = 75^{th}$ *Quantile*	n=10	n=100	n=10	n=100	n=10	n=100
2b(n-1) d.f. =	36	396	72	792	108	1188
Simulated Alpha =	0.038	0.034	0.037	0.035	0.037	0.035
Simulated Power(diff=1) =	0.400	0.564	0.717	0.878	0.883	0.968

Looking at Tables 3.2.1 and 3.2.2, it is apparent that replication increases the power of the test. Tests using the Satterthwaite-type adjusted error degrees of freedom seem to be much more appropriate than in the case of no replication. Whereas it was recommended not to use the Satterthwaite-type adjusted error degrees of freedom in the case of no replication, except when needing to assure adherence to α_0, in the case of under-replication, the Satterthwaite-type method or using $\rho_0 = 75^{th}$ quantile both seem to be recommendable options.

3.3: THE T-TREATMENT CASE

It is reasonable to assume the researcher will want to work with more than two treatments. Refer to the model (1.1.1) in section 1.1 for notation. Let $N = \sum_{i=1}^{t}\sum_{j=1}^{b_i} n_{ij}$; Y be an $N \times 1$ observable vector of random variables; X be a $N \times t$ design matrix, that is, a matrix of 0's and 1's denoting the design's treatment structure; β be a $t \times 1$ vector consisting of the treatment means $\beta_1, \beta_2, ..., \beta_t$; E be a $N \times 1$ vector of unknown random variables. Then we can write the Generalized Linear Model can be written as

$$Y = X\beta + E \qquad (3.3.1)$$

Let H be a $q \times t$ matrix and h be a $q \times 1$ vector, as defined in Graybill (1976, p. 184). Let V be an $N \times N$ known positive definite matrix such that $Var(Y) = \sigma_\varepsilon^2 V$.

$$V = \begin{bmatrix} A_{11} & 0 & \dots & \dots & 0 \\ 0 & A_{12} & \dots & \dots & 0 \\ . & & . & & . \\ . & & & . & . \\ 0 & 0 & \dots & & A_{tb_t} \end{bmatrix}, \; A_{ij} = \frac{1}{1-\rho}\begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} \qquad (3.3.2)$$

A_{ij} is an $n_{ij} \times n_{ij}$ matrix. The best linear unbiased estimate of β is $\hat{\beta} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}Y$. This is $\hat{\beta} = \left[\bar{y}_{1\cdot\cdot}, \bar{y}_{2\cdot\cdot}, ..., \bar{y}_{t\cdot\cdot}\right]$ in the case where $n = n_{ij}$ and $b = b_i$. An estimate of σ_ε^2 is

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{b_i}\left(n_{ij}-1\right)s_{ij}^2}{\sum_{i=1}^{t}\sum_{j=1}^{b_i}\left(n_{ij}-1\right)} \qquad (3.3.3)$$

where

$$s_{ij}^2 = \frac{\sum_{k=1}^{n_{ij}}\left(y_{ijk}-\bar{y}_{ij\cdot}\right)^2}{n_{ij}-1}. \qquad (3.3.4)$$

Equation 3.3.3 is not equivalent to Equation 2.3.4, except in the case of no replication. As explained previously, Equation 2.3.4 depends on the unknown value ρ. So, Equation 3.3.3 will be used instead. The test statistic for testing H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$ is of the form

$$W = \frac{\left(H\hat{\beta}-h\right)'\left[H\left(X'V^{-1}X\right)^{-1}H'\right]^{-1}\left(H\hat{\beta}-h\right)}{q\hat{\sigma}_\varepsilon^2} \qquad (3.3.5)$$

Reject H_0 at level α_0 when $W > f_{\alpha_0,q,\upsilon}$, where $f_{\alpha_0,q,\upsilon}$ is the $1-\alpha_0$ quantile of the F distribution with q and $\upsilon = \sum_{i=1}^{t}\sum_{j=1}^{b_i}\left(n_{ij}-1\right)$ degrees of freedom.

Properties

For simplicity, we consider the case where $n = n_{ij}$ and $b = b_i$ for all values of i and j. We denote V_0 as the matrix V with $\rho = \rho_0$. It can be shown that

$$\left(X'V_0^{-1}X\right)^{-1} = \frac{1}{b}\left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n}\right)I_t = \left[\frac{1+(n-1)\rho_0}{(1-\rho_0)}\right] \times \left(\frac{1}{bn}\right)I_t$$

$$= d_0\left(X'X\right)^{-1}$$

where

$$d_0 = \frac{1+(n-1)\rho_0}{(1-\rho_0)}. \qquad (3.3.6)$$

Let $\upsilon = tb(n-1)$. The probability of rejection of the null hypothesis can be computed as follows:

$$P\left(W > f_{0.05,q,\upsilon}\right) = P\left(W^* > \left(\frac{d_0}{d}\right) f_{0.05,q,\upsilon}\right) \qquad (3.3.7)$$

$$= P\left(W^* > \frac{\left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n}\right)}{\left(\frac{\rho}{1-\rho} + \frac{1}{n}\right)} f_{0.05,q,\upsilon}\right)$$

where W^* has a non-central F distribution with q and υ degrees of freedom with non-centrality parameter $\lambda = \frac{1}{\sigma_\varepsilon^2}(H\beta - h)'\left[H\left(X'V^{-1}X\right)^{-1}H'\right]^{-1}(H\beta - h)$.

Since d_0 is an increasing function of ρ_0, increasing ρ_0 decreases (*3.3.7*) for fixed ρ.

Next, we look at the effect of the number of treatments on the significance level. The following two figures display significance levels for testing the equality of multiple treatment

means in minimal replication settings. In both cases, the number of students per class is determined such that the total error degrees of freedom, $tb(n-1)$, is equal to 96. The true value ρ is set at 0.1, a likely value for many applications, and the values of ρ_0 range from 0.025 to 0.20 to demonstrate the consequence of discrepancy between the researchers choice, ρ_0, and the actual value, ρ.

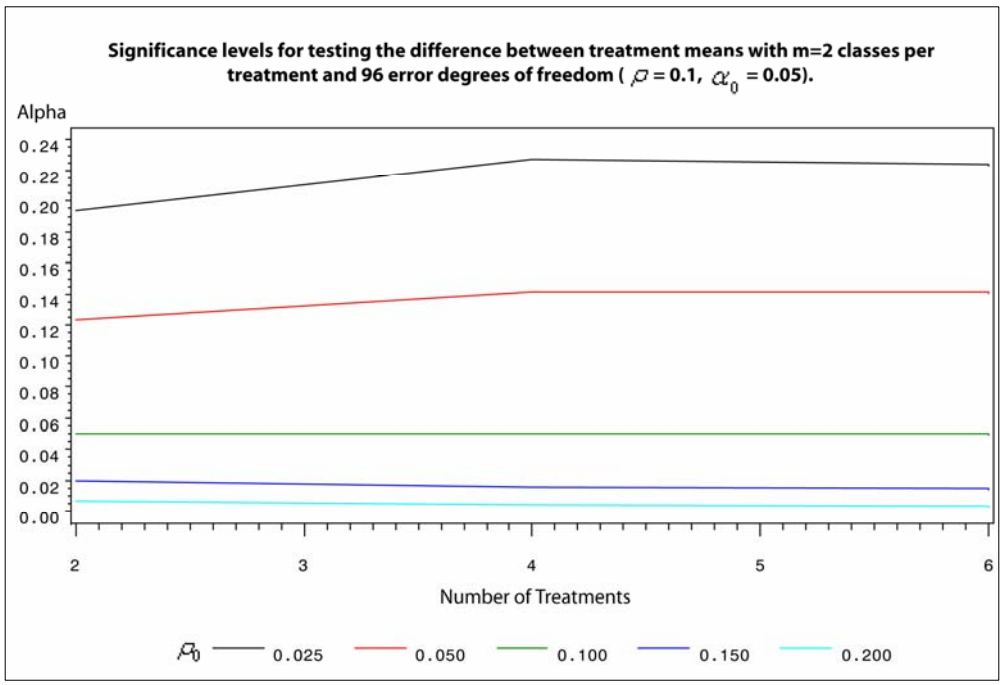


Figure 3.3.1. Significance levels with 2 classes per treatment and 96 error degrees of freedom.

In both figures it can be seen that the number of treatments involved does little to affect the significance level of the test.

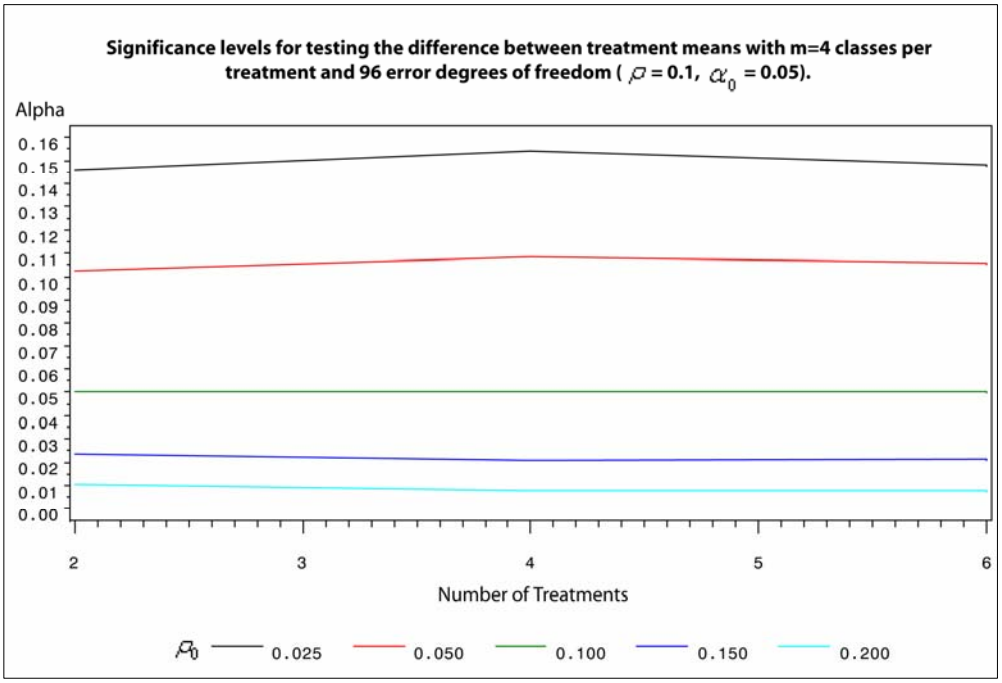


Figure 3.3.2. Significance levels with 4 classes per treatment and 96 error degrees of freedom.

We now compare two tests in the case when $n = n_{ij}$ and $b = b_i$. The first is a traditional F-test of the difference among treatment means using class means as the response variable for the design. The second test is an F-test of the difference between treatment means using the plug-in method.

Traditional F-test on Class Means

Recall the model defined previously (3.1.3).

$$w_{ij} = \mu_i + \varepsilon_{ij}^*$$

This can be rewritten in matrix form as follows:

$$W = X_w \beta + E_W \qquad (3.3.8)$$

where W is a $tb \times 1$ observation vector, X_w is a $tb \times t$ design matrix, β is a $t \times 1$ vector of treatment means, and E_w is a $tb \times 1$ error vector distributed as $E_w \sim N\left(0, \left[\sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{n}\right] I_{tb}\right)$.

The test statistic to test $H_0 : H\beta = h$ vs. $H_A : H\beta \neq h$ has the form

$$U = \frac{\left(H\hat{\beta} - h\right)'\left[H\left(X_w{}'X_w\right)^{-1}H'\right]^{-1}\left(H\hat{\beta} - h\right)}{W'\left[I - X_w\left(X_w{}'X_w\right)^{-1}X_w{}'\right]W}\left(\frac{tb - t}{q}\right) \qquad (3.3.9)$$

Let $\upsilon^* = t(b-1)$, $\sigma^2 = \sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{n}$, $\lambda^* = \frac{1}{\sigma^2}(H\beta - h)'\left[H\left(X_w{}'X_w\right)^{-1}H'\right]^{-1}(H\beta - h)$, and let F be a random variable having a noncentral F-distribution with degrees of freedom q and υ^*, and non-centrality parameter λ^*. The power of the traditional F-test can be found using the following formula:

$$Power = P\left[F > \left(F_{0.05,q,\upsilon^*,\lambda^*}\right)\right] \qquad (3.3.10)$$

F-test Using Plug-in Method

Recall that in the case of the plug-in method, $d_0 = \frac{1 + (n-1)\rho_0}{(1-\rho_0)}$. Also, let $\upsilon = tb(n-1)$. The probability of rejection of the null hypothesis can be computed as follows:

$$P\left(W > F_{0.05,q,\upsilon,\lambda}\right) = P\left(W_0 > \left(\frac{d_0}{d}\right)F_{0.05,q,\upsilon,\lambda}\right) \qquad (3.3.11)$$

where W_0 has a non-central F distribution with q and υ degrees of freedom with non-centrality parameter $\lambda = \frac{1}{\sigma_\varepsilon^2}(H\beta - h)'\left[H\left(X'V^{-1}X\right)^{-1}H'\right]^{-1}(H\beta - h)$. Note that $\lambda = \lambda^*$.

In the optimal case, $\rho_0 = \rho$, the power of the test (*3.3.11*) differs from that in (*3.3.10*) only by the degrees of freedom where we have υ^* for the traditional F-test on the class means and $\upsilon > \upsilon^*$ for the F-test using the plug-in method. As the sample size increases, eventually a point will be reached at which the power based on analyzing the test using the plug-in method

is not appreciably greater than the power based on analyzing the test using the traditional F-test.

The following four figures demonstrate the difference between the power using Equation 3.3.10 versus using Equation 3.3.11. In the figures depicting power, the "standardized difference of 1.0" in the t-treatment case means that all treatment means are equal to μ_T say, except for one treatment mean, μ_1 which equals $\mu_1 = \mu_T + \sigma(StDiff)$, where σ is the standard deviation of the measurements on students in each of the classes. Thus, $StDiff = \dfrac{\mu_1 - \mu_T}{\sigma}$.

Figure 3.3.3 illustrates the difference the number of classes (b) per treatment makes for 5 students per class. When there are five or six classes per treatment, there is not much of a difference between the power using the traditional method compared with the plug-in method. For ρ =0.2, the difference in power is 0.071.

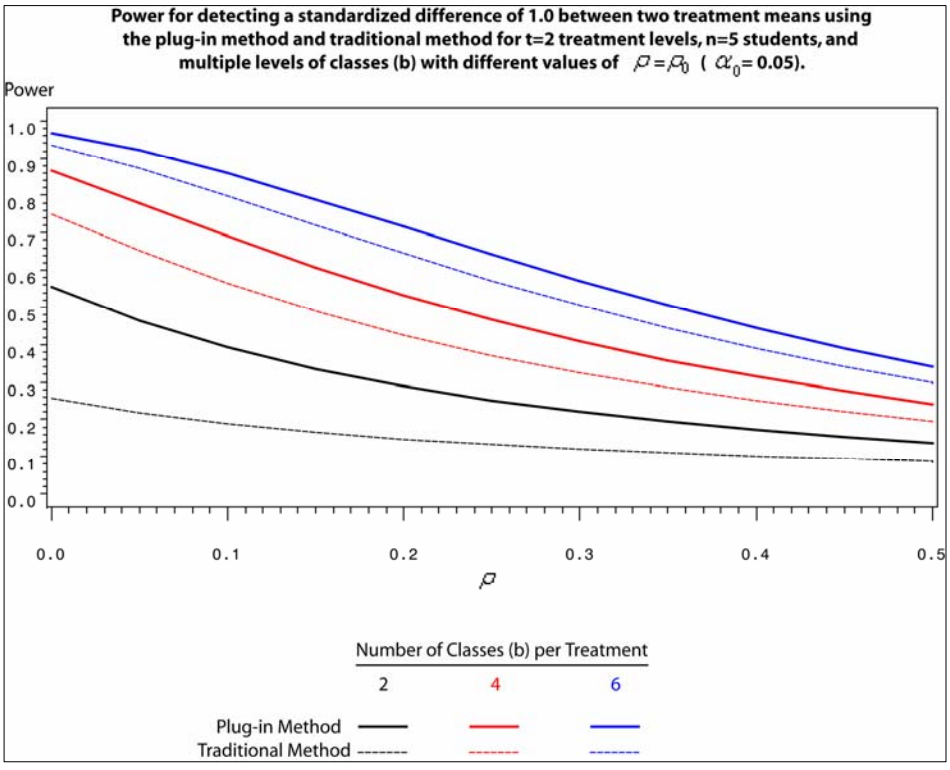


Figure 3.3.3. Power curves; 2 treatments; 5 students per class; multiple levels of classes per treatment.

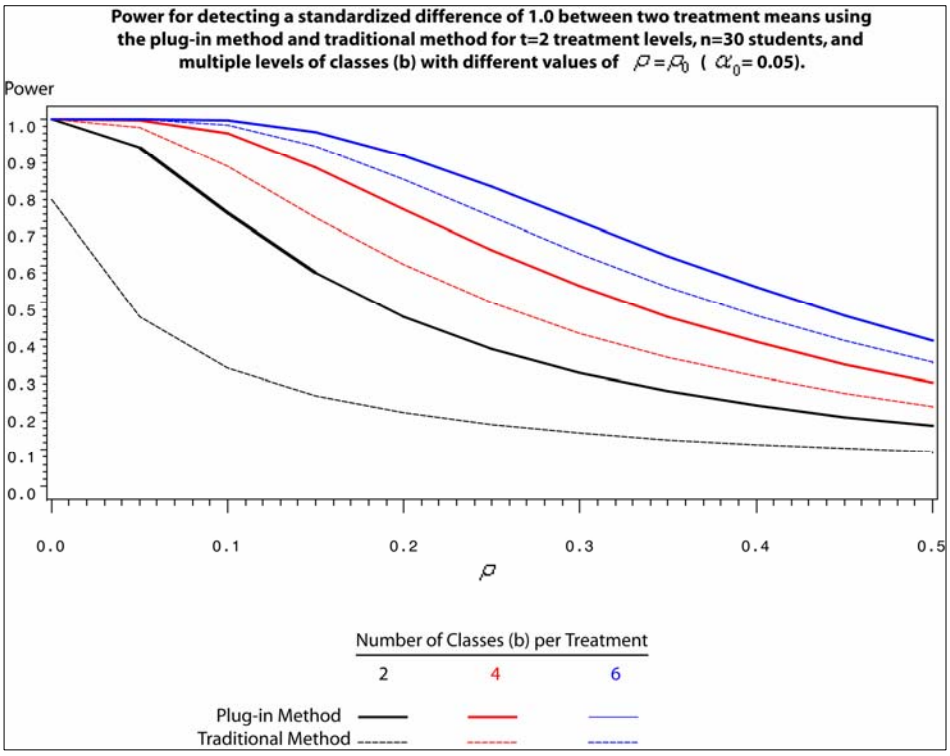


Figure 3.3.4. Power curves; 2 treatments; 30 students per class; multiple levels of classes per treatment.

