The Merging of Statistics Education, Consulting and Research: A Case Study

Daniel R. Jeske, Scott M. Lesch and Hongjie Deng
University of California - Riverside

Abstract

It is shown how student participation in a real consulting project can be leveraged to achieve the dual goals of (i) developing statistical consulting skills in graduate students, and (ii) enhancing the instructional effectiveness of statistical methodology. Achieving these goals is the primary mission of the Statistical Consulting Collaboratory at the University of California, Riverside. The paper gives a detailed illustration of the how the goals were achieved by reporting on an interesting case study, with special emphasis given to describing the involvement of students and the alternative ways in which the project found its way into classrooms.

1. Introduction

The Department of Statistics at the University of California at Riverside formally established a Statistical Consulting Collaboratory in the Fall of 2003. Agreeing with Carter, Scheaffer and Marks (1986), the first priority of the Collaboratory is to contribute effectively to the academic objectives of the Statistics Department. The Collaboratory is uniquely positioned to do this through the development and application of statistical methods to real world problems. Specific contributions the Collaboratory is making include: 1) curriculum material for the department's graduate-level statistical consulting class that addresses traditional pedagogical objectives [see, for example, Hertzberg, Clark and Brogan (2000), Taplin (2003), Johnson and Warner (2004), and Birch and Morgan (2005)], 2) curriculum material that both reinforces and broadens student knowledge in statistical methodology, 3) consulting opportunities for undergraduate and graduate students, 4) research opportunities that can develop into PhD dissertation topics, and 5) resume building activities for students through publication opportunities and industry internships made available through the Collaboratory client network.

The importance of skills on the non-technical side of consulting has been discussed elsewhere [see, for example, Boen and Zahn (1982), Kirk (1991), Derr (2000)]. While these skills are essential, an equally (arguably more) important skill is broad technical expertise that enables choosing a correct analysis on the basis of informed judgments. A distinguishing characteristic of the Collaboratory is its ability to promote a statistical consulting pedagogy that goes beyond pragmatic solutions to consulting problems by exploring additional statistical methodology that is related to the client problem. Through this influence, the Collaboratory enhances the students' ability to select appropriate methodology for a given problem. Moreover, it cultivates a curiosity and a self-sufficiency, which are attributes Russell (2001) discusses as crucial for a statistical consultant. In this paper, a case study is described that illustrates how the Collaboratory achieves these objectives. While case studies on statistical consulting have appeared in other work [see, for example, Tweedie (1998) and Cabrera and McDougall (2002)], our case study is different in that it specifically highlights how the Collaboratory influenced curriculum for the statistical consulting class. For example, references are given throughout to exercises included in Appendix A that were used as class assignments. The exercises by themselves are interesting and could be used as a supplement for a variety of statistics classes.

In the remainder of this Section, a more detailed overview about the mission of the Collaboratory is provided, and the specific consulting problem is introduced. In Sections 2-4, the major tasks associated with the consulting problem are discussed and the opportunities they provided to enhance graduate student training are highlighted. Within each of these sections, the statistical methods used to solve the consulting problem are first introduced and then accounts of the educational benefits generated by the consulting problem are detailed. The paper concludes with discussion in Section 5.

1.1 Statistical Consulting Collaboratory

Clients of the Statistical Consulting Collaboratory include professors, graduate students and University administrators. Clients to-date have been affiliated both with UC-Riverside and also other local Universities. In addition, the Collaboratory attracts industry clients through both personal networking and referrals. In the framework of Does and Zempleni (2001), the Collaboratory is a hybridization of a noncommercial and commercial consulting unit, though to-date the Collaboratory does not aggressively market itself to off-campus clients. The Collaboratory is directed by a tenured faculty member and employs a full-time Associate Director with an M.S. degree in Statistics. While the Director position is ultimately responsible for all of the activities of the Collaboratory, his/her primary emphasis is to create and nurture opportunities within the Collaboratory that make it look and operate like an academic unit. The primary role of the Associate Director is to lend his/her technical consulting skills to projects, but other significant responsibilities include supervising Collaboratory Research Assistants (CRAs) and managing some administrative aspects of the Collaboratory. The Collaboratory has typically supported 2-3 CRAs during the academic year with partial research assistantships. During the summer months a larger number of opportunities for part-time employment are available. While the majority of the CRAs are graduate students in Statistics, undergraduate students from both Statistics and other departments (e.g., Computer Science, Business and Mathematics) within the University have made contributions to some of the industry client projects along the lines of data base construction and development of customized data processing routines. The Director and Associate Director hire CRAs based on the level of their experience with applied statistics and/or computer skills, their demonstrated work ethic, and their interest for gaining experience with statistical consulting.

Projects that are taken on by the Collaboratory loosely fall into two categories: Service or Collaboration. Service describes projects that utilize standard statistical methods, both well-known and less well-known to the clients. Collaboration describes projects where there is some aspect of novelty either in the development or application of statistical methodology. To fund the support that is provided to students

working in the Collaboratory, fees are assessed for service projects. While the University provides the salary and benefits for the Associate Director, the fees are also intended to cover miscellaneous expenses such as software licenses and office supplies. Some projects start out as service but evolve into collaboration. When the transition to collaboration occurs and it is reasonable to expect the research could eventually be published in a Statistics journal, fees no longer apply. In the best collaborative relationships, a joint grant proposal would also be submitted.

The Statistics Department at UC-Riverside has a mandatory three quarter class on Statistical Consulting for both MS and PhD graduate students. The Consulting Class (CC) is taught by the Collaboratory Director and usually has 10-15 graduate students enrolled who are at least in the second year of their program. A great majority of the material covered in the CC is related to Collaboratory projects. Client visitations provide opportunities for the students to gain experience listening to clients and eliciting information that helps formulate objectives for the projects. Students are assigned to work on consulting projects independently and also in small groups. Lectures provide the students the necessary background they need to complete tasks associated with the projects. Throughout the duration of their work on the projects, students schedule meetings with the Director and/or Associate Director for additional direction and advice. Typically, students will have at least one interim meeting with the client before delivering a final presentation to them. The Director formulates homework exercises relating to each of the projects being addressed in the class. Appendix A contains the exercises that were extracted from the project being presented in this case study. The CC is a letter grade class, and includes a final exam that covers the statistical methodology relating to the consulting projects that were discussed during the quarter.

1.2 Consulting Problem

In order to respect proprietary issues, the client in this case study is referred to as Organization X. Organization X was an off-campus client that had t competing plans for a product's architecture, which in this paper will be indexed by the integer values 1 to t. With t alternative plans, there are C_2^t pairs of plans and a panel of independent judges was enlisted to evaluate each pair of plans. Individual judges could only serve on one panel, a logistics constraint that arose due to the fact judges were animals. Let n_{ij} denote the number of judges in the panel that compare the (i, j) pair of plans.

Each judge within a panel was able to express a preference as to which of the two plans is preferable. Prior to meeting with the Collaboratory, the data were analyzed by Organization X to construct approximate confidence intervals for the probabilities that a given plan is preferable to another plan. Let π_{ij} denote the probability that plan i is preferred over plan j when a judge compares the two plans. Assuming that the judges within a panel are a random sample from the targeted population for the product, the number of expressed preferences for plan i when it is compared with plan j, say S_{ij}, follows a binomial distribution with trial parameter n_{ij} and success probability π_{ij}. The $\{S_{ij}\}_{i<j}^t$ observations from an Organization X experiment with $t = 5$ are shown in Table 1. (The numbers enclosed by parentheses in Table 1 represent expected cell frequencies and will be discussed in Section 3.1.) The intent was to use 30 judges for each panel. However, only 29 judges participated in the $(1,2)$, $(3,4)$ and $(3,5)$ panels. Approximate confidence intervals for $\{\pi_{ij}\}_{i<j}^t$ were constructed by the client from the formula $\tilde{\pi}_{ij} \pm z_{\alpha/2}\sqrt{\tilde{\pi}_{ij}(1-\tilde{\pi}_{ij})/n_{ij}}$, where $\tilde{\pi}_{ij} = S_{ij}/n_{ij}$ [see, for example, Mendenhall, Beaver and Beaver (2006)].

| | Plan j | | | | |
Plan i	1	2	3	4	5
1	-	20 (23.4)	22 (20.3)	20 (16.6)	1 (2.7)
2	9 (5.6)	-	6 (9.9)	7 (6.8)	1 (0.7)
3	8 (9.8)	24 (20.0)	-	8 (10.8)	2 (1.3)
4	10 (13.4)	23 (23.2)	21 (18.7)	-	3 (2.2)
5	29 (27.2)	29 (29.3)	27 (27.7)	27 (27.8)	-