Figure 3.3.4 illustrates the difference the number of classes (b) per treatment makes with 30 students per class. When looking at values of ρ between 0.05 and 0.20, again not much difference exists between the power using the traditional method and the power using the plug-in method.

The number of treatment levels also affects the difference in power, but not much. To illustrate this, the following two figures show the power of detecting a standardized difference of 1.0 with 20 treatment levels. Figure 3.3.5 has five students per class and the second figure (3.3.6) has 30 students per class.

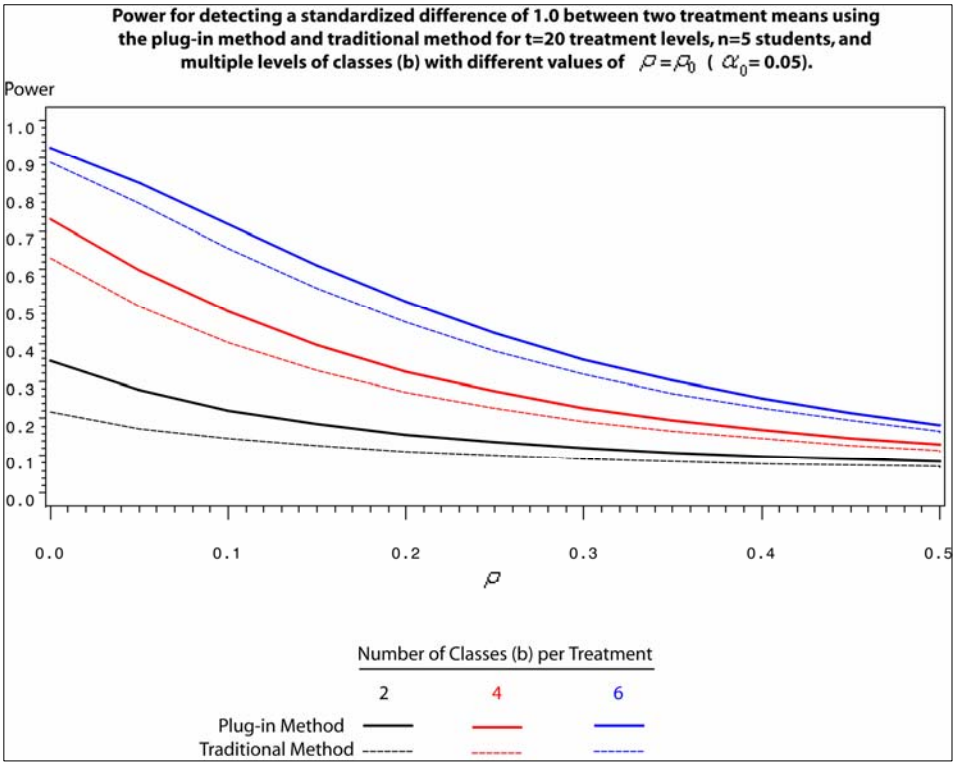


Figure 3.3.5. Power curves; 20 treatments; 5 students per class; multiple levels of classes per treatment.

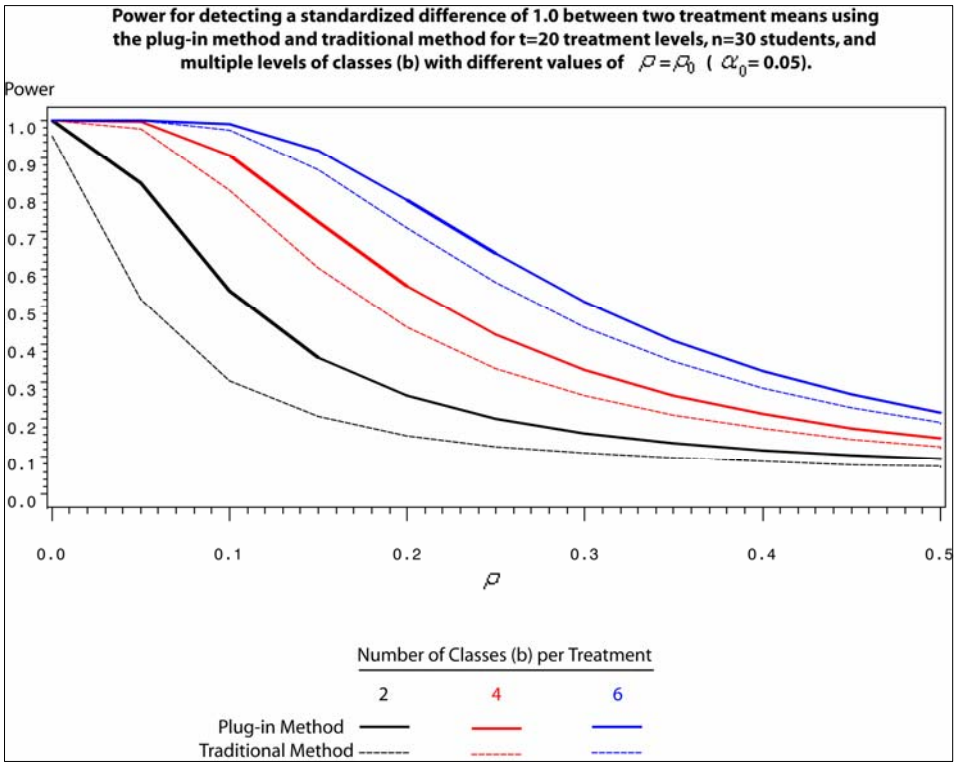


Figure 3.3.6. Power curves; 20 treatments; 30 students per class; multiple levels of classes per treatment.

In both cases (Figures 3.3.5 and 3.3.6), the traditional method will have $20(b-1)$ error degrees of freedom. The following table indicates the error degrees of freedom for each of the lines in each of the graphs:

Table 3.3.1. Error degrees of freedom totals.

Table	b	t	n	Traditional Error d.f.	Plug-in Error d.f.
3.3.3	2	2	5	2	16
3.3.3	4	2	5	6	32
3.3.3	6	2	5	10	48
3.3.4	2	2	30	2	116
3.3.4	4	2	30	6	232
3.3.4	6	2	30	10	348
3.3.5	2	20	5	20	160
3.3.5	4	20	5	60	320
3.3.5	6	20	5	100	480
3.3.6	2	20	30	20	1160
3.3.6	4	20	30	60	2320
3.3.6	6	20	30	100	3480

It can be seen in Figure 3.3.6 that although there are at least 20 error degrees of freedom, a noticeable difference in power exists between the traditional and plug-in methods of analyzing data when only two classes per treatment are considered. In fact, little difference is detectable among the two graphs of power curves with two treatment levels and those with 20 treatment levels. This can also be seen in the following two graphs (3.3.7 and 3.3.8). With b=2 classes per treatment and n=30 students the difference in power for a comparison of 20 treatments is 0.31 when rho=0.05. For the same comparison with b=6 classes per treatment, the difference in power is less than 0.001. So, in this case, it is not the error degrees of freedom, but number of classrooms (replication) that are important. It is recommended, if the researcher has more than four classes per treatment, the traditional method of analysis should be used over the plug-in method.

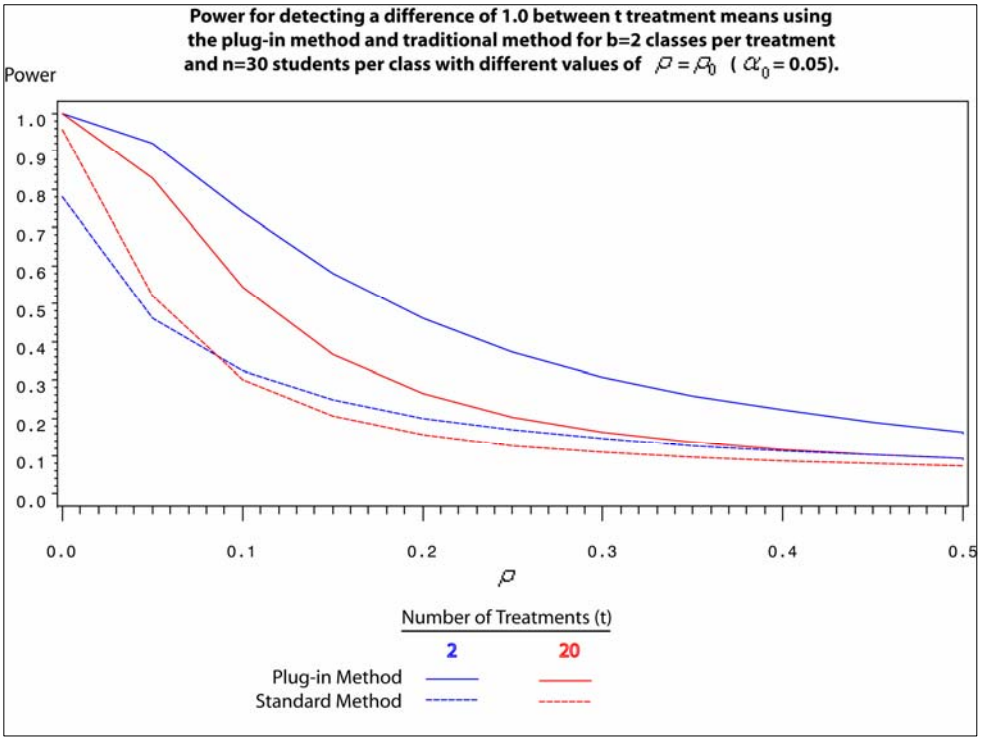


Figure 3.3.7. Power curves; 2 and 20 treatments; 2 classes per treatment; 30 students per class.

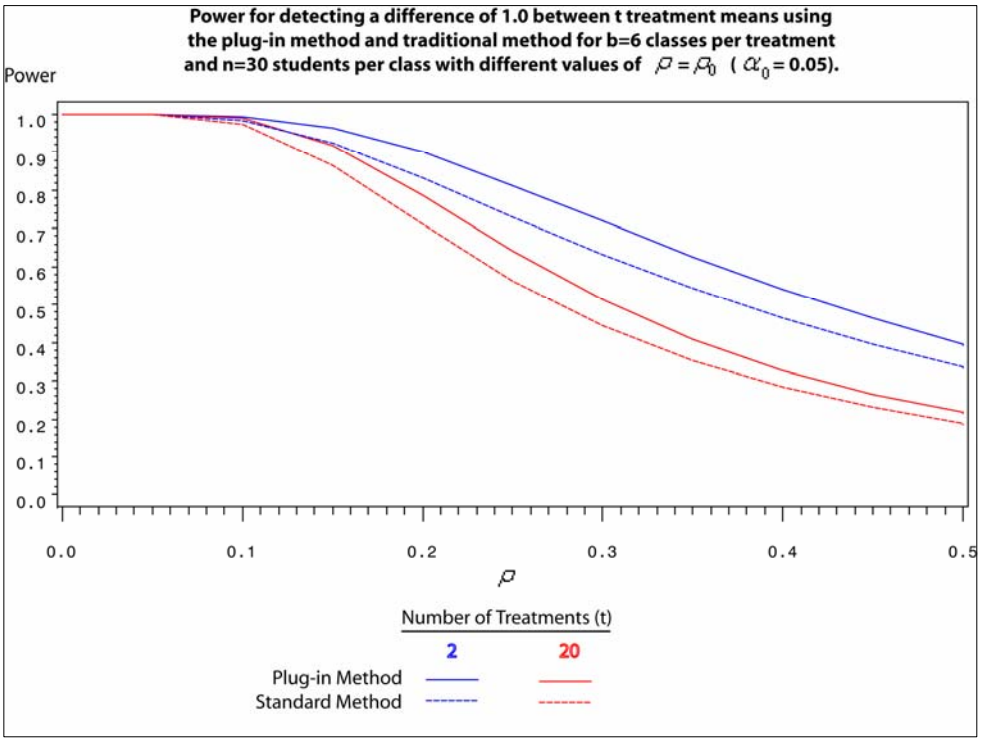


Figure 3.3.8. Power curves; 2 and 20 treatments; 6 classes per treatment; 30 students per class.

T-Treatment Strategies for Choosing ρ_0

The same strategies (Maximum Rho, Weighted P-value, etc.) used in the t-treatment un-replicated case and the two-treatment under-replicated case extend to the t-treatment case. Again we make note of the change in the formula for the Satterthwaite-type error degrees of freedom calculation.

In testing hypotheses of the form H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$, the Satterthwaite-type adjustment is given in Equation 2.3.18 as

$$d.f. = \frac{2\left[w\sigma_\varepsilon^2\right]^2}{\hat{Var}\left[w\sigma_\varepsilon^2\right]},$$

where

$$w = \frac{\rho}{1-\rho} + \frac{1}{n}$$

Likewise, Equation 2.3.19 indicated that

$$\hat{Var}\left[w\sigma_\varepsilon^2\right] = Var\left[E\left(w\sigma_\varepsilon^2 \mid \rho\right)\right] + E\left[Var\left(\sigma_\varepsilon^2 w \mid \rho\right)\right].$$

In the case of replication, this is equal to

$$= \sigma_\varepsilon^4 Var(w) + \frac{2\sigma_\varepsilon^4}{bt(n-1)} E\left(w^2\right) = \frac{\sigma_\varepsilon^4}{bt(n-1)}\left\{bt(n-1)\cdot Var(w) + 2\left[E\left(w^2\right)\right]\right\}$$

$$= \frac{\sigma_\varepsilon^4}{bt(n-1)}\left\{bt(n-1)\cdot Var(w) + 2Var(w) + \left[E(w)\right]^2\right\}$$

So,

$$d.f. = \frac{2w^2\sigma_\varepsilon^4}{\frac{\sigma_\varepsilon^4}{bt(n-1)}\left\{bt(n-1)\cdot Var(w) + 2Var(w) + 2[E(w)]^2\right\}}$$

$$= \frac{2w^2 bt(n-1)}{(btn - bt + 2)\cdot Var(w) + 2[E(w)]^2} \qquad (3.3.12)$$

and is estimated by $d.f. = \dfrac{2[E(w)]^2 bt(n-1)}{(btn - bt + 2)\cdot Var(w) + 2[E(w)]^2}$. (3.3.13)

As $Var(W) \to 0$, the error degrees of freedom using the Satterthwaite-type adjustment approaches $bt(n-1)$, the full d.f. As $n \to \infty$, the error degrees of freedom using the Satterthwaite-type adjustment approaches

$$d.f.(Satterthwaite - type) = 2\left\{\frac{\left[E\left(\frac{\rho}{1-\rho}\right)\right]^2}{Var\left(\frac{\rho}{1-\rho}\right)}\right\}.$$

CHAPTER 4: USING THE PLUG-IN METHOD ON MODELS WITH A SPLIT-PLOT DESIGN STRUCTURES

The discussion up to this point has focused on models with Completely Random design (CRD) structures. The focus of this section is on models with a split-plot design structure.

Description

The model for a split-plot design with a CRD whole plot is

$$y_{ijk} = \mu + A_i + w_{ik} + B_j + AB_{ij} + \varepsilon_{ijk} \qquad (4.1.1)$$

where y_{ijk} is the measurement taken on the kth subplot in the jth whole plot given the ith treatment, A_i is the ith whole plot treatment main effect, w_{ik} is the whole plot error effect, B_j is the jth subplot treatment main effect, AB_{ij} is the interaction effect of the ith whole plot and the jth subplot, ε_{ijk} is the subplot error effect; i = 1, 2, …,t, j = 1, 2, …, b_i, k = 1, 2, …, n_{ij}. Assume $w_{ik} \sim n\left(0, \sigma_w^2\right)$, where σ_w^2 represents the whole-plot variability and assume $\varepsilon_{ijk} \sim n\left(0, \sigma_\varepsilon^2\right)$, where σ_ε^2 represents the sub-plot variability. It is assumed that w_{ik} and ε_{ijk} are independent. Model 4.1.1 can also be written as

$$y_{ijk} = \mu_{ij} + w_{ik} + \varepsilon_{ijk} \qquad (4.1.2)$$

where $\mu_{ij} = A_i + B_j + AB_{ij}$ is the mean of the ijth treatment. The variance of an observation is $\text{var}\left(y_{ijk}\right) = \sigma_w^2 + \sigma_\varepsilon^2$. We believe that researchers who have done similar experiments may have a good idea about the value of the error variance. Similar to the case of the completely randomized design we let $\rho = \dfrac{\sigma_w^2}{\sigma_w^2 + \sigma_\varepsilon^2}$.

Example:

A researcher would like to compare the effect of three different fertilizers on the yield of corn. The researcher has four different plots of land available for the study. Each plot of land has a different irrigation system. The plot of land is divided into thirty subplots, and to each subplot, one of the three types of fertilizer is chosen at random and applied to the subplot. There is no replication on the whole plot, but subplots are replicated in a completely random fashion within each whole plot. From similar previous studies on similar fields, the researcher should have an idea of the ratio of the whole plot to the subplot variances.

Consider the equation of a split-plot model in matrix form.

$$Y = X\beta + E$$

It is the same as the completely randomized design model 3.3.1. Let V be an $N \times N$ known positive definite matrix such that $Var(Y) = \sigma_\varepsilon^2 V$.

$$V = \begin{bmatrix} A_1 & 0 & \dots & \dots & 0 \\ 0 & A_2 & \dots & \dots & 0 \\ . & & . & & . \\ . & & & . & . \\ 0 & 0 & \dots & & A_t \end{bmatrix}, \; A_i = \begin{bmatrix} B_{11} & \rho B_{12} & \dots & \rho B_{1b_i} \\ \rho B_{21} & B_{22} & \dots & \rho B_{2b_i} \\ \vdots & & \ddots & \vdots \\ \rho B_{b_i 1} & \rho B_{b_i 2} & \dots & B_{b_i b_i} \end{bmatrix},$$

$$B_{ij} = \frac{1}{1-\rho} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \qquad (4.1.4)$$

A_i is a $\sum_{j=1}^{b_i} n_{ij} \times \sum_{j=1}^{b_i} n_{ij}$ matrix, B_{ij} is a $n_{ij} \times n_{ij}$ matrix.

Let $\hat{\beta} = \left(X'V^{-1}X\right)^{-1}X'V^{-1}Y$ be the best linear unbiased estimate of β. This is equal to $\hat{\beta} = [\bar{y}_{1\cdot\cdot}, \bar{y}_{2\cdot\cdot}, ..., \bar{y}_{t\cdot\cdot}]$ in the no replication case. Again, we use Equation 3.3.3 to compute the estimate of σ_ε^2.