Table 1. Observed and Estimated Expected Cell Frequencies for Organization X Experiment

While Table 1 is the only data set provided to the Collaboratory by Organization X, they in fact do many experiments of the same type. The main goal of Organization X was to learn about the most appropriate type of analyses for data sets of this kind so that they could perform future analyses themselves. Organization X expressed specific interest in using the data to rank order the plans and to identify which plans were the "best" in a statistically significant way. Clearly, the confidence intervals they computed stop short of a formal ranking procedure. Organization X also expressed an interest in exploring the quality of their experimental design. In particular, they wanted to know if an alternative design could be employed that utilized fewer panels yet still provided enough information to adequately compare the alternative plans. A design that utilizes fewer panels would be attractive from the standpoint that it would be simpler to manage. The client noted that any proposed alternative design needed to adhere to the constraint that no judge could be asked to compare more than two plans.

1.3 Proposal Process

Organization X requested a proposal from the Collaboratory that outlined tasks and deliverables associated with their stated goals. During the winter quarter of 2005, the objectives of the consulting problem were introduced to a CC, along with the existing mode of data analysis being carried out by Organization X. The students were first asked to participate in a brainstorm discussion about what could be done for this client. To guide the discussion and provide some relevant technical background, a detailed introduction to the Bradley-Terry (1952) model for analyzing a paired comparison experiment was provided to the students. The students were then asked on a homework assignment to individually write a summary of this discussion and identify open issues concerning the proposed analysis techniques.

The Director then led a class discussion on basic proposal writing concepts and workload estimation techniques and used the submitted homework assignments to develop a draft of the proposal. Although the Director ultimately wrote the proposal, the students were able to observe firsthand what this activity involves. For most of the students, it was their first exposure to the challenge of setting a realistic project schedule that includes milestones and estimated costs. The students also observed the importance of performing background technical work (i.e., acquiring fundamental knowledge about Bradley-Terry models) that can help make a proposal more compelling to a client.

After the proposal was submitted to Organization X, an iterative feedback loop with the client began and the CC was kept abreast of the proposal progress. In one instance, the client requested customized software that would automate the proposed analysis methods to the extent they could import their data into one computer program and get every aspect of their analysis as the output. In a sense, the request was for an expert system, which was more than the Director wanted to commit to. As an alternative, it was suggested to the client that the analyses be done with off-the-shelf statistical software packages such as R or SAS, but not necessarily with one program and not necessarily without some human oversight. The students observed this decision-making process, and were also exposed to some of the important issues that arose during the proposal negotiating process. For example, they saw that it is acceptable, and even necessary, to declare some requests beyond the scope of the project. Furthermore, they gained a

better appreciation for why a proposal must contain well organized, clearly defined tasks and how to avoid the pitfall of being too vague or overextending when writing the scope-of-work.

The final proposal was ultimately approved by the client in late Spring, 2005. The remaining project work described in this paper began in the late summer of 2005 when a CRA began working on the estimation analyses.

2. Estimation

2.1 Consulting Application

Because ranking the plans based on evaluations from judges is of interest, it is natural to think of the Bradley-Terry (1952) modeling framework. As discussed in the previous section, the design used by Organization X is not the classic case where each judge evaluates every one of the C_2^t pairs of plans. Nevertheless, the key idea associated with the Bradley-Terry model can be used in conjunction with a logistic regression model for the independent observations $\{S_{ij}\}_{i<j}^t$. [Readers wanting a refresher on logistic regression are referred to Dobson (2002) or Agresti (2002).] In particular, suppose the plans have true (fixed and unobservable) merits $\{\alpha_i\}_{i=1}^t$ and suppose the following link function is assumed for π_{ij}:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_i - \alpha_j \ . \tag{1}$$

It follows from equation (1) that $\pi_{ij} = \exp(\alpha_i - \alpha_j)/\left[1+\exp(\alpha_i - \alpha_j)\right]$ so that if the i-th and j-th plans have the same merit then $\pi_{ij} = 0.5$, and otherwise larger $\alpha_i - \alpha_j$ contrasts imply larger π_{ij} values. It is also clear that it is not the values of $\{\alpha_i\}_{i=1}^t$ that are important, but only their relative differences $\alpha_i - \alpha_j$. In fact, only the differences $\alpha_i - \alpha_j$ are identifiable in this model. The link function (1) may look a little

peculiar for logistic regression contexts, but looks more familiar when written as $\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x'_{ij}\underset{\sim}{\alpha}$,

where $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \dots \alpha_t)'$ and x_{ij} is the $t \times 1$ vector with $+1$ in the i-th position and -1 in the j-th position.

It follows that the likelihood function for $\{\pi_{ij}\}_{i<j}^t$ based on the $\{S_{ij}\}_{i<j}^t$ is

$$L(\pi_{12}, \pi_{13}, \dots \pi_{t-1,t}) \ \propto \ \prod_{i=1}^t \prod_{i<j}^t \pi_{ij}^{S_{ij}} (1-\pi_{ij})^{n_{ij}-S_{ij}} \tag{2}$$

where constants of proportionality have been neglected. Using equation (1), an equivalent representation of the likelihood in terms of $\{\alpha_i\}_{i=1}^t$ is

$$L(\alpha_1, \alpha_2, \dots \alpha_t) \ \propto \ \prod_{i=1}^t \prod_{i<j}^t \frac{e^{\alpha_i S_{ij}} e^{\alpha_j (n_{ij}-S_{ij})}}{(e^{\alpha_i} + e^{\alpha_j})^{n_{ij}}} \ . \tag{3}$$

Since only the differences $\alpha_i - \alpha_j$ are identifiable, an identifiability constraint is required when seeking the maximizing values $\{\hat{\alpha}_i\}_{i=1}^t$ from equation (3). The constraint $\hat{\alpha}_1 = 0$ is used by the R package, $\hat{\alpha}_t = 0$

is used by SAS and $\sum_{i=1}^{t} \hat{\alpha}_i = 1$ is used in some of the Bradley-Terry model literature. Strauss (1992) discusses how standard logistic regression software packages can be used to compute $\{\hat{\alpha}_i\}_{i=1}^{t}$.

Once the $\{\hat{\alpha}_i\}_{i=1}^{t}$ have been obtained, maximum likelihood (ML) estimates of $\{\pi_{ij}\}_{i<j}^{t}$ are obtained as $\hat{\pi}_{ij} = \exp(\hat{\alpha}_i - \hat{\alpha}_j)/\left[1 + \exp(\hat{\alpha}_i - \hat{\alpha}_j)\right]$. The scores used for ranking the alternative plans are $\{\hat{\kappa}_i\}_{i=1}^{t}$, where $\hat{\kappa}_i = (\sum_{j \neq i}^{t} \hat{\pi}_{ij})/(t-1)$. The ranking scores are ML estimates of the average preference probability of each plan when it is pair-wise compared to the other $t-1$ plans.

The R code provided in Appendix B was used with the data $\{S_{ij}\}_{i<j}^{5}$ shown in Table 1 to arrive at the $\{\hat{\alpha}_i\}_{i=1}^{5}$ values for the client data set. Table 2 shows these values along with the estimated ranking scores $\{\hat{\kappa}_i\}_{i=1}^{5}$ that are derived from the estimated pair-wise preference probabilities $\{\hat{\pi}_{ij}\}_{i<j}^{5}$ that are shown in Table 3. The values $\{\hat{\kappa}_i\}_{i=1}^{5}$ are simply the row means of Table 3. A preliminary conclusion, based on the ranking scores, is that it appears Plan 5 is the most favored and that Plan 2 is the least favored.

Plan	$\hat{\alpha}_i$	$\hat{\kappa}_i$	Ranking
1	0	0.53	2
2	-1.437	0.19	5
3	-0.731	0.36	4
4	-0.213	0.48	3
5	2.302	0.94	1

Table 2. Ranking Analysis of Alternative Plans

	Plan j				
Plan i	1	2	3	4	5
1	-	0.81	0.67	0.55	0.09
2	0.19	-	0.33	0.23	0.02
3	0.33	0.67	-	0.37	0.05
4	0.45	0.77	0.63	-	0.07
5	0.91	0.98	0.95	0.93	-

Table 3. Estimated Pair-wise Preference Probabilities $\hat{\pi}_{ij}$

2.2 Educational Opportunities

The client expressed a preference for using R since it is freeware, and through internet searching the CRA identified the R function BTm (Appendix B) to facilitate the ranking analysis shown in Table 2. Understanding how to use BTm in connection with the notation and parameterization presented in Bradley and Terry (1952) was not a trivial task, and provided the CRA with an appreciation for how to link theory to packaged statistical software.