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{b_i}(n_{ij}-1)s_{ij}^2}{\sum_{i=1}^{t}\sum_{j=1}^{b_i}(n_{ij}-1)} \qquad (4.1.5)$$

where

$$s_{ij}^2 = \frac{\sum_{k=1}^{n_{ij}}(y_{ijk}-\bar{y}_{ij\cdot})^2}{n_{ij}-1}. \qquad (4.16)$$

The test statistic for testing H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$ is of the form

$$W = \frac{\left(H\hat{\beta}-h\right)'\left[H\left(X'V_0^{-1}X\right)^{-1}H'\right]^{-1}\left(H\hat{\beta}-h\right)}{q\hat{\sigma}_\varepsilon^2} \qquad (4.1.7)$$

where V_0 is equal to the matrix V with $\rho = \rho_0$. Let H_0 be rejected at level α_0 when $W > f_{\alpha_0,q,\upsilon}$, where $f_{\alpha_0,q,\upsilon}$ is the $1-\alpha_0$ quantile of the F distribution with q and $\upsilon = \sum_{i=1}^{t}\sum_{j=1}^{b_i}(n_{ij}-1)$ degrees of freedom.

Properties

Two-Treatment Case

We now look at the split-plot situation where there is under-replication of the whole plot. The no-replication case is the same as the replication case with b=1. Therefore, specific discussion of the no-replication two-treatment case will be omitted. For simplicity, we consider the case

where $n = n_{ij}$ for all values of i and j. Consider model (4.1.2). With a split-plot analysis, there are four different types of tests that are of interest.

The first, which we will call "Test1," tests $H_0 : \mu_{1\cdot} = \mu_{2\cdot}$ vs. $H_A : \mu_{1\cdot} \neq \mu_{2\cdot}$. This test compares the effect of whole plot treatment level one with whole plot treatment level two, averaged across all levels of the subplot. The variance of the difference of the means is

$$Var\left(\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}\right) = \frac{2}{b}\left(\sigma_w^2 + \frac{\sigma_\varepsilon^2}{n}\right), \tag{4.1.8}$$

which can be rewritten as

$$Var\left(\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}\right) = \frac{2\sigma_\varepsilon^2}{b}\left(\frac{\rho}{1-\rho} + \frac{1}{n}\right) \tag{4.1.9}$$

which is exactly the same as in the CRD two-treatment under-replication case. The test statistic for this test using the plug-in method is

$$T1 = \frac{\left(\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}\right) - \left(\mu_{01\cdot\cdot} - \mu_{02\cdot\cdot}\right)}{\sqrt{\frac{2\hat{\sigma}_\varepsilon^2}{b}\left(\frac{\rho_0}{1-\rho_0}\right)}} \tag{4.1.10}$$

and the probability of rejecting the null hypothesis is the same as in Equation 3.1.4.

The second test we define as "Test2" and write as $H_0 : \mu_{\cdot j} = \mu_{\cdot j'}$ vs. $H_A : \mu_{\cdot j} \neq \mu_{\cdot j'}$. This test compares the effect of subplot treatment level j with subplot treatment level j', averaged across all levels of the whole plot. The variance of the difference of the means is

$$Var\left(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot j'\cdot}\right) = \frac{\sigma_\varepsilon^2}{bn}, \tag{4.1.11}$$

which does not depend on ρ at all, and therefore is not affected by use of the plug-in method. The test statistic for this test is

$$T2 = \frac{\left(\bar{y}_{\cdot 1\cdot} - \bar{y}_{\cdot 2\cdot}\right) - \left(\mu_{0\cdot 1\cdot} - \mu_{0\cdot 2\cdot}\right)}{\sqrt{\frac{\hat{\sigma}_{\varepsilon}^2}{bn}}} \tag{4.1.12}$$

and the probability of rejecting the null hypothesis is

$$P\left(T2 > t_{\alpha,2b(n-1)}\right) = P\left(T > t_{\alpha,2b(n-1)}\right) \text{ where } T \sim t_{\alpha,2b(n-1)}. \tag{4.1.13}$$

The third test we refer to as "Test3" and write as $H_0 : \mu_{1j} = \mu_{2j}$ vs. $H_A : \mu_{1j} \neq \mu_{2j}$. This test compares the effect of the whole plot treatment levels, at the j^{th} subplot level. The variance of the difference of the means is

$$Var\left(\bar{y}_{1j\cdot} - \bar{y}_{2j\cdot}\right) = \frac{2\sigma_{\varepsilon}^2}{bn}\left(\frac{1}{1-\rho}\right), \tag{4.1.14}$$

The test statistic for this test using the plug-in method is:

$$T3 = \frac{\left(\bar{y}_{1j\cdot} - \bar{y}_{2j\cdot}\right) - \left(\mu_{01j\cdot} - \mu_{02j\cdot}\right)}{\sqrt{\frac{2\hat{\sigma}_{\varepsilon}^2}{bn}\left(\frac{1}{1-\rho_0}\right)}} \tag{4.1.15}$$

and the probability of rejecting the null hypothesis is

$$P\left(T3 > t_{0.05,2b(n-1)}\right) = P\left(T' > \left(t_{0.05,2b(n-1)}\right)\sqrt{\frac{1-\rho}{1-\rho_0}}\right) \tag{4.1.16}$$

where T' has a non-central t-distribution with non-centrality parameter $\lambda = \dfrac{(\mu_{1j} - \mu_{2j}) - (\mu_{01j} - \mu_{02j})}{\sqrt{\dfrac{2\sigma_\varepsilon^2}{bn}\left(\dfrac{1}{1-\rho}\right)}}$

and degrees of freedom $\upsilon = 2b(n-1)$.

The fourth test we define as "Test4" and write as $H_0 : \mu_{ij} = \mu_{ij'}$ vs. $H_A : \mu_{ij} \neq \mu_{ij'}$. This test compares the effect of subplot treatment levels j and j' at the i^{th} whole plot level. The variance of the difference of the means is

$$Var(\bar{y}_{ij\cdot} - \bar{y}_{ij'\cdot}) = \frac{2\sigma_\varepsilon^2}{bn}, \qquad (4.1.17)$$

which does not depend on ρ at all, and therefore is not affected by use of the plug-in method. The test statistic for this test is:

$$T4 = \frac{(\bar{y}_{ij\cdot} - \bar{y}_{ij'\cdot}) - (\mu_{0ij\cdot} - \mu_{0ij'\cdot})}{\sqrt{\dfrac{2\hat{\sigma}_\varepsilon^2}{bn}}} \qquad (4.1.18)$$

and the probability of rejecting the null hypothesis is

$$P(T4 > t_{0.05,2b(n-1)}) = P(T > t_{0.05,2b(n-1)}) \text{ where } T \sim t_{0.05,2b(n-1)}. \qquad (4.1.19)$$

No-replication T-treatment Case

Consider the matrix form of the model (*4.1.2*) and the test statistic *W* that follows it (*4.1.7*). It can be shown that $(X'V^{-1}X)^{-1} = \left[\dfrac{1}{(1-\rho)}\right] \times \left(\dfrac{1}{n}\right) I_t$. Also, $(X'X)^{-1} = \left(\dfrac{1}{n}\right) I_t$. So, the contribution of the *V* matrix is

$$d = \frac{1}{(1-\rho)}. \qquad (4.2.20)$$

Let $d_0 = \frac{1}{(1-\rho_0)}$. The probability of rejection of the null hypothesis can be computed as follows:

$$P(W > F_{0.05,q,N-t}) = P\left(W^* > \left(\frac{d_0}{d}\right)F_{0.05,q,N-t}\right) \qquad (4.1.21)$$

where *W** has a non-central F distribution with *q* and *N-t* degrees of freedom with non-centrality parameter $\lambda = \frac{1}{\sigma_\varepsilon^2}(H\beta - h)'\left[H\left(X'V^{-1}X\right)^{-1}H'\right]^{-1}(H\beta - h)$.

Since d_0 is an increasing function of ρ_0, increasing ρ_0 decreases (*4.1.21*) for fixed ρ.

Replication T-treatment Case

Consider the general test of hypothesis H_0: $H\beta = h$ vs. H_A: $H\beta \neq h$. In the CRD case, it was possible to find a constant d_0 (see Equation 3.3.5) such that $\left(X'V_0^{-1}X\right)^{-1} = d_0\left(X'X\right)^{-1}$. This allowed for the probability of rejection to be computed analytically using a constant multiple of a standard F-distribution (*3.3.7*). This is not true in the split-plot case. In the split-plot case, the probability of rejection must be found using simulation.

CHAPTER 5: AN EXAMPLE

5.1: DESCRIPTION

A researcher would like to test the effectiveness of four different treatments on the control of two-spotted spider mites (TSM) in commercial greenhouses. Four greenhouses were used for the study. Within each greenhouse, eight potted plants were inoculated with TSM. One of the four treatments was applied to all the pots in each of the four greenhouses. At the end of the treatment period, the number of TSM was counted on each of the potted plants in each of the greenhouses. The data are listed in Appendix C as data set C.3.

Treatment 1 was a control. No methods of pest control were performed. Treatment 2 was a biological control agent (PP). Two weeks after the plants were inoculated with TSM, sampling was done on ivy geranium plants according to an existing sampling plan. Using the sampling plan, the number of TSM per leaf was determined. The ratio of PP:TSM known to effectively control TSM is 1:4. To determine how many PP to release, the total number of TSM in the area occupied by plants was divided by four—PP was released only once, two weeks after plants were inoculated using TSM. Treatment 3 was also a biological control agent. Two weeks after plants were inoculated with TSM, PP was released at a rate of $50/m^2$ - a rate recommended by insectaries that sell natural enemies. Release of PP was done on a weekly basis for four weeks. Treatment 4 involves application of a chemical insecticide. Two weeks after plants were inoculated with TSM, a single chemical application was done.

The response variable was the number of spider mites recorded on a given potted plant in a given greenhouse with a given treatment. The following box plots represent the data for the different treatments:

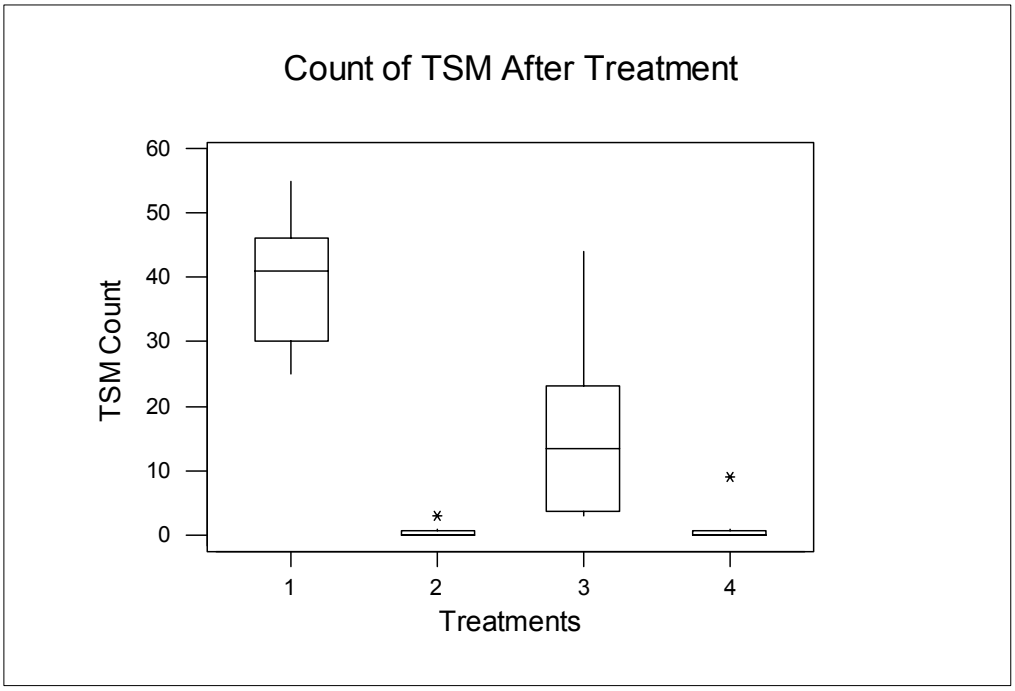


Figure 5.1.1. Count of TSM after treatment.

As can be seen in the Figure 5.1.1, a convincing difference seems to exist between the count of TSM in the greenhouse treated with treatment 1 and those treated with treatments 2 and 4. However, there is question about the difference between the TSM counts under treatment 3 versus the others. Specifically, treatments 2 and 4 appear to reduce significantly the count of TSM. Treatment 3 is questionable. Both Bartlett's and Levene's tests of homogeneity of variance indicate unequal variances with p-values < .001 and = .012, respectively. This leads to a square-root transformation on the response variable TSM count. Once a square-root transformation was performed neither Bartlett's or Levene's tests of homogeneity of variance indicated significant heterogeneity of variances with p-values = .081 and .205, respectively. The following analyses use the transformed data rather than the original counts.

5.2: CONDITIONAL P-VALUE PLOTS

The first step in an analysis using plug-in methods is to look at the conditional p-value plots. If the p-value plots are very conclusive, it may eliminate the need for certain hypothesis tests.

The first plot (5.2.1) shows the p-value of the test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. H_A : *At least two means differ*. The p-values are plotted as a function of the plug-in value ρ_0.

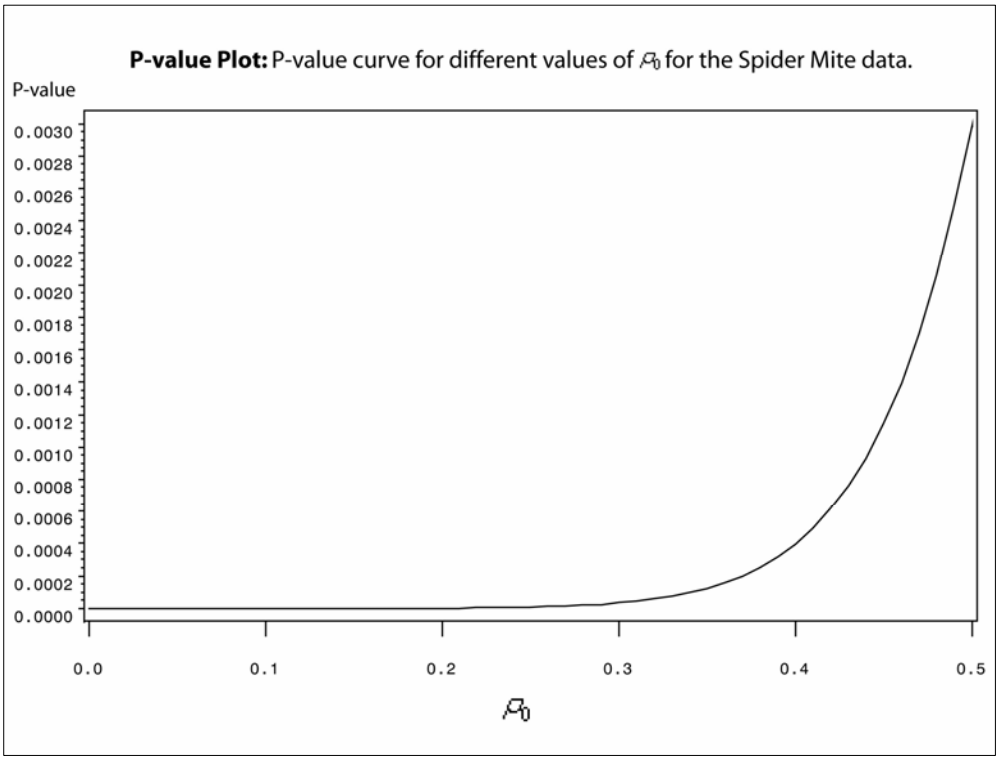


Figure 5.2.1. P-value plot for testing equality of treatment means.

The p-value is less than α_0 =0.05 for every value of ρ_0 <0.5. So, using any of the plug-in methods described in this paper with any value of ρ_0 <0.5 will lead to rejection of the null hypothesis. However, for the purpose of illustration, we will give an example of each strategy proposed in this paper as it applies to this overall F-test.

5.3: WHAT DOES THE RESEARCHER KNOW ABOUT ρ

Before treatments were applied to the greenhouses (week two), the count of TSM for each potted plant in each greenhouse was determined on several different varieties of plants. That

data (C.4) was used to compute estimated values of ρ for each variety. The following values were obtained:

Table 5.3.1. Estimated ICC for different plant varieties.

Variety	**Estimated ρ**
Cajun Cranberry	0.07275
Cajun White	0.21830
Impulse Orange	0.15886
Impulse Orange White	0.16666
Ivy Geranium	0.05467
Summer Rose Lilac	0.06480
Summer Rose Red	0.30069

The values in the table have a maximum value just over 0.30 and a minimum value just over 0.05. Next we look at applying each of the plug-in strategies to the overall F-test.

5.4: OVERALL F-TEST

Strategy 1: Maximum Rho

The estimates of ρ lead us to select ρ_0=0.30. Using this value leads to a p-value less than 0.0001. The result is to reject the null hypothesis at α_0=0.05. The following is the SAS® output for this test.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	3	28	11.95	<.0001

Strategy 2: Acceptable Interval (Best-Worst-Case Scenario)

Using ρ_0=0.05 and ρ_0=0.30 leads to p-values less than 0.0001. The result is to reject the null hypothesis at α_0=0.05. The following is the SAS® output for this test.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	3	28	36.05	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	3	28	11.95	<.0001

Strategy 3: ADFMR

We probably do not have sufficient distributional data to compute Satterthwaite-type adjusted error degrees of freedom. However, strictly for illustration purposes, we will use the seven estimates of ρ to compute the adjusted error degrees of freedom, 7.292. The full error degrees of freedom are 4(8-1)=28.

Once the error degrees of freedom have been determined, the choice of ρ_0 must be determined. Using the mean of the seven estimates of ρ we get ρ_0=0.148104. Using this value and the full 28 error degrees of freedom leads to a p-value less than 0.0001. The result is to reject the null hypothesis at α_0=0.05. The following is the SAS® output for the test using 28 error degrees of freedom.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	3	28	21.93	<.0001

Using the reduced 7.292 error degrees of freedom leads to a p-value of 0.0006. The result is to reject the null hypothesis at α_0=0.05. The following is the output for the test using 7.292 error degrees of freedom.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
trt	3	7.29	20.62	0.0006

Strategy 4: Weighted P-value

Using the seven estimates of ρ and weights equal to 1/7, the weighted p-value strategy yields a weighted p-value of <0.0002. The result is to reject the null hypothesis at α_0=0.05.