By this time the 2005 fall quarter had begun and a new CC was available to participate in the work. The CRA prepared a draft of slides that summarized the ML analysis and, in preparation for a client meeting, presented them to the CC for peer review. Through this experience, the both the CRA and CC learned the importance of practicing presentations while preparing for a client meeting. In particular, the students learned how to assemble material that would answer client questions and also teach the client about statistical methods relevant to their problem. Especially in academic consulting environments, teaching is a strong element of the client-consultant relationship.

The students in the CC were asked to write their own Newton-Raphson algorithm in R to verify the ML analysis carried out by the CRA. The primary purposes of this assignment were to have the students confront the issue that only the differences $\alpha_i - \alpha_j$ are identifiable in the model, and to show them more clearly the necessity of a constraint such as $\hat{\alpha}_1 = 0$ on the solution to the likelihood equations. In addition, the assignment reviewed a fundamental numerical optimization technique, and asked the students to think through the details of implementing the technique in a programming language. For a few students, this was their first experience with practical details of implementing an optimization technique. The assignment also asked the students to check and compare the computational results across two software packages (R and SAS).

The literature associated with the Bradley-Terry model frequently expresses the likelihood shown in equation (3) in terms of quantities $\{r_{ijk}\}$, where r_{ijk} is the rank (1 or 2, with 1 corresponding to "preferred") of the i-th plan when compared to the j-th plan by the k-th judge. Exercise 1 in Appendix A was used to guide the student through a translation that connects equation (3) to the classic notation used for the Bradley-Terry model.

3. Hypothesis Testing

3.1 Consulting Application

The logistic regression model that uses the Bradley-Terry link function is a reduction of a saturated model that has a separate binomial parameter for each of the C_2^t panels. The likelihood function for the saturated model would simply be equation (2) without the assumed link function given by equation (1). The ML estimates of the saturated model are easily seen to be $\tilde{\pi}_{ij} = S_{ij}/n_{ij}$. A goodness-of-fit test [see, for example, Dobson (2002) or Agresti (2002)] can be made using the statistic $X^2 = -2\log\left[L(\{\hat{\pi}_{ij}\}_{i<j}^t)/L(\{\tilde{\pi}_{ij}\}_{i<j}^t)\right]$. Under the null hypothesis that the logistic regression model with the Bradley-Terry link function is an adequate reduced model, X^2 follows a chi-square distribution with $v = (t-1)(t-2)/2$ degrees of freedom. For the Organization X data, it can be shown that $\log L(\{\hat{\pi}_{ij}\}_{i<j}^5) = -133.84$ and $\log L(\{\tilde{\pi}_{ij}\}_{i<j}^5) = -128.65$ and therefore $X^2 = 10.39$. With $t = 5$ the null degrees of freedom for the chi-square distribution are $v = 6$, and hence the p-value for model adequacy is 0.11, suggesting the reduced model offers an adequate fit.

A visual way to illustrate the adequacy of the reduced model is to compare the observed and expected cell frequencies corresponding to Table 1. The numbers in Table 1 that are in parentheses are the expected cell frequencies $n_{ij}\hat{\pi}_{ij}$ according to the fitted model, and the model adequacy is again reflected by the closeness of the observed and expected cell frequencies.

The $\{\hat{\kappa}_i\}_{i=1}^5$ scores (shown in Table 2) provide a ranking of the alternative plans, but do not by themselves give an indication as to which of the differences $\alpha_i - \alpha_j$ are non-zero. Holm's (1979) sequential Bonferroni (SB) procedure was used to determine which of the estimated differences, $\hat{\alpha}_i - \hat{\alpha}_j$, are significantly different from zero in a statistical sense. Table 4 shows the ML estimates of each contrast, their asymptotic standard errors, the z-scores for the hypotheses $H_0(i, j)$, and the corresponding unadjusted and SB adjusted p-values. Significance levels of 5% (*) and 1% (**) are also indicated in the table.