5.5: MULTIPLE COMPARISONS

Next, it is determined that each of the treatment means be compared. Again, each of the strategies will be implemented for the pairwise comparisons.

Strategy 1: Maximum Rho

Again using ρ_0=0.30, the following SAS® output results:

Differences of Least Squares Means

Effect	trt	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
trt	1	2	5.8571	1.1852	28	4.94	<.0001	Bonferroni	0.0002
trt	1	3	2.5904	1.1852	28	2.19	0.0374	Bonferroni	0.2242
trt	1	4	5.6986	1.1852	28	4.81	<.0001	Bonferroni	0.0003
trt	2	3	-3.2668	1.1852	28	-2.76	0.0102	Bonferroni	0.0610
trt	2	4	-0.1585	1.1852	28	-0.13	0.8946	Bonferroni	1.0000
trt	3	4	3.1083	1.1852	28	2.62	0.0140	Bonferroni	0.0838
trt	1	2	5.8571	1.1852	28	4.94	<.0001	Tukey	0.0002
trt	1	3	2.5904	1.1852	28	2.19	0.0374	Tukey	0.1520
trt	1	4	5.6986	1.1852	28	4.81	<.0001	Tukey	0.0003
trt	2	3	-3.2668	1.1852	28	-2.76	0.0102	Tukey	0.0472
trt	2	4	-0.1585	1.1852	28	-0.13	0.8946	Tukey	0.9991
trt	3	4	3.1083	1.1852	28	2.62	0.0140	Tukey	0.0632

trt	1	2	5.8571	1.1852	28	4.94	<.0001	Simulate	0.0002
trt	1	3	2.5904	1.1852	28	2.19	0.0374	Simulate	0.1520
trt	1	4	5.6986	1.1852	28	4.81	<.0001	Simulate	0.0003
trt	2	3	-3.2668	1.1852	28	-2.76	0.0102	Simulate	0.0472
trt	2	4	-0.1585	1.1852	28	-0.13	0.8946	Simulate	0.9991
trt	3	4	3.1083	1.1852	28	2.62	0.0140	Simulate	0.0632

The following differences are considered significant for the following methods at α_0 =0.05 using ρ_0 =0.30:

Table 5.5.1. Significant differences in treatment level means for different multiple comparison methods using the Maximum Rho method.

Method	Significant Treatment Level Means
Fisher's Protected LSD	1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 3 vs. 4
Bonferroni	1 vs. 2, 1 vs. 4
Tukey	1 vs. 2, 1 vs. 4, 2 vs. 3
Simulate	1 vs. 2, 1 vs. 4, 2 vs. 3

Strategy 2: Acceptable Interval (Best-Worst-Case Scenario)

Using ρ_0 =0.05, the following SAS® output results:

Differences of Least Squares Means

Effect	trt	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
trt	1	2	5.8571	0.6714	28	8.72	<.0001	Bonferroni	<.0001
trt	1	3	2.5904	0.6714	28	3.86	0.0006	Bonferroni	0.0037
trt	1	4	5.6986	0.6714	28	8.49	<.0001	Bonferroni	<.0001
trt	2	3	-3.2668	0.6714	28	-4.87	<.0001	Bonferroni	0.0002
trt	2	4	-0.1585	0.6714	28	-0.24	0.8151	Bonferroni	1.0000
trt	3	4	3.1083	0.6714	28	4.63	<.0001	Bonferroni	0.0005
trt	1	2	5.8571	0.6714	28	8.72	<.0001	Tukey	<.0001
trt	1	3	2.5904	0.6714	28	3.86	0.0006	Tukey	0.0032
trt	1	4	5.6986	0.6714	28	8.49	<.0001	Tukey	<.0001
trt	2	3	-3.2668	0.6714	28	-4.87	<.0001	Tukey	0.0002
trt	2	4	-0.1585	0.6714	28	-0.24	0.8151	Tukey	0.9953
trt	3	4	3.1083	0.6714	28	4.63	<.0001	Tukey	0.0004
trt	1	2	5.8571	0.6714	28	8.72	<.0001	Simulate	<.0001
trt	1	3	2.5904	0.6714	28	3.86	0.0006	Simulate	0.0032
trt	1	4	5.6986	0.6714	28	8.49	<.0001	Simulate	<.0001
trt	2	3	-3.2668	0.6714	28	-4.87	<.0001	Simulate	0.0002
trt	2	4	-0.1585	0.6714	28	-0.24	0.8151	Simulate	0.9953
trt	3	4	3.1083	0.6714	28	4.63	<.0001	Simulate	0.0004

Using ρ_0 =0.30, the following SAS® output results:

Differences of Least Squares Means

Effect	trt	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
trt	1	2	5.8571	1.1852	28	4.94	<.0001	Bonferroni	0.0002
trt	1	3	2.5904	1.1852	28	2.19	0.0374	Bonferroni	0.2242
trt	1	4	5.6986	1.1852	28	4.81	<.0001	Bonferroni	0.0003
trt	2	3	-3.2668	1.1852	28	-2.76	0.0102	Bonferroni	0.0610
trt	2	4	-0.1585	1.1852	28	-0.13	0.8946	Bonferroni	1.0000
trt	3	4	3.1083	1.1852	28	2.62	0.0140	Bonferroni	0.0838
trt	1	2	5.8571	1.1852	28	4.94	<.0001	Tukey	0.0002
trt	1	3	2.5904	1.1852	28	2.19	0.0374	Tukey	0.1520
trt	1	4	5.6986	1.1852	28	4.81	<.0001	Tukey	0.0003
trt	2	3	-3.2668	1.1852	28	-2.76	0.0102	Tukey	0.0472
trt	2	4	-0.1585	1.1852	28	-0.13	0.8946	Tukey	0.9991
trt	3	4	3.1083	1.1852	28	2.62	0.0140	Tukey	0.0632
trt	1	2	5.8571	1.1852	28	4.94	<.0001	Simulate	0.0002
trt	1	3	2.5904	1.1852	28	2.19	0.0374	Simulate	0.1520
trt	1	4	5.6986	1.1852	28	4.81	<.0001	Simulate	0.0003
trt	2	3	-3.2668	1.1852	28	-2.76	0.0102	Simulate	0.0472
trt	2	4	-0.1585	1.1852	28	-0.13	0.8946	Simulate	0.9991
trt	3	4	3.1083	1.1852	28	2.62	0.0140	Simulate	0.0632

Under ρ_0 =0.05 for all methods, all treatment level means differ significantly except for 2 vs. 4. So, we conclude those means considered different under ρ_0 =0.30 are different, mean 2 vs. mean 4 is not different, and the other differences are questionable.

Strategy 3: ADFMR

The Satterthwaite-type adjusted error degrees of freedom are 7.292. The full error degrees of freedom are 4(8-1)=28. Using the mean of the seven estimates of ρ we get ρ_0 =0.148104. Using ρ_0 =0.148104 and full error degrees of freedom=28 leads to the following SAS® output:

Differences of Least Squares Means

Effect	trt	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
trt	1	2	5.8571	0.8708	28	6.73	<.0001	Bonferroni	<.0001
trt	1	3	2.5904	0.8708	28	2.97	0.0060	Bonferroni	0.0359
trt	1	4	5.6986	0.8708	28	6.54	<.0001	Bonferroni	<.0001
trt	2	3	-3.2668	0.8708	28	-3.75	0.0008	Bonferroni	0.0049
trt	2	4	-0.1585	0.8708	28	-0.18	0.8569	Bonferroni	1.0000
trt	3	4	3.1083	0.8708	28	3.57	0.0013	Bonferroni	0.0079
trt	1	2	5.8571	0.8708	28	6.73	<.0001	Tukey	<.0001
trt	1	3	2.5904	0.8708	28	2.97	0.0060	Tukey	0.0288
trt	1	4	5.6986	0.8708	28	6.54	<.0001	Tukey	<.0001

trt	2	3	-3.2668	0.8708	28	-3.75	0.0008	Tukey	0.0043
trt	2	4	-0.1585	0.8708	28	-0.18	0.8569	Tukey	0.9978
trt	3	4	3.1083	0.8708	28	3.57	0.0013	Tukey	0.0068
trt	1	2	5.8571	0.8708	28	6.73	<.0001	Simulate	<.0001
trt	1	3	2.5904	0.8708	28	2.97	0.0060	Simulate	0.0288
trt	1	4	5.6986	0.8708	28	6.54	<.0001	Simulate	<.0001
trt	2	3	-3.2668	0.8708	28	-3.75	0.0008	Simulate	0.0043
trt	2	4	-0.1585	0.8708	28	-0.18	0.8569	Simulate	0.9978
trt	3	4	3.1083	0.8708	28	3.57	0.0013	Simulate	0.0068

All the methods agree. The differences considered statistically significant at α_0 =0.05 are 1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 3 vs. 4.

Using ρ_0 =0.148104 and the reduced 7.292 error degrees of freedom leads to the following SAS® output:

Differences of Least Squares Means

Effect	trt	trt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
trt	1	2	5.8571	0.8708	7.29	6.73	0.0002	Bonferroni	0.0014
trt	1	3	2.5904	0.8708	7.29	2.97	0.0197	Bonferroni	0.1183
trt	1	4	5.6986	0.8708	7.29	6.54	0.0003	Bonferroni	0.0016
trt	2	3	-3.2668	0.8708	7.29	-3.75	0.0066	Bonferroni	0.0399
trt	2	4	-0.1585	0.8708	7.29	-0.18	0.8605	Bonferroni	1.0000
trt	3	4	3.1083	0.8708	7.29	3.57	0.0085	Bonferroni	0.0510
trt	1	2	5.8571	0.8708	7.29	6.73	0.0002	Tukey	0.0010
trt	1	3	2.5904	0.8708	7.29	2.97	0.0197	Tukey	0.0747
trt	1	4	5.6986	0.8708	7.29	6.54	0.0003	Tukey	0.0012
trt	2	3	-3.2668	0.8708	7.29	-3.75	0.0066	Tukey	0.0268
trt	2	4	-0.1585	0.8708	7.29	-0.18	0.8605	Tukey	0.9977
trt	3	4	3.1083	0.8708	7.29	3.57	0.0085	Tukey	0.0340
trt	1	2	5.8571	0.8708	7.29	6.73	0.0002	Simulate	0.0010
trt	1	3	2.5904	0.8708	7.29	2.97	0.0197	Simulate	0.0747
trt	1	4	5.6986	0.8708	7.29	6.54	0.0003	Simulate	0.0012
trt	2	3	-3.2668	0.8708	7.29	-3.75	0.0066	Simulate	0.0268
trt	2	4	-0.1585	0.8708	7.29	-0.18	0.8605	Simulate	0.9977
trt	3	4	3.1083	0.8708	7.29	3.57	0.0085	Simulate	0.0340

The following differences are considered significant for the following methods at α_0 =0.05 using the reduced 7.292 error degrees of freedom:

Table 5.5.2. Significant differences in treatment level means for different multiple comparison methods using the ADFMR method.

Method	Significant Treatment Level Means
Fisher's Protected LSD	1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 3 vs. 4
Bonferroni	1 vs. 2, 1 vs. 4, 2 vs. 3
Tukey	1 vs. 2, 1 vs. 4, 2 vs. 3, 3 vs. 4
Simulate	1 vs. 2, 1 vs. 4, 2 vs. 3, 3 vs. 4

Strategy 4: Weighted P-value

Using the seven estimates of ρ and weights equal to 1/7, the weighted p-value strategy yields weighted p-values according to the following table:

Table 5.5.3. Significant differences in treatment level means for different multiple comparison methods using the Weighted P-value method.

	Multiple Comparison Adjustment Method				
Comparison	**Bonferroni**	**Simulate**	**Tukey**	**LSD**	**Result**
trt1 vs. trt2	0.000248	0.000225	0.000225	0.000041	Reject H_0
trt1 vs. trt3	0.022488	0.017489	0.017489	0.003748	Reject H_0
trt1 vs. trt4	0.000312	0.000282	0.000282	0.000052	Reject H_0
trt2 vs. trt3	0.245456	0.153403	0.153403	0.040909	FTR H_0
trt2 vs. trt4	1.000000	0.999373	0.999373	0.908709	FTR H_0
trt3 vs. trt4	0.300138	0.182297	0.182297	0.050023	FTR H_0

According to table 5.5.3, treatment means 2, 3, and 4 are all significantly different than the control. Fisher's Protected LSD also shows treatment means 2 and 3 differ significantly from each other. However, as stated earlier, Fisher's Protected LSD does not control the family-wise error rate in cases with more than three treatment levels—which is the case here.

5.6: FINAL COMMENTS REGARDING THE ANALYSIS

It is obvious from Figure 5.1.1 that a difference exists between the means for the control and some of the treatment levels. However, without replication, standard methods fail. 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.

Insufficient data exists for an exact choice for the value of ρ_0. However, the results make it clear the value of ρ should be below 0.5. This would indicate a situation in which the variability among greenhouses is small compared with the variability between pots and the overall variability. This is a reasonable assumption, supported by the data.

When it comes to the pairwise comparisons, again, Figure 5.1.1 shows a clear drop in the number of TSM on the potted plants treated with all three treatments versus the control, with extreme drops with treatments 2 and 4. There is a questionable difference between treatment levels 2 and 4 vs. 3 that is supported by some of the methods. However, there is some overlap in Figure 5.1.1. Based on the analysis, one should recommend to the researcher that treatments 2 and 4 controlled the TSM infestation. Treatment 3 seems to reduce the TSM infestation, just not as much or as clearly as the other two methods.

CHAPTER 6: FUTURE DIRECTION AND SUMMARY

6.1: RANDOMIZED COMPLETE BLOCK DESIGN

Another design structure to look at is the randomized complete block design (RCBD). The focus of this section is on models with RCBD design structures.

Description

The model for a RCBD can be expressed as follows:

$$y_{ijk} = \mu_i + r_j + \delta_{ij} + \varepsilon_{ijk} \qquad (6.1.1)$$

where y_{ijk} is the measurement taken on the k^{th} individual in the j^{th} block given the i^{th} treatment, μ_i is the fixed effect of treatment i, r_j is the random effect of the j^{th} block, δ_{ij} is the block × treatment interaction, and ε_{ijk} is the random effect of individual k in block j given treatment i; i = 1, 2, …,t j = 1, 2, …, b_i k = 1, 2, …, n_{ij}. Assume $\delta_{ij} \sim n\left(0, \sigma_\delta^2\right)$, where σ_δ^2 represents the unit-to-unit variability of a block × treatment combination; assume $\varepsilon_{ijk} \sim n\left(0, \sigma_\varepsilon^2\right)$, where σ_ε^2 represents the subunit-to-subunit variability within a unit × block × treatment combination; assume $r_j \sim n\left(0, \sigma_R^2\right)$, where σ_R^2 represents the block-to-block variability. It is assumed that r_j is independent of δ_{ij} and ε_{ijk}.

The variance of an observation is $\text{var}\left(y_{ijk}\right) = \sigma_\delta^2 + \sigma_\varepsilon^2 + \sigma_R^2$. The variance of the difference between two treatment means is

$$\text{var}\left(\bar{y}_{1\cdot\cdot} - \bar{y}_{2\cdot\cdot}\right) = \frac{b_1 + b_2}{b_1 b_2}\left(\sigma_R^2 + \sigma_\delta^2\right) + \frac{n_{1\cdot} + n_{2\cdot}}{n_{1\cdot} n_{2\cdot}}\sigma_\varepsilon^2. \qquad (6.1.2)$$

However, in the balanced case, $\text{var}(\bar{y}_{1..} - \bar{y}_{2..}) = 2\left(\sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{bn}\right)$. We believe that researchers may have reliable information on the value of $\rho = \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma_\varepsilon^2}$ in many situations. One might also consider the quantity

$$\rho^* = \frac{\text{cov}(y_{ijk}, y_{ijk'})}{\sqrt{\text{var}(y_{ijk})\text{var}(y_{ijk'})}} = \frac{\sigma_\delta^2 + \sigma_R^2}{\sqrt{(\sigma_\delta^2 + \sigma_\varepsilon^2 + \sigma_R^2)(\sigma_\delta^2 + \sigma_\varepsilon^2 + \sigma_R^2)}} = \frac{\sigma_\delta^2 + \sigma_R^2}{\sigma_\delta^2 + \sigma_\varepsilon^2 + \sigma_R^2}.$$ However, in designing experiments, blocks are chosen to separate classifiable error from random error. Thus, block-to-block variability σ_R^2 may be small or large, relative to the other components of variance and depending on what the researcher decided to use as blocks. So, quantities involving σ_R^2 may not be constant even in similar studies. In the case of balanced data, σ_R^2 drops out of Equation 6.1.2. So, ρ may be fairly constant in many situations. Knowing reliable information on the value of ρ allows for hypothesis testing in cases of un- and under-replicated experiments.

Example

A researcher is interested in testing the effects of four different treatments on hospital patients. Each treatment is given to one hundred different patients for a total of four hundred patients per hospital. The researcher would like the results to apply to different hospitals. Only three hospitals agree to allow the research. Similar studies have been conducted in the past with the involvement of many hospitals. In this case the ratio of the variability between hospitals and the variability within a hospital may be a knowable quantity, and that it may be fairly constant for studies of a similar nature.

We will consider the two-treatment under-replicated balanced case. The test statistic for testing H_0: $\mu_1 = \mu_2$ vs. H_A: $\mu_1 \neq \mu_2$ is of the form

$$T = \frac{(\bar{y}_{1\cdot} - \bar{y}_{2\cdot}) - (\mu_{01} - \mu_{02})}{\sqrt{\frac{2\hat{\sigma}_{\varepsilon}^2}{b}\left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n}\right)}} \qquad (6.1.3)$$

where $\hat{\sigma}_{\varepsilon}^2$ represents a pooled estimate of the within class variance. Let

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\sum_{i=1}^{2}\sum_{j=1}^{b} s_{ij}^2}{2b} \qquad (6.1.4)$$

$$\text{where } s_{ij}^2 = \frac{\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2}{n-1}. \qquad (6.1.5)$$

Properties

In a balanced case, the test statistic for the RCBD case (*6.1.3*) is the same as the test statistic for the CRD case (*2.1.1*) and the probability of rejecting the null hypothesis is the same as in Equation 2.1.2. Likewise, the full t-treatment RCBD case should closely resemble the full t-treatment CRD case discussed in this paper. The difference between the two will be the form of the V matrix.

6.2: USING A POSTERIOR DISTRIBUTION OF THE ICC TO INCREASE THE ACCURACY OF THE CHOICE OF THE ICC

In the case of minimal replication it is possible to construct a hybrid test procedure that combines a Bayesian prior on ρ with under-replicated data to find a posterior mean ($\vec{\rho}$) to use as the plug-in value ρ_0 in one of the strategies described in this paper.