Contrast	ML Est.	Std. Error	z-score	Unadjusted p-value	Sequential Bonferroni Adjusted p-value
$\alpha_1 - \alpha_2$	1.437	0.294	4.88	1.06×10^{-6}	6.39×10^{-6} **
$\alpha_1 - \alpha_3$	0.731	0.269	2.71	6.77×10^{-3}	2.71×10^{-2} *
$\alpha_1 - \alpha_4$	0.213	0.264	0.81	4.20×10^{-1}	4.20×10^{-1}
$\alpha_1 - \alpha_5$	-2.302	0.416	5.53	3.12×10^{-8}	2.18×10^{-7} **
$\alpha_2 - \alpha_3$	-0.706	0.280	2.52	1.18×10^{-2}	3.53×10^{-2} *
$\alpha_2 - \alpha_4$	-1.223	0.288	4.24	2.19×10^{-5}	1.10×10^{-4} **
$\alpha_2 - \alpha_5$	-3.739	0.450	8.31	9.62×10^{-17}	9.62×10^{-16} **
$\alpha_3 - \alpha_4$	-0.517	0.267	1.94	5.25×10^{-2}	1.05×10^{-1}
$\alpha_3 - \alpha_5$	-3.033	0.431	7.03	2.07×10^{-12}	1.86×10^{-11} **
$\alpha_4 - \alpha_5$	-2.516	0.420	5.98	2.17×10^{-9}	1.74×10^{-8} **

Table 4. Sequential Bonferroni Multiple Comparison Procedure

It can be seen from Table 4 that the only contrasts that are not significantly different at the 5% level are $\alpha_1 - \alpha_4$ and $\alpha_3 - \alpha_4$. Figure 1 shows the grouping of the plans based on the 5% significance level, with the usual interpretation that plans that are connected by a line are not significantly different.
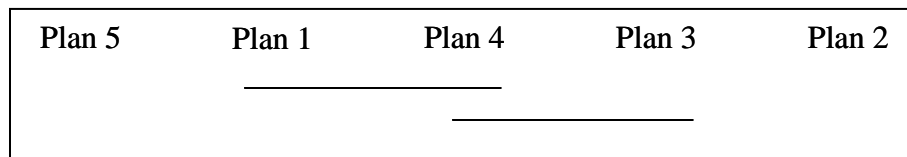

Figure 1. Multiple Comparison Groupings of Plans (5% Significance Level)

3.2 Educational Opportunities

A detailed discussion of the goodness-of-fit test for generalized linear models, with particular emphasis on how it applies to the consulting problem, was provided in the CC. Exercise 2 in Appendix A was assigned to the students to ensure they understand how to compute the degrees of freedom associated with the null distribution of X^2 and interpret the results. The time spent discussing the goodness-of-fit test set a good example for the students of how consultants need to pay attention to the adequacy of models presented to their clients, as they are typically the only ones in a position to make such evaluations.

The sequential Bonferroni procedure was not the first method considered for doing the multiple comparison test of the pair-wise contrasts. Instead, an alternative method was developed in the CC based

on the fact that under the null hypothesis $H_0: \alpha_1 = \ldots = \alpha_t$ the $\{S_{ij}\}_{i<j}^t$ are independently distributed binomial random variables with parameters $(n_{ij}, 0.5)$. Hence, the null distribution of $Q = \underset{1 \le i < j \le t}{Max} |\hat\alpha_i - \hat\alpha_j|$ can be determined to an arbitrary precision via Monte Carlo simulation as it depends only on the sample sizes $\{n_{ij}\}_{i<j}^t$. Two plans would be declared different if and only if $|\hat\alpha_i - \hat\alpha_j| > q_\alpha$, where q_α denotes the upper $\alpha-$percentile of the null distribution of Q. Students in the CC were asked to develop a simulation algorithm to verify that for the client data set $q_{.05} \Box 0.651$. The motivation for this exercise was to reinforce the role and usefulness of simulation studies when solving applied problems.

The multiple comparison procedure based upon Q has the property that the probability of at least one false positive under $H_0: \alpha_1 = \ldots = \alpha_t$ is *exactly* .05, and as such would seem to offer something stronger than other conservative procedures. However, it turns out that while this method exhibits weak control of the Type-1 family-wise error rate, it does not exhibit strong control. The distinction between weak and strong control for a multiple comparison procedure [see, for example, Westfall and Young (1993), Romano and Wolf (2005)] is very important, but is not well known. The consulting problem provided a natural context to expose the concept in a lucid and accessible manner. Exercise 3 in Appendix A guided the students through this learning process.

4. Experimental Design Comparison

4.1 Consulting Application

The experimental design used by Organization X was balanced in the sense that all 10 pairs of plans were evaluated by a panel of judges. Organization X expressed an interest in knowing if there was a viable alternative to running a panel for each of the pairs of plans, while at the same time still being able to rank the plans and assess significance. The CRA suggested an alternative "cyclic" design that employs only four panels comparing the following plan pairs: $(1,2)$, $(2,3)$, $(3,4)$ and $(4,5)$. Within the environment of Organization X, the cyclic design would be significantly simpler to manage. If the Bradley-Terry link function is assumed to hold for all 10 pairs of plans, then all the contrasts $\alpha_i - \alpha_j$ ($i \ne j$) remain estimable with the cyclic design.

One way to compare the balanced and cyclic designs is to evaluate and compare the power of the likelihood ratio test statistic of $H_0: \alpha_1 = \cdots = \alpha_5$ under each design. For the balanced design, the full likelihood is $L_B(\alpha_1, \alpha_2, \ldots \alpha_5) \propto \prod_{i=1}^{5} \prod_{i<j}^{5} \dfrac{e^{\alpha_i S_{ij}} e^{\alpha_j (n_{ij} - S_{ij})}}{(e^{\alpha_i} + e^{\alpha_j})^{n_{ij}}}$. For the cyclic design, the full likelihood is

$$L_C(\alpha_1, \alpha_2, \ldots \alpha_5) \propto$$

$$\left(\frac{e^{\alpha_1}}{e^{\alpha_1} + e^{\alpha_2}}\right)^{S_{12}} \left(\frac{e^{\alpha_2}}{e^{\alpha_1} + e^{\alpha_2}}\right)^{n_{12} - S_{12}} \left(\frac{e^{\alpha_2}}{e^{\alpha_2} + e^{\alpha_3}}\right)^{S_{23}} \left(\frac{e^{\alpha_3}}{e^{\alpha_2} + e^{\alpha_3}}\right)^{n_{23} - S_{23}}$$

$$\times \left(\frac{e^{\alpha_3}}{e^{\alpha_3} + e^{\alpha_4}}\right)^{S_{34}} \left(\frac{e^{\alpha_4}}{e^{\alpha_3} + e^{\alpha_4}}\right)^{n_{34} - S_{34}} \left(\frac{e^{\alpha_4}}{e^{\alpha_4} + e^{\alpha_5}}\right)^{S_{45}} \left(\frac{e^{\alpha_5}}{e^{\alpha_4} + e^{\alpha_5}}\right)^{n_{45} - S_{45}}$$

The likelihood ratio test (LRT) statistics under the balanced and cyclic designs are $\Lambda_B = -2\log\left[L_B(0) / \underset{\hat\alpha \in \Theta}{Max} L_B(\hat\alpha)\right]$ and $\Lambda_C = -2\log\left[L_C(0) / \underset{\hat\alpha \in \Theta}{Max} L_C(\hat\alpha)\right]$, respectively, where

$\underset{\sim}{0} = (0,0,0,0,0)'$ and $\Theta = \{\hat{\underset{\sim}{\alpha}} : \hat{\alpha}_1 = 0\}$. In both cases, the null distribution of the LRT is approximately chi-square with 4 degrees of freedom.

Power for the balanced design was computed for the case where each of the 10 panels had 30 judges (the approximate panel sizes utilized by Organization X) and the power of the cyclic design was computed for the case where the 4 panels each had 75 judges (i.e., 300 judges for both designs). For a given alternative $(\alpha_1, \ldots, \alpha_5)$, power for the balanced design was computed by: 1) simulating 1000 data sets consisting of observations $\{S_{ij}\}_{i<j}^{5}$ that are independent binomial distributions that have trial parameter equal to 30 and success parameters equal to $\exp(\alpha_i - \alpha_j)/[1 + \exp(\alpha_i - \alpha_j)]$, and 2) computing the fraction of the data sets for which Λ_B is greater than $\chi^2_{4,0.05}$. For the same alternatives, power for the cyclic design was computed by: 1) simulating 1000 data sets consisting of observations $(S_{12}, S_{23}, S_{34}, S_{45})$ that are independent binomial distributions that have trial parameter equal to 75 and respective success parameters equal to $\exp(\alpha_1 - \alpha_2)/[1 + \exp(\alpha_1 - \alpha_2)]$, $\exp(\alpha_2 - \alpha_3)/[1 + \exp(\alpha_2 - \alpha_3)]$, $\exp(\alpha_3 - \alpha_4)/[1 + \exp(\alpha_3 - \alpha_4)]$ and $\exp(\alpha_4 - \alpha_5)/[1 + \exp(\alpha_4 - \alpha_5)]$, and 2) computing the fraction of the data sets for which Λ_C is greater than $\chi^2_{4,.05}$.