Essentially, $\vec{\rho}$ will be a weighted average of prior information and current data. If the prior information is considered to be reliable, more weight will be given to the prior information. If the prior information is considered less reliable, more weight should be given to the current

data. Because in an under-replication situation very little data exists to incorporate into $\ddot{\rho}$, most of the weight will be given to the prior information.

This paper has already explored various prior distributions on ρ. One such prior is the Beta(5,30) distribution. This distribution has a mean of 0.14286. Suppose a set of under-replication data yielded an estimated ICC of 0.20. If in prior studies of a similar nature the ICC has been found to follow approximately a Beta(5,30) distribution, it would not be unreasonable to use $\ddot{\rho}$ with a value between 0.14286 and 0.20. That value could then be used in the ADFMR strategy. The probability of rejecting the null hypothesis using such a method would not differ considerably using $\ddot{\rho}$ from which it would be using $\overline{\rho}$.

6.3: INTRODUCING THE PLUG-IN METHOD TO THE LITERATURE

The utility of the methods in this study is dependent upon their introduction to practicioners. If deemed an acceptable method of analysis, the plug-in method may be commonly used as a solution to the inability to conduct hypothesis tests on treatments means in situations in which replication is prohibitive. I intend to present methods at conferences and submit articles to subject area journals to introduce the methodology to researchers in various areas of application.

6.4: SUMMARY

In this study, we investigated the use of plug-in values of the intraclass correlation coefficient and weighted p-values in unreplicated and under-replicated experiments. We defined a test in the two-treatment no replication case and found it to be a feasible method of analysis. We proposed strategies for implementation in this case. We looked at properties of each of the strategies and provided an example of implementation. We investigated the t-treatment no replication case for which we also looked at properties and provided an example of

implementation. We explained and demonstrated the implementation of several multiple comparison procedures in the context of the plug-in method. We then extended the plug-in method and strategies to the general under-replication t-treatment case.

This study provides a method of data analysis in situations that otherwise prohibit analysis by traditional methods. These methods result in hypothesis tests that can maintain significance levels near or below nominal levels with power to detect differences in treatment means. The strategies presented in this study can easily be implemented by practicioners. Examples of implementation have been given for each of the strategies using the widely available SAS® software. A complete example is given of the analysis of a set of real-world data using each of the strategies presented in this study.

6.5 IMPLICATIONS AND DRAWBACKS

The methodology described in this study allows researchers to conduct tests of hypothesis with no replication and with very minimal replication and still obtain statistical results. In the examples and explanations given for lack of abundant replication it is suggested that the plug-in methods be used when replication is prohibitive. The plug-in methodology might also be used by researchers in an effort to save money and resources, where those resources are available, but, the researcher could benefit from using less. Suppose a company conducts quality control tests on a regular basis that include subsampling and costly replication. Over time, the company may have a very accurate and consistent estimate of the ICC for this process. The company may choose to use this ICC estimate to enable future quality control tests without replication. This could save the company considerable time and resources.

The ICC must be <0.5 in order for the methods described in this paper to work. An ICC value <0.05 indicates a situation in which the between-class variability is less than the within-class variability. As $\rho \to 1,\ w = \frac{\rho}{1-\rho} + \frac{1}{n} \to \infty$. This leads to small test statistics and large

p-values. Tests based on ICC values >0.5 will have little ability to detect differences in treatment means. This restriction, ICC<0.5, excludes this methodology from being used in many situations in which ICC>0.5.

APPENDICES

APPENDIX A: TABLES

Table 1: Probability of Rejection (α=0.05, no replication, t=2, b=1)

The following table contains the two-tail probability of rejecting the hypothesis H_0: $\mu_1 = \mu_2$ vs. H_A: $\mu_1 \neq \mu_2$ (i.e.- t=2 treatments) where b=1 experimental unit per treatment and significance level α=0.05. Both alpha (the probability of rejecting the null hypothesis when it is in fact correct) and power (the probability of rejecting the null hypothesis when it is in fact incorrect) can be obtained from this table.

The table includes the following variables:

n = the number of *observational units* per experimental unit

Rho = the value of the actual ICC (ρ)

Rho* = the value the researcher uses for the ICC (which is not generally the same as the value of the actual ICC, because it must be chosen from prior information)

St.Diff. = the standardized difference to be tested, $\text{St.Diff.} = \frac{\mu_1 - \mu_2}{\sigma_\varepsilon}$

Probt = the probability of rejecting the null hypothesis in favor of the two-tail alternative, given the values of the other variables.

St.Diff = 0

		Rho*										
n	**Rho**	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0.00	0.0500	0.0183	0.0069	0.0026	0.0010	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
	0.05	0.1062	0.0500	0.0237	0.0112	0.0052	0.0023	0.0010	0.0004	0.0002	0.0001	0.0000
	0.10	0.1654	0.0909	0.0500	0.0272	0.0145	0.0075	0.0038	0.0018	0.0008	0.0004	0.0001
	0.15	0.2225	0.1359	0.0830	0.0500	0.0295	0.0170	0.0094	0.0050	0.0026	0.0012	0.0005
	0.20	0.2762	0.1823	0.1201	0.0782	0.0500	0.0312	0.0188	0.0109	0.0061	0.0032	0.0016
	0.25	0.3262	0.2284	0.1598	0.1106	0.0752	0.0500	0.0323	0.0201	0.0120	0.0068	0.0036
	0.30	0.3729	0.2737	0.2009	0.1460	0.1045	0.0733	0.0500	0.0330	0.0209	0.0126	0.0072
	0.35	0.4166	0.3179	0.2426	0.1837	0.1372	0.1006	0.0720	0.0500	0.0335	0.0214	0.0130
	0.40	0.4578	0.3610	0.2848	0.2232	0.1728	0.1316	0.0981	0.0712	0.0500	0.0337	0.0216
	0.45	0.4969	0.4029	0.3271	0.2641	0.2110	0.1661	0.1283	0.0968	0.0709	0.0500	0.0337
	0.50	0.5344	0.4440	0.3695	0.3062	0.2514	0.2038	0.1625	0.1269	0.0964	0.0709	0.0500
30	0.00	0.0500	0.0021	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.2176	0.0500	0.0120	0.0028	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.3402	0.1280	0.0500	0.0190	0.0069	0.0023	0.0007	0.0002	0.0000	0.0000	0.0000
	0.15	0.4282	0.2052	0.1021	0.0500	0.0235	0.0105	0.0043	0.0016	0.0006	0.0002	0.0000
	0.20	0.4951	0.2748	0.1583	0.0903	0.0500	0.0265	0.0132	0.0062	0.0026	0.0010	0.0003
	0.25	0.5485	0.3365	0.2140	0.1354	0.0837	0.0500	0.0285	0.0153	0.0076	0.0034	0.0014
	0.30	0.5928	0.3914	0.2676	0.1826	0.1224	0.0797	0.0500	0.0298	0.0167	0.0087	0.0040
	0.35	0.6307	0.4408	0.3186	0.2302	0.1642	0.1144	0.0772	0.0500	0.0307	0.0176	0.0093
	0.40	0.6639	0.4858	0.3670	0.2777	0.2079	0.1528	0.1094	0.0756	0.0500	0.0312	0.0181
	0.45	0.6935	0.5273	0.4131	0.3245	0.2530	0.1942	0.1458	0.1064	0.0747	0.0500	0.0314
	0.50	0.7205	0.5659	0.4572	0.3708	0.2989	0.2380	0.1860	0.1419	0.1049	0.0744	0.0500
50	0.00	0.0500	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.3003	0.0500	0.0090	0.0015	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.4402	0.1429	0.0500	0.0170	0.0054	0.0015	0.0004	0.0001	0.0000	0.0000	0.0000
	0.15	0.5281	0.2305	0.1082	0.0500	0.0221	0.0091	0.0034	0.0012	0.0003	0.0001	0.0000
	0.20	0.5904	0.3059	0.1698	0.0937	0.0500	0.0254	0.0121	0.0053	0.0021	0.0007	0.0002
	0.25	0.6379	0.3705	0.2296	0.1421	0.0859	0.0500	0.0276	0.0143	0.0068	0.0029	0.0011
	0.30	0.6761	0.4265	0.2860	0.1921	0.1269	0.0813	0.0500	0.0291	0.0158	0.0079	0.0035
	0.35	0.7081	0.4759	0.3386	0.2420	0.1708	0.1177	0.0784	0.0500	0.0301	0.0169	0.0086
	0.40	0.7356	0.5202	0.3880	0.2911	0.2163	0.1578	0.1119	0.0766	0.0500	0.0307	0.0174
	0.45	0.7598	0.5604	0.4344	0.3390	0.2628	0.2006	0.1498	0.1085	0.0755	0.0500	0.0310
	0.50	0.7817	0.5976	0.4785	0.3859	0.3098	0.2456	0.1912	0.1452	0.1067	0.0751	0.0500
100	0.00	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.4317	0.0500	0.0067	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	0.5716	0.1577	0.0500	0.0153	0.0043	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000
	0.15	0.6484	0.2545	0.1136	0.0500	0.0209	0.0081	0.0028	0.0009	0.0002	0.0000	0.0000
	0.20	0.6994	0.3343	0.1799	0.0965	0.0500	0.0245	0.0112	0.0046	0.0017	0.0005	0.0001
	0.25	0.7368	0.4007	0.2429	0.1477	0.0877	0.0500	0.0270	0.0135	0.0062	0.0025	0.0009
	0.30	0.7662	0.4570	0.3013	0.2000	0.1305	0.0826	0.0500	0.0286	0.0152	0.0073	0.0031
	0.35	0.7903	0.5059	0.3552	0.2516	0.1761	0.1203	0.0794	0.0500	0.0297	0.0163	0.0081
	0.40	0.8108	0.5492	0.4051	0.3018	0.2230	0.1617	0.1140	0.0774	0.0500	0.0303	0.0169
	0.45	0.8287	0.5882	0.4517	0.3505	0.2705	0.2057	0.1528	0.1101	0.0762	0.0500	0.0306
	0.50	0.8446	0.6239	0.4955	0.3978	0.3183	0.2516	0.1953	0.1478	0.1081	0.0757	0.0500
∞	0.00	0.0500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	1.0000	0.0500	0.0044	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	1.0000	0.1774	0.0500	0.0135	0.0033	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	0.15	1.0000	0.2845	0.1199	0.0500	0.0197	0.0071	0.0023	0.0006	0.0001	0.0000	0.0000
	0.20	1.0000	0.3685	0.1913	0.0996	0.0500	0.0236	0.0103	0.0040	0.0014	0.0004	0.0001
	0.25	1.0000	0.4361	0.2578	0.1538	0.0896	0.0500	0.0263	0.0127	0.0056	0.0021	0.0007
	0.30	1.0000	0.4922	0.3183	0.2085	0.1344	0.0839	0.0500	0.0280	0.0145	0.0068	0.0028
	0.35	1.0000	0.5400	0.3733	0.2618	0.1817	0.1231	0.0804	0.0500	0.0292	0.0157	0.0076
	0.40	1.0000	0.5818	0.4236	0.3133	0.2301	0.1658	0.1161	0.0782	0.0500	0.0299	0.0164
	0.45	1.0000	0.6191	0.4701	0.3627	0.2786	0.2109	0.1560	0.1118	0.0769	0.0500	0.0302
	0.50	1.0000	0.6530	0.5135	0.4103	0.3271	0.2578	0.1995	0.1504	0.1095	0.0763	0.0500

St.Diff = 0.5

		Rho*										
n	**Rho**	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0.00	0.1851	0.0927	0.0452	0.0214	0.0098	0.0044	0.0019	0.0008	0.0003	0.0001	0.0000
	0.05	0.2352	0.1375	0.0788	0.0440	0.0238	0.0124	0.0062	0.0030	0.0013	0.0006	0.0002
	0.10	0.2780	0.1782	0.1128	0.0698	0.0420	0.0245	0.0137	0.0073	0.0037	0.0018	0.0008
	0.15	0.3174	0.2167	0.1467	0.0977	0.0635	0.0400	0.0244	0.0142	0.0079	0.0041	0.0020
	0.20	0.3549	0.2540	0.1809	0.1271	0.0875	0.0587	0.0381	0.0238	0.0142	0.0080	0.0042
	0.25	0.3910	0.2906	0.2154	0.1579	0.1138	0.0802	0.0549	0.0363	0.0230	0.0138	0.0078
	0.30	0.4259	0.3267	0.2504	0.1902	0.1423	0.1044	0.0747	0.0518	0.0345	0.0220	0.0132
	0.35	0.4597	0.3625	0.2859	0.2238	0.1729	0.1313	0.0975	0.0704	0.0491	0.0328	0.0207
	0.40	0.4927	0.3981	0.3219	0.2587	0.2056	0.1609	0.1234	0.0923	0.0669	0.0467	0.0310
	0.45	0.5251	0.4335	0.3584	0.2948	0.2402	0.1931	0.1525	0.1178	0.0884	0.0641	0.0445
	0.50	0.5568	0.4689	0.3956	0.3324	0.2770	0.2281	0.1849	0.1470	0.1139	0.0856	0.0618
30	0.00	0.4779	0.1128	0.0194	0.0027	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.4936	0.2203	0.0903	0.0334	0.0111	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	0.10	0.5192	0.2811	0.1500	0.0760	0.0359	0.0156	0.0062	0.0022	0.0007	0.0002	0.0000
	0.15	0.5502	0.3294	0.2000	0.1182	0.0668	0.0356	0.0177	0.0080	0.0033	0.0012	0.0004
	0.20	0.5816	0.3729	0.2452	0.1591	0.1001	0.0604	0.0345	0.0184	0.0090	0.0040	0.0015
	0.25	0.6116	0.4135	0.2880	0.1993	0.1350	0.0886	0.0556	0.0331	0.0183	0.0093	0.0042
	0.30	0.6398	0.4519	0.3292	0.2393	0.1712	0.1194	0.0805	0.0519	0.0316	0.0179	0.0092
	0.35	0.6662	0.4884	0.3691	0.2791	0.2086	0.1528	0.1087	0.0746	0.0487	0.0300	0.0170
	0.40	0.6909	0.5232	0.4081	0.3190	0.2471	0.1883	0.1401	0.1011	0.0700	0.0460	0.0283
	0.45	0.7141	0.5565	0.4462	0.3589	0.2867	0.2259	0.1745	0.1313	0.0955	0.0664	0.0436
	0.50	0.7362	0.5887	0.4837	0.3989	0.3273	0.2656	0.2119	0.1654	0.1253	0.0914	0.0635
50	0.00	0.6969	0.1097	0.0078	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.6169	0.2549	0.0928	0.0292	0.0078	0.0018	0.0003	0.0000	0.0000	0.0000	0.0000
	0.10	0.6208	0.3180	0.1620	0.0774	0.0339	0.0134	0.0047	0.0014	0.0004	0.0001	0.0000
	0.15	0.6426	0.3661	0.2161	0.1240	0.0676	0.0344	0.0161	0.0068	0.0025	0.0008	0.0002
	0.20	0.6679	0.4093	0.2635	0.1678	0.1034	0.0608	0.0336	0.0171	0.0079	0.0032	0.0011
	0.25	0.6927	0.4495	0.3078	0.2101	0.1404	0.0906	0.0558	0.0323	0.0173	0.0084	0.0036
	0.30	0.7160	0.4873	0.3499	0.2517	0.1783	0.1231	0.0818	0.0519	0.0309	0.0170	0.0084
	0.35	0.7376	0.5229	0.3904	0.2928	0.2172	0.1578	0.1113	0.0755	0.0487	0.0294	0.0163
	0.40	0.7577	0.5566	0.4296	0.3335	0.2569	0.1946	0.1439	0.1030	0.0707	0.0459	0.0277
	0.45	0.7766	0.5887	0.4676	0.3740	0.2974	0.2334	0.1795	0.1343	0.0970	0.0669	0.0434
	0.50	0.7943	0.6194	0.5047	0.4143	0.3387	0.2739	0.2179	0.1694	0.1278	0.0927	0.0639
100	0.00	0.9404	0.0881	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.7481	0.2901	0.0945	0.0252	0.0054	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000
	0.10	0.7305	0.3524	0.1728	0.0785	0.0322	0.0116	0.0036	0.0009	0.0002	0.0000	0.0000
	0.15	0.7428	0.3988	0.2300	0.1288	0.0681	0.0334	0.0148	0.0058	0.0020	0.0006	0.0001
	0.20	0.7608	0.4411	0.2791	0.1750	0.1061	0.0611	0.0328	0.0162	0.0071	0.0027	0.0009
	0.25	0.7791	0.4806	0.3243	0.2190	0.1447	0.0922	0.0559	0.0317	0.0166	0.0078	0.0032
	0.30	0.7964	0.5175	0.3670	0.2618	0.1841	0.1259	0.0829	0.0519	0.0304	0.0163	0.0078
	0.35	0.8124	0.5521	0.4078	0.3037	0.2240	0.1618	0.1134	0.0762	0.0486	0.0289	0.0157
	0.40	0.8272	0.5846	0.4469	0.3451	0.2646	0.1996	0.1469	0.1045	0.0712	0.0458	0.0273
	0.45	0.8409	0.6154	0.4847	0.3859	0.3058	0.2392	0.1833	0.1367	0.0982	0.0673	0.0433
	0.50	0.8538	0.6448	0.5214	0.4263	0.3475	0.2804	0.2224	0.1725	0.1297	0.0937	0.0642
∞	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	0.3379	0.0957	0.0203	0.0032	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	0.10	1.0000	0.3946	0.1855	0.0796	0.0301	0.0098	0.0026	0.0006	0.0001	0.0000	0.0000
	0.15	1.0000	0.4375	0.2460	0.1343	0.0687	0.0322	0.0135	0.0049	0.0015	0.0004	0.0001
	0.20	1.0000	0.4779	0.2964	0.1830	0.1090	0.0613	0.0320	0.0152	0.0064	0.0023	0.0007
	0.25	1.0000	0.5160	0.3424	0.2287	0.1494	0.0939	0.0560	0.0311	0.0158	0.0071	0.0027
	0.30	1.0000	0.5516	0.3855	0.2726	0.1901	0.1290	0.0840	0.0519	0.0298	0.0157	0.0073
	0.35	1.0000	0.5848	0.4265	0.3154	0.2312	0.1660	0.1155	0.0770	0.0485	0.0284	0.0151
	0.40	1.0000	0.6158	0.4655	0.3573	0.2727	0.2048	0.1500	0.1061	0.0717	0.0457	0.0269
	0.45	1.0000	0.6450	0.5030	0.3984	0.3145	0.2452	0.1873	0.1390	0.0994	0.0677	0.0431
	0.50	1.0000	0.6727	0.5392	0.4389	0.3567	0.2870	0.2272	0.1756	0.1316	0.0946	0.0644

St.Diff = 1

		Rho*										
n	**Rho**	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0.00	0.5620	0.3837	0.2451	0.1475	0.0839	0.0452	0.0230	0.0110	0.0050	0.0021	0.0008
	0.05	0.5515	0.4027	0.2813	0.1878	0.1198	0.0729	0.0421	0.0230	0.0118	0.0056	0.0025
	0.10	0.5458	0.4164	0.3079	0.2200	0.1514	0.1001	0.0632	0.0379	0.0215	0.0114	0.0056
	0.15	0.5442	0.4280	0.3294	0.2469	0.1795	0.1260	0.0851	0.0549	0.0336	0.0193	0.0103
	0.20	0.5462	0.4394	0.3482	0.2704	0.2049	0.1507	0.1071	0.0732	0.0476	0.0293	0.0168
	0.25	0.5516	0.4516	0.3661	0.2922	0.2286	0.1745	0.1292	0.0924	0.0633	0.0412	0.0252
	0.30	0.5601	0.4652	0.3840	0.3134	0.2515	0.1977	0.1515	0.1125	0.0805	0.0550	0.0355
	0.35	0.5713	0.4804	0.4028	0.3347	0.2744	0.2211	0.1742	0.1336	0.0991	0.0705	0.0477
	0.40	0.5848	0.4975	0.4228	0.3569	0.2979	0.2451	0.1978	0.1559	0.1193	0.0880	0.0620
	0.45	0.6004	0.5165	0.4443	0.3803	0.3226	0.2703	0.2227	0.1798	0.1414	0.1076	0.0786
	0.50	0.6178	0.5372	0.4677	0.4055	0.3490	0.2972	0.2495	0.2057	0.1658	0.1298	0.0979
30	0.00	0.9677	0.7404	0.3987	0.1533	0.0438	0.0096	0.0016	0.0002	0.0000	0.0000	0.0000
	0.05	0.8777	0.6597	0.4337	0.2501	0.1264	0.0556	0.0211	0.0068	0.0018	0.0004	0.0001
	0.10	0.8183	0.6258	0.4483	0.2986	0.1838	0.1036	0.0529	0.0240	0.0096	0.0032	0.0009
	0.15	0.7824	0.6075	0.4575	0.3296	0.2252	0.1447	0.0864	0.0474	0.0234	0.0102	0.0038
	0.20	0.7624	0.5986	0.4658	0.3526	0.2573	0.1793	0.1183	0.0730	0.0416	0.0214	0.0097
	0.25	0.7530	0.5968	0.4752	0.3722	0.2841	0.2095	0.1480	0.0992	0.0622	0.0360	0.0187
	0.30	0.7508	0.6005	0.4866	0.3909	0.3082	0.2368	0.1759	0.1252	0.0844	0.0532	0.0307
	0.35	0.7534	0.6085	0.5004	0.4100	0.3316	0.2628	0.2027	0.1511	0.1077	0.0726	0.0455
	0.40	0.7595	0.6199	0.5166	0.4304	0.3552	0.2886	0.2295	0.1773	0.1321	0.0939	0.0627
	0.45	0.7680	0.6340	0.5351	0.4523	0.3799	0.3152	0.2569	0.2045	0.1579	0.1172	0.0825
	0.50	0.7783	0.6503	0.5556	0.4762	0.4062	0.3432	0.2857	0.2333	0.1856	0.1428	0.1051
50	0.00	0.9986	0.8821	0.4747	0.1355	0.0217	0.0021	0.0001	0.0000	0.0000	0.0000	0.0000
	0.05	0.9432	0.7383	0.4860	0.2692	0.1249	0.0481	0.0151	0.0038	0.0007	0.0001	0.0000
	0.10	0.8838	0.6836	0.4896	0.3216	0.1922	0.1033	0.0492	0.0203	0.0071	0.0020	0.0004
	0.15	0.8452	0.6545	0.4921	0.3521	0.2373	0.1490	0.0862	0.0451	0.0209	0.0083	0.0027
	0.20	0.8232	0.6392	0.4958	0.3737	0.2706	0.1864	0.1208	0.0727	0.0399	0.0195	0.0082
	0.25	0.8122	0.6333	0.5021	0.3919	0.2976	0.2180	0.1524	0.1006	0.0618	0.0346	0.0173
	0.30	0.8085	0.6343	0.5113	0.4093	0.3217	0.2460	0.1815	0.1280	0.0852	0.0527	0.0296
	0.35	0.8095	0.6404	0.5235	0.4275	0.3447	0.2723	0.2092	0.1550	0.1096	0.0730	0.0449
	0.40	0.8137	0.6503	0.5386	0.4471	0.3681	0.2984	0.2365	0.1821	0.1349	0.0951	0.0628
	0.45	0.8201	0.6630	0.5561	0.4685	0.3926	0.3250	0.2644	0.2099	0.1615	0.1192	0.0834
	0.50	0.8281	0.6780	0.5758	0.4919	0.4187	0.3531	0.2935	0.2392	0.1898	0.1456	0.1066
100	0.00	1.0000	0.9812	0.5813	0.0936	0.0039	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	0.05	0.9793	0.8028	0.5342	0.2861	0.1222	0.0409	0.0105	0.0020	0.0003	0.0000	0.0000
	0.10	0.9333	0.7307	0.5248	0.3413	0.1990	0.1026	0.0459	0.0173	0.0053	0.0013	0.0002
	0.15	0.8994	0.6925	0.5206	0.3707	0.2472	0.1525	0.0858	0.0431	0.0189	0.0069	0.0020
	0.20	0.8796	0.6720	0.5201	0.3908	0.2814	0.1921	0.1227	0.0723	0.0385	0.0180	0.0072
	0.25	0.8694	0.6629	0.5236	0.4076	0.3085	0.2247	0.1559	0.1017	0.0614	0.0336	0.0162
	0.30	0.8656	0.6618	0.5309	0.4239	0.3323	0.2533	0.1860	0.1303	0.0858	0.0523	0.0288
	0.35	0.8656	0.6664	0.5418	0.4412	0.3550	0.2798	0.2143	0.1581	0.1111	0.0733	0.0445
	0.40	0.8683	0.6751	0.5559	0.4602	0.3781	0.3059	0.2420	0.1858	0.1371	0.0961	0.0629
	0.45	0.8727	0.6867	0.5726	0.4811	0.4024	0.3327	0.2701	0.2141	0.1643	0.1208	0.0840
	0.50	0.8783	0.7007	0.5917	0.5041	0.4284	0.3608	0.2995	0.2437	0.1931	0.1477	0.1078
∞	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	0.8691	0.5927	0.3062	0.1171	0.0321	0.0060	0.0007	0.0000	0.0000	0.0000
	0.10	1.0000	0.7803	0.5641	0.3636	0.2065	0.1014	0.0420	0.0141	0.0037	0.0007	0.0001
	0.15	1.0000	0.7330	0.5515	0.3911	0.2580	0.1561	0.0852	0.0409	0.0167	0.0056	0.0014
	0.20	1.0000	0.7070	0.5461	0.4092	0.2930	0.1981	0.1247	0.0718	0.0370	0.0165	0.0061
	0.25	1.0000	0.6947	0.5463	0.4242	0.3200	0.2318	0.1595	0.1028	0.0610	0.0325	0.0150
	0.30	1.0000	0.6916	0.5516	0.4392	0.3434	0.2608	0.1907	0.1326	0.0864	0.0518	0.0278
	0.35	1.0000	0.6946	0.5611	0.4556	0.3657	0.2876	0.2196	0.1613	0.1126	0.0736	0.0440
	0.40	1.0000	0.7020	0.5741	0.4738	0.3885	0.3137	0.2477	0.1896	0.1393	0.0971	0.0630
	0.45	1.0000	0.7125	0.5900	0.4942	0.4125	0.3405	0.2760	0.2183	0.1671	0.1224	0.0846
	0.50	1.0000	0.7253	0.6083	0.5167	0.4383	0.3686	0.3056	0.2483	0.1964	0.1498	0.1090

St.Diff = 1.5

		Rho*										
n	**Rho**	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0.00	0.8870	0.7675	0.6214	0.4700	0.3322	0.2193	0.1351	0.0775	0.0413	0.0202	0.0090
	0.05	0.8411	0.7284	0.6021	0.4743	0.3551	0.2522	0.1692	0.1067	0.0629	0.0344	0.0173
	0.10	0.8048	0.7002	0.5888	0.4774	0.3721	0.2779	0.1979	0.1337	0.0851	0.0506	0.0277
	0.15	0.7755	0.6786	0.5788	0.4798	0.3855	0.2989	0.2227	0.1586	0.1071	0.0680	0.0401
	0.20	0.7520	0.6616	0.5711	0.4820	0.3965	0.3167	0.2445	0.1815	0.1285	0.0861	0.0540
	0.25	0.7333	0.6484	0.5653	0.4842	0.4061	0.3323	0.2641	0.2027	0.1494	0.1048	0.0691
	0.30	0.7191	0.6387	0.5615	0.4869	0.4149	0.3463	0.2819	0.2227	0.1697	0.1237	0.0854
	0.35	0.7091	0.6322	0.5599	0.4905	0.4236	0.3595	0.2987	0.2418	0.1895	0.1429	0.1025
	0.40	0.7030	0.6290	0.5605	0.4953	0.4328	0.3726	0.3150	0.2603	0.2092	0.1624	0.1206
	0.45	0.7005	0.6291	0.5637	0.5020	0.4430	0.3861	0.3313	0.2788	0.2290	0.1824	0.1397
	0.50	0.7015	0.6324	0.5697	0.5110	0.4549	0.4008	0.3485	0.2980	0.2494	0.2032	0.1600
30	0.00	0.9999	0.9938	0.9392	0.7686	0.5002	0.2479	0.0922	0.0256	0.0053	0.0008	0.0001
	0.05	0.9909	0.9449	0.8426	0.6859	0.4999	0.3209	0.1786	0.0846	0.0334	0.0107	0.0027
	0.10	0.9662	0.8926	0.7836	0.6478	0.4998	0.3562	0.2315	0.1352	0.0695	0.0307	0.0113
	0.15	0.9363	0.8493	0.7436	0.6245	0.4998	0.3783	0.2681	0.1757	0.1047	0.0556	0.0255
	0.20	0.9081	0.8144	0.7142	0.6081	0.4999	0.3941	0.2956	0.2087	0.1368	0.0818	0.0436
	0.25	0.8841	0.7867	0.6919	0.5961	0.5001	0.4063	0.3174	0.2363	0.1658	0.1080	0.0639
	0.30	0.8652	0.7655	0.6753	0.5874	0.5009	0.4165	0.3356	0.2601	0.1921	0.1333	0.0855
	0.35	0.8513	0.7501	0.6635	0.5817	0.5026	0.4257	0.3515	0.2812	0.2161	0.1577	0.1077
	0.40	0.8418	0.7399	0.6562	0.5792	0.5058	0.4349	0.3663	0.3006	0.2385	0.1813	0.1303
	0.45	0.8362	0.7344	0.6531	0.5799	0.5110	0.4449	0.3809	0.3192	0.2601	0.2044	0.1532
	0.50	0.8340	0.7331	0.6542	0.5841	0.5188	0.4565	0.3963	0.3379	0.2815	0.2275	0.1766
50	0.00	1.0000	0.9998	0.9888	0.8819	0.5803	0.2408	0.0594	0.0086	0.0007	0.0000	0.0000
	0.05	0.9981	0.9736	0.8950	0.7462	0.5459	0.3408	0.1768	0.0740	0.0242	0.0059	0.0010
	0.10	0.9844	0.9262	0.8265	0.6910	0.5347	0.3776	0.2395	0.1336	0.0639	0.0253	0.0079
	0.15	0.9620	0.8822	0.7796	0.6589	0.5285	0.3986	0.2794	0.1792	0.1031	0.0518	0.0219
	0.20	0.9383	0.8453	0.7453	0.6370	0.5245	0.4129	0.3079	0.2150	0.1383	0.0802	0.0408
	0.25	0.9175	0.8157	0.7193	0.6210	0.5216	0.4235	0.3299	0.2441	0.1694	0.1083	0.0624
	0.30	0.9007	0.7928	0.6998	0.6094	0.5200	0.4323	0.3478	0.2686	0.1970	0.1353	0.0853
	0.35	0.8883	0.7760	0.6858	0.6014	0.5198	0.4402	0.3632	0.2900	0.2220	0.1609	0.1087
	0.40	0.8798	0.7648	0.6768	0.5971	0.5214	0.4483	0.3774	0.3094	0.2449	0.1854	0.1323
	0.45	0.8747	0.7584	0.6725	0.5965	0.5254	0.4573	0.3914	0.3278	0.2667	0.2091	0.1561
	0.50	0.8724	0.7564	0.6725	0.5996	0.5322	0.4681	0.4063	0.3463	0.2883	0.2326	0.1801
100	0.00	1.0000	1.0000	0.9998	0.9733	0.6929	0.2098	0.0206	0.0006	0.0000	0.0000	0.0000
	0.05	0.9997	0.9880	0.9313	0.7961	0.5874	0.3582	0.1732	0.0635	0.0167	0.0030	0.0003
	0.10	0.9936	0.9482	0.8584	0.7256	0.5637	0.3956	0.2457	0.1316	0.0587	0.0209	0.0056
	0.15	0.9790	0.9055	0.8068	0.6859	0.5516	0.4152	0.2885	0.1818	0.1015	0.0486	0.0191
	0.20	0.9617	0.8681	0.7689	0.6595	0.5438	0.4278	0.3178	0.2199	0.1393	0.0788	0.0385
	0.25	0.9457	0.8375	0.7402	0.6403	0.5384	0.4371	0.3397	0.2502	0.1722	0.1086	0.0611
	0.30	0.9327	0.8137	0.7185	0.6263	0.5348	0.4446	0.3573	0.2753	0.2009	0.1368	0.0851
	0.35	0.9230	0.7962	0.7029	0.6165	0.5330	0.4514	0.3723	0.2968	0.2265	0.1634	0.1095
	0.40	0.9163	0.7844	0.6927	0.6109	0.5334	0.4585	0.3860	0.3161	0.2499	0.1886	0.1339
	0.45	0.9122	0.7775	0.6874	0.6092	0.5364	0.4668	0.3995	0.3344	0.2718	0.2127	0.1583
	0.50	0.9103	0.7750	0.6867	0.6115	0.5424	0.4769	0.4138	0.3527	0.2934	0.2365	0.1828
∞	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	0.9961	0.9621	0.8495	0.6375	0.3786	0.1661	0.0499	0.0093	0.0010	0.0000
	0.10	1.0000	0.9666	0.8891	0.7617	0.5956	0.4158	0.2523	0.1287	0.0527	0.0163	0.0035
	0.15	1.0000	0.9273	0.8339	0.7139	0.5762	0.4330	0.2982	0.1844	0.0995	0.0450	0.0161
	0.20	1.0000	0.8904	0.7927	0.6826	0.5641	0.4436	0.3282	0.2251	0.1402	0.0772	0.0360
	0.25	1.0000	0.8595	0.7613	0.6600	0.5558	0.4512	0.3500	0.2566	0.1750	0.1087	0.0597
	0.30	1.0000	0.8352	0.7375	0.6435	0.5500	0.4573	0.3672	0.2821	0.2049	0.1383	0.0848
	0.35	1.0000	0.8173	0.7203	0.6319	0.5465	0.4629	0.3816	0.3038	0.2312	0.1659	0.1102
	0.40	1.0000	0.8050	0.7090	0.6249	0.5456	0.4690	0.3947	0.3230	0.2549	0.1918	0.1355
	0.45	1.0000	0.7978	0.7028	0.6221	0.5476	0.4764	0.4077	0.3411	0.2770	0.2164	0.1605
	0.50	1.0000	0.7949	0.7014	0.6236	0.5528	0.4859	0.4215	0.3591	0.2986	0.2405	0.1855

St.Diff = 2

		Rho*										
n	**Rho**	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0.00	0.9882	0.9602	0.9037	0.8132	0.6925	0.5535	0.4126	0.2851	0.1816	0.1057	0.0557
	0.05	0.9692	0.9280	0.8632	0.7745	0.6661	0.5457	0.4232	0.3088	0.2104	0.1326	0.0764
	0.10	0.9459	0.8965	0.8290	0.7446	0.6465	0.5400	0.4315	0.3277	0.2347	0.1571	0.0971
	0.15	0.9212	0.8673	0.7999	0.7204	0.6312	0.5357	0.4381	0.3433	0.2558	0.1796	0.1174
	0.20	0.8967	0.8406	0.7748	0.7002	0.6186	0.5321	0.4437	0.3567	0.2745	0.2004	0.1372
	0.25	0.8732	0.8163	0.7528	0.6830	0.6080	0.5292	0.4485	0.3683	0.2912	0.2197	0.1566
	0.30	0.8515	0.7946	0.7335	0.6681	0.5989	0.5267	0.4528	0.3787	0.3064	0.2379	0.1755
	0.35	0.8318	0.7753	0.7166	0.6552	0.5912	0.5248	0.4568	0.3882	0.3204	0.2550	0.1940
	0.40	0.8146	0.7586	0.7022	0.6444	0.5848	0.5234	0.4606	0.3971	0.3336	0.2714	0.2121
	0.45	0.8003	0.7448	0.6903	0.6356	0.5798	0.5229	0.4647	0.4057	0.3463	0.2873	0.2300
	0.50	0.7890	0.7340	0.6813	0.6292	0.5768	0.5236	0.4695	0.4145	0.3588	0.3030	0.2479
30	0.00	1.0000	1.0000	0.9996	0.9935	0.9553	0.8333	0.6070	0.3443	0.1451	0.0438	0.0092
	0.05	0.9998	0.9973	0.9853	0.9497	0.8732	0.7456	0.5745	0.3879	0.2227	0.1051	0.0392
	0.10	0.9970	0.9847	0.9553	0.9010	0.8159	0.6997	0.5594	0.4095	0.2687	0.1539	0.0744
	0.15	0.9890	0.9640	0.9218	0.8590	0.7746	0.6700	0.5501	0.4232	0.3003	0.1924	0.1083
	0.20	0.9760	0.9396	0.8894	0.8238	0.7430	0.6485	0.5435	0.4331	0.3240	0.2237	0.1393
	0.25	0.9600	0.9144	0.8595	0.7939	0.7177	0.6319	0.5385	0.4407	0.3427	0.2498	0.1673
	0.30	0.9431	0.8899	0.8325	0.7682	0.6967	0.6184	0.5344	0.4468	0.3582	0.2722	0.1929
	0.35	0.9268	0.8673	0.8085	0.7461	0.6790	0.6072	0.5312	0.4520	0.3715	0.2918	0.2162
	0.40	0.9120	0.8474	0.7878	0.7273	0.6642	0.5980	0.5286	0.4567	0.3831	0.3095	0.2379
	0.45	0.8994	0.8306	0.7705	0.7117	0.6521	0.5907	0.5270	0.4612	0.3938	0.3256	0.2581
	0.50	0.8893	0.8173	0.7569	0.6997	0.6430	0.5855	0.5267	0.4661	0.4041	0.3409	0.2774
50	0.00	1.0000	1.0000	1.0000	0.9997	0.9921	0.9254	0.6989	0.3572	0.1090	0.0184	0.0016
	0.05	1.0000	0.9994	0.9945	0.9738	0.9163	0.8000	0.6218	0.4114	0.2210	0.0915	0.0274
	0.10	0.9991	0.9922	0.9717	0.9280	0.8517	0.7387	0.5935	0.4317	0.2764	0.1504	0.0664
	0.15	0.9948	0.9761	0.9410	0.8848	0.8047	0.7008	0.5774	0.4433	0.3111	0.1944	0.1046
	0.20	0.9860	0.9546	0.9092	0.8475	0.7688	0.6742	0.5666	0.4513	0.3357	0.2287	0.1389
	0.25	0.9739	0.9309	0.8790	0.8155	0.7402	0.6539	0.5585	0.4572	0.3546	0.2565	0.1694
	0.30	0.9605	0.9072	0.8512	0.7879	0.7166	0.6377	0.5521	0.4618	0.3698	0.2798	0.1965
	0.35	0.9471	0.8850	0.8263	0.7641	0.6968	0.6243	0.5469	0.4658	0.3826	0.2999	0.2210
	0.40	0.9349	0.8653	0.8047	0.7438	0.6802	0.6132	0.5428	0.4693	0.3937	0.3176	0.2434
	0.45	0.9244	0.8485	0.7866	0.7270	0.6666	0.6044	0.5398	0.4727	0.4037	0.3337	0.2641
	0.50	0.9160	0.8352	0.7724	0.7139	0.6563	0.5980	0.5382	0.4767	0.4134	0.3487	0.2836
100	0.00	1.0000	1.0000	1.0000	1.0000	0.9999	0.9876	0.8210	0.3608	0.0548	0.0024	0.0000
	0.05	1.0000	0.9999	0.9980	0.9868	0.9458	0.8444	0.6645	0.4325	0.2170	0.0775	0.0179
	0.10	0.9998	0.9959	0.9814	0.9461	0.8781	0.7696	0.6219	0.4504	0.2823	0.1464	0.0592
	0.15	0.9977	0.9834	0.9538	0.9030	0.8271	0.7248	0.5993	0.4597	0.3197	0.1957	0.1011
	0.20	0.9923	0.9645	0.9231	0.8647	0.7882	0.6939	0.5846	0.4657	0.3450	0.2326	0.1383
	0.25	0.9841	0.9425	0.8930	0.8314	0.7571	0.6707	0.5739	0.4701	0.3639	0.2618	0.1708
	0.30	0.9745	0.9198	0.8649	0.8025	0.7316	0.6523	0.5656	0.4735	0.3788	0.2858	0.1994
	0.35	0.9647	0.8982	0.8395	0.7775	0.7102	0.6372	0.5589	0.4763	0.3912	0.3061	0.2247
	0.40	0.9556	0.8788	0.8174	0.7561	0.6922	0.6247	0.5535	0.4788	0.4018	0.3239	0.2476
	0.45	0.9478	0.8623	0.7988	0.7384	0.6775	0.6147	0.5494	0.4814	0.4113	0.3398	0.2686
	0.50	0.9415	0.8492	0.7841	0.7246	0.6663	0.6073	0.5469	0.4846	0.4204	0.3547	0.2883
∞	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	0.9995	0.9950	0.9708	0.8910	0.7162	0.4583	0.2086	0.0590	0.0087
	0.10	1.0000	0.9981	0.9888	0.9619	0.9037	0.8017	0.6530	0.4713	0.2883	0.1410	0.0508
	0.15	1.0000	0.9892	0.9650	0.9202	0.8494	0.7495	0.6225	0.4772	0.3289	0.1966	0.0969
	0.20	1.0000	0.9732	0.9360	0.8813	0.8074	0.7141	0.6034	0.4809	0.3549	0.2366	0.1375
	0.25	1.0000	0.9532	0.9064	0.8470	0.7740	0.6878	0.5898	0.4834	0.3736	0.2672	0.1723
	0.30	1.0000	0.9319	0.8782	0.8169	0.7466	0.6671	0.5794	0.4854	0.3881	0.2919	0.2022
	0.35	1.0000	0.9112	0.8525	0.7908	0.7235	0.6502	0.5710	0.4870	0.3999	0.3125	0.2285
	0.40	1.0000	0.8925	0.8300	0.7684	0.7043	0.6363	0.5643	0.4885	0.4100	0.3303	0.2520
	0.45	1.0000	0.8765	0.8110	0.7499	0.6885	0.6251	0.5591	0.4902	0.4189	0.3461	0.2732
	0.50	1.0000	0.8638	0.7960	0.7353	0.6763	0.6167	0.5557	0.4926	0.4275	0.3607	0.2930

St.Diff = 2.5

		Rho*										
n	**Rho**	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
10	0.00	0.9995	0.9972	0.9890	0.9677	0.9239	0.8494	0.7420	0.6080	0.4617	0.3210	0.2017
	0.05	0.9969	0.9896	0.9731	0.9417	0.8899	0.8140	0.7139	0.5946	0.4653	0.3382	0.2253
	0.10	0.9907	0.9771	0.9530	0.9147	0.8591	0.7846	0.6918	0.5842	0.4684	0.3526	0.2460
	0.15	0.9808	0.9611	0.9310	0.8884	0.8315	0.7596	0.6736	0.5759	0.4710	0.3648	0.2643
	0.20	0.9678	0.9428	0.9084	0.8633	0.8066	0.7381	0.6583	0.5690	0.4733	0.3754	0.2808
	0.25	0.9524	0.9231	0.8858	0.8396	0.7841	0.7191	0.6450	0.5630	0.4753	0.3849	0.2959
	0.30	0.9353	0.9027	0.8636	0.8173	0.7635	0.7021	0.6333	0.5578	0.4772	0.3935	0.3098
	0.35	0.9174	0.8822	0.8420	0.7963	0.7446	0.6867	0.6228	0.5532	0.4789	0.4014	0.3228
	0.40	0.8993	0.8621	0.8214	0.7766	0.7271	0.6727	0.6132	0.5490	0.4805	0.4087	0.3351
	0.45	0.8816	0.8428	0.8020	0.7583	0.7111	0.6599	0.6047	0.5453	0.4821	0.4156	0.3468
	0.50	0.8651	0.8250	0.7842	0.7416	0.6965	0.6485	0.5970	0.5422	0.4838	0.4222	0.3581
30	0.00	1.0000	1.0000	1.0000	1.0000	0.9996	0.9954	0.9682	0.8710	0.6634	0.3894	0.1626
	0.05	1.0000	1.0000	0.9996	0.9974	0.9888	0.9629	0.9022	0.7888	0.6193	0.4174	0.2297
	0.10	0.9999	0.9990	0.9955	0.9859	0.9640	0.9215	0.8489	0.7401	0.5967	0.4324	0.2716
	0.15	0.9989	0.9948	0.9854	0.9667	0.9343	0.8828	0.8078	0.7069	0.5823	0.4422	0.3012
	0.20	0.9958	0.9865	0.9700	0.9436	0.9042	0.8487	0.7750	0.6821	0.5718	0.4494	0.3239
	0.25	0.9899	0.9743	0.9513	0.9189	0.8754	0.8188	0.7480	0.6626	0.5638	0.4549	0.3420
	0.30	0.9813	0.9590	0.9305	0.8940	0.8483	0.7923	0.7250	0.6464	0.5572	0.4595	0.3572
	0.35	0.9706	0.9417	0.9087	0.8696	0.8231	0.7685	0.7051	0.6327	0.5517	0.4634	0.3702
	0.40	0.9586	0.9235	0.8869	0.8460	0.7997	0.7470	0.6875	0.6207	0.5469	0.4667	0.3817
	0.45	0.9463	0.9052	0.8657	0.8238	0.7780	0.7276	0.6718	0.6102	0.5428	0.4698	0.3920
	0.50	0.9343	0.8879	0.8459	0.8033	0.7584	0.7101	0.6578	0.6010	0.5393	0.4726	0.4015
50	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9951	0.9484	0.7567	0.4098	0.1247
	0.05	1.0000	1.0000	0.9999	0.9994	0.9959	0.9818	0.9390	0.8405	0.6680	0.4413	0.2243
	0.10	1.0000	0.9997	0.9980	0.9923	0.9773	0.9441	0.8812	0.7779	0.6313	0.4542	0.2763
	0.15	0.9996	0.9972	0.9907	0.9766	0.9500	0.9049	0.8353	0.7367	0.6096	0.4618	0.3098
	0.20	0.9979	0.9911	0.9779	0.9555	0.9205	0.8693	0.7987	0.7069	0.5946	0.4670	0.3339
	0.25	0.9941	0.9809	0.9609	0.9318	0.8914	0.8376	0.7686	0.6837	0.5834	0.4709	0.3526
	0.30	0.9879	0.9674	0.9413	0.9072	0.8637	0.8094	0.7433	0.6648	0.5744	0.4741	0.3678
	0.35	0.9798	0.9515	0.9202	0.8827	0.8376	0.7841	0.7214	0.6489	0.5670	0.4767	0.3806
	0.40	0.9704	0.9344	0.8986	0.8587	0.8133	0.7613	0.7021	0.6352	0.5606	0.4789	0.3916
	0.45	0.9606	0.9170	0.8775	0.8360	0.7907	0.7406	0.6849	0.6231	0.5551	0.4809	0.4015
	0.50	0.9509	0.9004	0.8577	0.8150	0.7702	0.7220	0.6696	0.6125	0.5503	0.4828	0.4104
100	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9935	0.8737	0.4285	0.0664
	0.05	1.0000	1.0000	1.0000	0.9998	0.9986	0.9915	0.9634	0.8821	0.7126	0.4633	0.2161
	0.10	1.0000	0.9999	0.9990	0.9956	0.9850	0.9590	0.9047	0.8077	0.6603	0.4728	0.2795
	0.15	0.9999	0.9984	0.9938	0.9827	0.9605	0.9205	0.8557	0.7599	0.6316	0.4778	0.3165
	0.20	0.9990	0.9937	0.9829	0.9635	0.9319	0.8841	0.8163	0.7258	0.6125	0.4811	0.3419
	0.25	0.9967	0.9852	0.9674	0.9408	0.9029	0.8514	0.7841	0.6997	0.5985	0.4835	0.3609
	0.30	0.9927	0.9731	0.9488	0.9166	0.8749	0.8221	0.7569	0.6787	0.5876	0.4853	0.3760
	0.35	0.9870	0.9585	0.9283	0.8921	0.8483	0.7957	0.7335	0.6611	0.5785	0.4869	0.3885
	0.40	0.9804	0.9423	0.9071	0.8680	0.8233	0.7718	0.7129	0.6460	0.5709	0.4882	0.3992
	0.45	0.9733	0.9258	0.8862	0.8450	0.8001	0.7502	0.6946	0.6328	0.5643	0.4893	0.4086
	0.50	0.9663	0.9099	0.8665	0.8237	0.7789	0.7308	0.6784	0.6212	0.5586	0.4905	0.4171
∞	0.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.05	1.0000	1.0000	1.0000	1.0000	0.9997	0.9972	0.9827	0.9246	0.7673	0.4912	0.2011
	0.10	1.0000	1.0000	0.9996	0.9977	0.9909	0.9718	0.9270	0.8386	0.6923	0.4939	0.2821
	0.15	1.0000	0.9991	0.9960	0.9877	0.9696	0.9350	0.8757	0.7836	0.6549	0.4952	0.3236
	0.20	1.0000	0.9958	0.9871	0.9705	0.9424	0.8984	0.8338	0.7451	0.6312	0.4959	0.3503
	0.25	1.0000	0.9888	0.9732	0.9491	0.9138	0.8647	0.7994	0.7159	0.6141	0.4965	0.3696
	0.30	1.0000	0.9783	0.9558	0.9255	0.8856	0.8344	0.7705	0.6927	0.6009	0.4969	0.3845
	0.35	1.0000	0.9650	0.9361	0.9012	0.8586	0.8071	0.7455	0.6733	0.5903	0.4972	0.3967
	0.40	1.0000	0.9501	0.9154	0.8770	0.8331	0.7822	0.7237	0.6568	0.5813	0.4975	0.4070
	0.45	1.0000	0.9346	0.8948	0.8539	0.8093	0.7597	0.7043	0.6424	0.5735	0.4978	0.4159
	0.50	1.0000	0.9196	0.8752	0.8323	0.7876	0.7395	0.6872	0.6298	0.5669	0.4982	0.4239

APPENDIX B: EQUATIONS

A test statistic for testing H_0: $\mu_i = \mu_{i'}$ vs. H_A: $\mu_i \neq \mu_{i'}$ can be computed from our data as follows: Let ρ denote the true ICC. Let ρ_0 denote the value the researcher chooses for the ICC. Let μ_{0i} and $\mu_{0i'}$ represent the hypothesized values of the mean for treatment i and the mean for treatment i' respectively. Let $n_{i\cdot} = \sum_{j=1}^{b_i} n_{ij}$ and $\langle n_{i\cdot} \rangle^2 = \sum_{j=1}^{b_i} n_{ij}^2$. Let T be defined as

$$T = \frac{(\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}) - (\mu_{0i} - \mu_{0i'})}{\sqrt{\hat{\sigma}_\varepsilon^2 \left\{ \left(\frac{\rho_0}{1-\rho_0} \right) \left[\left(\frac{1}{n_{i\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 + \left(\frac{1}{n_{i'\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 \right] + \left[\frac{n_{ij} + n_{i'j}}{n_{ij} n_{i'j}} \right] \right\}}} \qquad (B.1)$$

where $\hat{\sigma}_\varepsilon^2$ represents a pooled estimate of the within class variance.

The probability of rejecting H_0 for an upper-tail test at the 0.05 level can be determined using the following steps:

$$P\left[\frac{(\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}) - (\mu_{0i} - \mu_{0i'})}{\sqrt{\hat{\sigma}_\varepsilon^2 \left\{ \left(\frac{\rho_0}{1-\rho_0} \right) \left[\left(\frac{1}{n_{i\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 + \left(\frac{1}{n_{i'\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 \right] + \left[\frac{n_{ij} + n_{i'j}}{n_{ij} n_{i'j}} \right] \right\}}} > t_{0.05,\upsilon} \right] \qquad (B.2)$$

$$= P\left[\frac{(\bar{y}_{i\cdot\cdot} - \bar{y}_{i'\cdot\cdot}) - (\mu_{0i} - \mu_{0i'}) + (\mu_i - \mu_{i'}) - (\mu_i - \mu_{i'})}{\sqrt{\hat{\sigma}_\varepsilon^2 \left\{ \left(\frac{\rho_0}{1-\rho_0} \right) \left[\left(\frac{1}{n_{i\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 + \left(\frac{1}{n_{i'\cdot}} \right)^2 \langle n_{i\cdot} \rangle^2 \right] + \left[\frac{n_{ij} + n_{i'j}}{n_{ij} n_{i'j}} \right] \right\}}} \right].$$

$$\left.\frac{\sqrt{\sigma_\varepsilon^2\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}{\sqrt{\sigma_\varepsilon^2\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}>t_{0.05,\upsilon}\right]$$

$$=P\left[\frac{\frac{(\bar{y}_{i\cdot\cdot}-\bar{y}_{i'\cdot\cdot})-(\mu_1-\mu_2)}{\sqrt{\sigma_\varepsilon^2\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}+\frac{(\mu_1-\mu_2)-(\mu_{01}-\mu_{02})}{\sqrt{\sigma_\varepsilon^2\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}}{\frac{\sqrt{\hat{\sigma}_\varepsilon^2}}{\sqrt{\sigma_\varepsilon^2}}}>\right.$$

$$\left.\frac{\sqrt{\left\{\left(\frac{\rho_0}{1-\rho_0}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}{\sqrt{\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}\cdot\left(t_{0.05,\upsilon}\right)\right]$$

$$=P\left[\frac{Z+\lambda}{\sqrt{U/\upsilon}}>\frac{\sqrt{\left\{\left(\frac{\rho_0}{1-\rho_0}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}{\sqrt{\left\{\left(\frac{\rho}{1-\rho}\right)\left[\left(\frac{1}{n_{i\cdot}}\right)^2\langle n_{i\cdot}\rangle^2+\left(\frac{1}{n_{i'\cdot}}\right)^2\langle n_{i\cdot}\rangle^2\right]+\left[\frac{n_{ij}+n_{i'j}}{n_{ij}n_{i'j}}\right]\right\}}}\cdot\left(t_{0.05,\upsilon}\right)\right] \qquad (B.3)$$

where Z~n(0,1), λ is a constant (our non-centrality parameter), and $U\sim\chi_\upsilon^2$. To the left of the inequality is a random variable with a non-central t distribution with non-centrality parameter λ and degrees of freedom υ.

APPENDIX C: DATA SETS

Data Set C.1: Class Data

Two introductory statistics courses (class) were taught by two different methods (trt). Final grades were recorded for each student on a scale from 0 (Failure) to 4 (Excellent).

```
Data class;
Input class trt student y;
Datalines;
```

1	1	1	3
1	1	2	3
1	1	3	4
1	1	4	3
1	1	5	2
1	1	6	0
1	1	7	4
1	1	8	2
1	1	9	2
1	1	10	3
1	1	11	3
1	1	12	3
1	1	13	0
1	1	14	3
1	1	15	3
1	1	16	4
1	1	17	1
1	1	18	2
1	1	19	3
1	1	20	3
1	1	21	4
1	1	22	3
1	1	23	3
1	1	24	2
1	1	25	2
1	1	26	2
1	1	27	4
1	1	28	4
1	1	29	4
1	1	30	3
1	1	31	3
1	1	32	3
1	1	33	4

1	1	34	4
1	1	35	3
2	2	1	4
2	2	2	0
2	2	3	3
2	2	4	4
2	2	5	3
2	2	6	4
2	2	7	3
2	2	9	4
2	2	10	4
2	2	11	3
2	2	12	4
2	2	13	4
2	2	15	3
2	2	16	3
2	2	17	2
2	2	18	4
2	2	19	4
2	2	20	4
2	2	21	3
2	2	22	3
2	2	23	3
2	2	24	4
2	2	25	4
2	2	26	3
2	2	27	3
2	2	28	4
2	2	30	4
2	2	31	4
2	2	32	4
2	2	33	3
2	2	34	3
2	2	35	4
2	2	36	3
2	2	37	4
2	2	38	2

Data Set C.2: KSU Grades Data

The following data set contains estimated intraclass correlation coefficients for KSU grades given in a selection of undergraduate courses offered from 2001-2003. The experimental unit in the data is the course/per semester and the sub-sampling unit is the course sections for that semester. The total number of students (stud_tot), the total number of sections (class_tot), and the ratio of the two are also listed.

```
Data rho;
Input year semester course rho stud_tot class_tot stud_avg;
Datalines;
```

2002	F	ENGL100	0.0524	1252	63	19.9
2001	F	PSYCH110	0.0214	1186	19	62.4
2001	F	ENGL100	0.0718	1157	60	19.3
2002	F	PSYCH110	0.0203	1045	7	149.3
2003	SP	PSYCH110	0.0089	972	6	162.0
2002	F	CHM210	0.0981	907	5	181.4
2002	SP	PSYCH110	0.1296	906	11	82.4
2001	F	CIS101	0.0040	894	6	149.0
2002	F	CIS101	0.0001	856	7	122.3
2001	SP	PSYCH110	0.1918	829	9	92.1
2001	SP	CIS101	0.1459	813	8	101.6
2001	F	CHM210	0.0171	808	5	161.6
2002	SP	CIS101	0.0947	738	8	92.3
2003	SP	CIS101	0.1011	716	8	89.5
2002	SP	ENGL100	0.0829	702	42	16.7
2001	SP	ENGL100	0.0497	663	36	18.4
2003	SP	ENGL100	0.0525	562	33	17.0
2001	SP	CHM230	0.0103	536	4	134.0
2002	SP	CHM230	0.0327	512	4	128.0
2003	SP	CHM230	0.1096	502	4	125.5
2003	SP	CHM210	0.0832	448	2	224.0
2002	SP	CHM210	0.0117	378	2	189.0
2001	F	MATH010	0.0067	352	24	14.7
2002	F	MATH010	0.0014	336	26	12.9
2001	F	SPAN161	0.0724	307	14	21.9
2002	F	SPAN161	0.0893	247	11	22.5
2001	SP	PSYCH202	0.0895	244	2	122.0
2001	F	PSYCH202	0.2266	241	2	120.5
2002	F	MUSIC250	0.1341	208	3	69.3
2003	SP	SPAN161	0.0466	195	8	24.4
2002	SP	SPAN161	0.1380	187	8	23.4

2001	SP	SPAN161	0.0689	184	8	23.0
2003	SP	MUSIC255	0.2034	153	17	9.0
2001	F	MUSIC255	0.0247	145	17	8.5
2001	SP	MUSIC255	0.0699	130	17	7.6
2003	SP	MATH010	0.0439	117	15	7.8
2001	SP	MATH010	0.0863	116	15	7.7
2002	SP	MATH010	0.1012	103	15	6.9
2002	F	MATH100	0.0424	96	4	24.0
2001	SP	PSYCH350	0.0479	54	3	18.0
2001	SP	ENGL125	0.3358	35	4	8.8
2002	SP	ENGL125	0.1127	30	3	10.0
2001	F	CHM111	0.0102	18	2	9.0

Data Set C.3: Spider Mites Test Data

One of four different treatments (trt) was applied to the eight potted plants (pot) within each of four greenhouses (grnhs). At the end of treatment period, the number of spider mites was counted (count).

```
Data class;
Input grnhs trt pot count;
Datalines;
```

1	1	1	25
1	1	2	47
1	1	3	43
1	1	4	30
1	1	5	55
1	1	6	30
1	1	7	40
1	1	8	42
2	2	1	0
2	2	2	0
2	2	3	0
2	2	4	0
2	2	5	1
2	2	6	3
2	2	7	0
2	2	8	0
3	3	1	14
3	3	2	44
3	3	3	3
3	3	4	15
3	3	5	13
3	3	6	3
3	3	7	26
3	3	8	6
4	4	1	0
4	4	2	0
4	4	3	0
4	4	4	0
4	4	5	1
4	4	6	0
4	4	7	0
4	4	8	9

Data Set C.4: Spider Mites Pre-treatment Data

The number of two-spotted spider mites (count) was determined for each of eight potted plants (pot) for seven different varieties (variety) within each of four greenhouses (grnhs).

```
Data class;
Input grnhs variety pot count;
Datalines;
```

1	Cajun Cranberry	1	19
1	Cajun Cranberry	2	16
1	Cajun Cranberry	3	13
1	Cajun Cranberry	4	5
1	Cajun Cranberry	5	25
1	Cajun Cranberry	6	3
1	Cajun Cranberry	7	2
1	Cajun Cranberry	8	19
1	Cajun White	1	3
1	Cajun White	2	0
1	Cajun White	3	8
1	Cajun White	4	5
1	Cajun White	5	2
1	Cajun White	6	0
1	Cajun White	7	0
1	Cajun White	8	0
1	Impulse Orange White	1	0
1	Impulse Orange White	2	0
1	Impulse Orange White	3	0
1	Impulse Orange White	4	0
1	Impulse Orange White	5	0
1	Impulse Orange White	6	0
1	Impulse Orange White	7	15
1	Impulse Orange White	8	0
1	Impulse Orange	1	0
1	Impulse Orange	2	110
1	Impulse Orange	3	20
1	Impulse Orange	4	100
1	Impulse Orange	5	2
1	Impulse Orange	6	25
1	Impulse Orange	7	90
1	Impulse Orange	8	20
2	Cajun Cranberry	1	2
2	Cajun Cranberry	2	5
2	Cajun Cranberry	3	0
2	Cajun Cranberry	4	1

2	Cajun Cranberry	5	3
2	Cajun Cranberry	6	12
2	Cajun Cranberry	7	0
2	Cajun Cranberry	8	0
2	Cajun White	1	12
2	Cajun White	2	4
2	Cajun White	3	23
2	Cajun White	4	2
2	Cajun White	5	0
2	Cajun White	6	45
2	Cajun White	7	65
2	Cajun White	8	21
2	Impulse Orange White	1	0
2	Impulse Orange White	2	0
2	Impulse Orange White	3	16
2	Impulse Orange White	4	9
2	Impulse Orange White	5	1
2	Impulse Orange White	6	17
2	Impulse Orange White	7	18
2	Impulse Orange White	8	2
2	Impulse Orange	1	5
2	Impulse Orange	2	15
2	Impulse Orange	3	13
2	Impulse Orange	4	20
2	Impulse Orange	5	14
2	Impulse Orange	6	12
2	Impulse Orange	7	20
2	Impulse Orange	8	15
3	Cajun Cranberry	1	9
3	Cajun Cranberry	2	6
3	Cajun Cranberry	3	48
3	Cajun Cranberry	4	12
3	Cajun Cranberry	5	0
3	Cajun Cranberry	6	8
3	Cajun Cranberry	7	50
3	Cajun Cranberry	8	1
3	Cajun White	1	5
3	Cajun White	2	0
3	Cajun White	3	70
3	Cajun White	4	38
3	Cajun White	5	5
3	Cajun White	6	3
3	Cajun White	7	22
3	Cajun White	8	3
3	Impulse Orange White	1	1
3	Impulse Orange White	2	15

3	Impulse Orange White	3	0
3	Impulse Orange White	4	2
3	Impulse Orange White	5	0
3	Impulse Orange White	6	4
3	Impulse Orange White	7	0
3	Impulse Orange White	8	0
3	Impulse Orange	1	1
3	Impulse Orange	2	0
3	Impulse Orange	3	20
3	Impulse Orange	4	26
3	Impulse Orange	5	36
3	Impulse Orange	6	22
3	Impulse Orange	7	3
3	Impulse Orange	8	1
4	Cajun Cranberry	1	2
4	Cajun Cranberry	2	4
4	Cajun Cranberry	3	0
4	Cajun Cranberry	4	35
4	Cajun Cranberry	5	8
4	Cajun Cranberry	6	0
4	Cajun Cranberry	7	0
4	Cajun Cranberry	8	25
4	Cajun White	1	0
4	Cajun White	2	0
4	Cajun White	3	0
4	Cajun White	4	5
4	Cajun White	5	0
4	Cajun White	6	0
4	Cajun White	7	0
4	Cajun White	8	0
4	Impulse Orange White	1	23
4	Impulse Orange White	2	2
4	Impulse Orange White	3	9
4	Impulse Orange White	4	3
4	Impulse Orange White	5	5
4	Impulse Orange White	6	19
4	Impulse Orange White	7	6
4	Impulse Orange White	8	10
4	Impulse Orange	1	0
4	Impulse Orange	2	80
4	Impulse Orange	3	7
4	Impulse Orange	4	8
4	Impulse Orange	5	13
4	Impulse Orange	6	1
4	Impulse Orange	7	25
4	Impulse Orange	8	12

1	Summer Rose Lilac	1	55
1	Summer Rose Lilac	2	34
1	Summer Rose Lilac	3	36
1	Summer Rose Lilac	4	45
1	Summer Rose Lilac	5	18
1	Summer Rose Lilac	6	58
1	Summer Rose Lilac	7	38
1	Summer Rose Lilac	8	53
1	Summer Rose Red	1	1
1	Summer Rose Red	2	4
1	Summer Rose Red	3	8
1	Summer Rose Red	4	20
1	Summer Rose Red	5	30
1	Summer Rose Red	6	22
1	Summer Rose Red	7	9
1	Summer Rose Red	8	12
1	Ivy Geranium	1	5
1	Ivy Geranium	2	46
1	Ivy Geranium	3	3
1	Ivy Geranium	4	62
1	Ivy Geranium	5	37
1	Ivy Geranium	6	45
1	Ivy Geranium	7	43
1	Ivy Geranium	8	45
2	Summer Rose Lilac	1	87
2	Summer Rose Lilac	2	50
2	Summer Rose Lilac	3	37
2	Summer Rose Lilac	4	73
2	Summer Rose Lilac	5	55
2	Summer Rose Lilac	6	36
2	Summer Rose Lilac	7	12
2	Summer Rose Lilac	8	47
2	Summer Rose Red	1	0
2	Summer Rose Red	2	14
2	Summer Rose Red	3	0
2	Summer Rose Red	4	23
2	Summer Rose Red	5	18
2	Summer Rose Red	6	42
2	Summer Rose Red	7	12
2	Summer Rose Red	8	10
2	Ivy Geranium	1	27
2	Ivy Geranium	2	8
2	Ivy Geranium	3	25
2	Ivy Geranium	4	14
2	Ivy Geranium	5	55
2	Ivy Geranium	6	10

2	Ivy Geranium	7	10
2	Ivy Geranium	8	15
3	Summer Rose Lilac	1	43
3	Summer Rose Lilac	2	12
3	Summer Rose Lilac	3	7
3	Summer Rose Lilac	4	5
3	Summer Rose Lilac	5	3
3	Summer Rose Lilac	6	66
3	Summer Rose Lilac	7	22
3	Summer Rose Lilac	8	37
3	Summer Rose Red	1	1
3	Summer Rose Red	2	1
3	Summer Rose Red	3	0
3	Summer Rose Red	4	17
3	Summer Rose Red	5	1
3	Summer Rose Red	6	20
3	Summer Rose Red	7	0
3	Summer Rose Red	8	6
3	Ivy Geranium	1	42
3	Ivy Geranium	2	35
3	Ivy Geranium	3	43
3	Ivy Geranium	4	35
3	Ivy Geranium	5	42
3	Ivy Geranium	6	17
3	Ivy Geranium	7	38
3	Ivy Geranium	8	30
4	Summer Rose Lilac	1	7
4	Summer Rose Lilac	2	8
4	Summer Rose Lilac	3	101
4	Summer Rose Lilac	4	46
4	Summer Rose Lilac	5	52
4	Summer Rose Lilac	6	58
4	Summer Rose Lilac	7	16
4	Summer Rose Lilac	8	11
4	Summer Rose Red	1	26
4	Summer Rose Red	2	22
4	Summer Rose Red	3	32
4	Summer Rose Red	4	37
4	Summer Rose Red	5	14
4	Summer Rose Red	6	11
4	Summer Rose Red	7	60
4	Summer Rose Red	8	19
4	Ivy Geranium	1	37
4	Ivy Geranium	2	1
4	Ivy Geranium	3	70
4	Ivy Geranium	4	63

4	Ivy Geranium	5	26
4	Ivy Geranium	6	36
4	Ivy Geranium	7	7
4	Ivy Geranium	8	80

VERIFICATION OF RESULTS

Verification that the function listed as Equation 2.2.9 is a valid probability density function:

$$f_Y(y)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}c\left[cy-\frac{1}{n}\right]^{\alpha-1}\left[1+cy-\frac{1}{n}\right]^{-(\alpha+\beta)},\ \frac{1}{cn}<y<\infty \qquad (2.2.9)$$

where

$$c=\frac{\rho_0}{1-\rho_0}+\frac{1}{n}.$$

We generate values of x in the range of greatest probability: $\frac{1}{cn}$ through 10 by 0.0001 for $\alpha=1,\beta=10$, $\frac{1}{cn}$ through 10 by 0.001 for $\alpha=5,\beta=30$, and $\frac{1}{cn}$ through 200 by 0.001 for $\alpha=10,\beta=4$ to check for (1) values being positive, and (2) values integrating to 1.0 by use of Riehmann sums. This is repeated for the following situations:

α	β	ρ_0	n
1	10	0.1	10
1	10	0.1	1000
1	10	0.5	10
1	10	0.5	1000

α	β	ρ_0	n
5	30	0.1	10
5	30	0.1	1000
5	30	0.5	10
5	30	0.5	1000

α	β	ρ_0	n
10	4	0.1	10
10	4	0.1	1000
10	4	0.5	10
10	4	0.5	1000

In all cases, the function returned positive probabilities and integrated to 1.0.

Verification that the function listed as Equation 2.2.12 is a valid probability density function:

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}\left(\frac{1}{1+x}\right)^{\alpha+\beta}, 0 < x < \infty \qquad (2.2.12)$$

which is the equivalent to $\frac{\alpha}{\beta}F$, a constant times an F-distributed random variable with $\upsilon_1 = 2\alpha$ and $\upsilon_2 = 2\beta$.

We generate values of x in the range of greatest probability: 0 through 100 by 0.001 for $\alpha = 1, \beta = 10$, 0 through 5 by 0.001 for $\alpha = 5, \beta = 30$, and 0 through 100 by 0.001 for $\alpha = 10, \beta = 4$ to check for (1) values being positive, and (2) values integrating to 1.0 by use of Riehmann sums. This is repeated for the following situations:

α	β	ρ_0	n
1	10	0.1	10
1	10	0.1	1000
1	10	0.5	10
1	10	0.5	1000

α	β	ρ_0	n
5	30	0.1	10
5	30	0.1	1000
5	30	0.5	10
5	30	0.5	1000

α	β	ρ_0	n
10	4	0.1	10
10	4	0.1	1000
10	4	0.5	10
10	4	0.5	1000

In all cases, the function returned positive probabilities and integrated to 1.0.

Verification that $\hat{\sigma}_\varepsilon^2$ from Equation 3.1.2 is an unbiased estimator of σ_ε^2.

A simulation of was conducted to determine whether Equation 3.1.2 is an unbiased estimator of σ_ε^2. A data set was generated with the value of σ_ε^2 being 1.0 and the value of ρ being 0.10. An estimate of σ_ε^2 was computed using both Equation 3.1.2 (pooled) and Equation 2.3.4 (UMVUE). The value 0.20 was used as a plug-in value in place of ρ in Equation 2.3.4. These steps were iterated 10,000 times and a mean and standard deviation were computed on the estimates. The following table shows the results.

	t	4			
	b	1		3	
	n	5	20	5	20
Pooled	**Mean**	1.0067340	1.0048189	1.0057520	0.9996724
	St. Dev.	0.3485084	0.1645745	0.2020840	0.0913511
UMVUE	**Mean**	1.0067340	1.0048189	0.9573357	0.9842369
	St. Dev.	0.3485084	0.1645745	0.1763341	0.0885420

As can be seen, Equation 3.1.2 (pooled) provides an unbiased estimate of σ_ε^2, and the estimate using Equation 2.3.4 (UMVUE) with a plug-in value in place of ρ can lead to an estimate of σ_ε^2 that is biased.

BIBLIOGRAPHY

Bansal, N. K. (2000). Bayes Estimation of Intraclass Correlation Coefficient. Communications in Statistics – Theory and Methods, 29, 79-93.

Barcikowski, R. S. (1981). Statistical Power with Group Mean as the Unit of Analysis. Journal of Educational Statistics, 6, 267-285.

Blair, R C., Higgins, J. J., Topping, M. E. H., & Mortimer, A. L. (1983). An Investigation of the Robustness of the t-test to Unit of Analysis Violations. Educational and Psychological Measurement, 43, 69-80.

Blair, R. C., & Higgins, J. J. (1986). Comment on "Statistical Power With Group Mean as the Unit of Analysis". Journal of Educational Statistics Summer, 2, 161-169.

Bond, M., & Higgins, J. J. (2001). A Note on 'A Comparison of Bayes and Maximum Likelihood Estimation of the Intraclass Correlation Coefficient'. Communications in Statistics – Theory and Methods, 30, 371-380.

Burch, B. D., & Harris, I. R. (1999). Bayesian Estimators of the Intraclass Correlation Coefficient in the One-Way Random Effects Model. Communications in Statistics – Theory and methods, 28, 1247-1272.

Casella, G., & Berger, R. L. (2002). Statistical Inference (2nd ed.). Pacific Grove, CA: Duxbury.

Graybill, F. A. (1976). *Theory and* Application of the Linear Model. Pacific Grove, CA: Wadsworth.

Groggel, D. J., Wackerly, D. D., & Rao, P. V. (1988). Nonparametric Estimation in One-Way Random Effects Models. Communications in Statistics – Simulation, 17, 887-903.

Little, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). SAS® System for Mixed Models. Cary, NC: SAS® Institute Inc.

Milliken, G. A., & Johnson, D. E. (1992). Analysis of Messy Data, Volume 1: Experimental Design. London, UK: Chapman & Hall.

Ostle, B., & Malone, L. C. (1988). Statistics in Research (4th ed.). Ames, IA: Iowa State University Press.

SAS® Institute Inc. (1999). SAS/STAT® User's Guide, Version 8. Cary, NC: Author.

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., & Hochberg, Y. (1999). Multiple Comparisons and Multiple Tests Using the SAS® System. Cary, NC: SAS® Institute Inc.