Results of the power simulations are shown in Table 5 for 14 different alternatives $(\alpha_1, \ldots, \alpha_5)$. It can be seen that for 12 of the alternatives considered, the balanced design has considerably more power than the cyclic design. For these 12 alternatives, the loss of information by using fewer panels is not compensated for by using more judges in each panel. For alternatives $(1, 1.5, 1, 1, 1)$ and $(1, 1.5, 1, 1.5, 1)$ (rows 8 and 9) the cyclic design has higher power. Higher power for the cyclic design in these two cases occurs because a higher proportion of the non-zero contrasts are cyclic and hence the larger panel sizes are able to wield a bigger impact. In the latter alternative, for example, four of the six non-zero contrasts are cyclic. Unfortunately for the cyclic design, its advantage for specific alternatives cannot be exploited in the absence of a-priori information about $(\alpha_1, \ldots, \alpha_5)$.

Alternative	Power	
$(\alpha_1, \ldots, \alpha_5)$	Balanced Design	Cyclic Design
$(1.0, 1.0, 1.0, 1.0, 1.0)$	.06	.05
$(1.0, 1.0, 1.0, 1.0, 1.2)$	.11	.08
$(1.0, 1.0, 1.0, 1.2, 1.2)$	.18	.10
$(1.0, 1.0, 1.2, 1.2, 1.2)$	.18	.078
$(1.0, 1.0, 1.0, 1.0, 1.5)$	.54	.38
$(1.0, 1.0, 1.0, 1.5, 1.5)$	.74	.34
$(1.0, 1.0, 1.5, 1.5, 1.5)$	.78	.38
$(1.0, 1.5, 1.0, 1.0, 1.0)$	.54	.65
$(1.0, 1.5, 1.0, 1.5, 1.0)$	.75	.95
$(1.0, 1.0, 1.0, 1.0, 1.8)$	.95	.77
$(1.0, 1.0, 1.0, 1.8, 1.8)$	.99	.75
$(1.0, 1.0, 1.8, 1.8, 1.8)$	.99	.80
$(1.0, 1.2, 1.2, 1.2, 1.2)$	.12	.074
$(1.0, 1.5, 1.5, 1.5, 1.5)$	.55	.37

Table 5. Monte-Carlo Power of Balanced and Cyclic Designs for Various Alternatives

The precision of contrast estimates can also be used to assess the sensitivity of competing designs. For $\underset{\sim}{\alpha} = (1,1,1,1,1)'$, the standard errors of all 10 pair-wise contrasts $\hat{\alpha}_i - \hat{\alpha}_j$ $(i < j \leq 5)$ were estimated under both the balanced and cyclic designs using an additional simulation study. Table 6 shows estimated standard errors from 1000 simulated data sets. As might be expected, in the balanced design all contrasts exhibit the same standard error while the same is not true for the cyclic design. In the cyclic design, the standard error of a contrast depends on how many panels have to be utilized in order to estimate the contrast. For example, contrasts comparing the plans (1,2), (2,3), (3,4) or (4,5) are estimated with the highest precision since these contrasts are directly estimable from the panels that were run. Contrasts comparing (1,3), (2,4) and (3,5) have less precision because they require utilizing two of the panels that were run. For example, the contrast estimate $\hat{\alpha}_1 - \hat{\alpha}_3$ can be viewed as $(\hat{\alpha}_1 - \hat{\alpha}_2) + (\hat{\alpha}_2 - \hat{\alpha}_3)$, where the two terms come from panels that were actually run in the cyclic design. Similarly, contrasts comparing (1,4) and (2,5) require utilizing three of the panels that were run and the contrast comparing (1,5) requires utilizing all four of the panels that were run. The fact that the cyclic design does not estimate all contrasts equally complicates how a practitioner would go about assigning labels to the plans.

| | Estimated Standard Errors | |
Contrast	Balanced Design	Cyclic Design
$\alpha_1 - \alpha_2$	.23	.24
$\alpha_1 - \alpha_3$	.23	.33
$\alpha_1 - \alpha_4$	.24	.42
$\alpha_1 - \alpha_5$	.24	.48
$\alpha_2 - \alpha_3$	.23	.24
$\alpha_2 - \alpha_4$	.23	.34
$\alpha_2 - \alpha_5$	.24	.41
$\alpha_3 - \alpha_4$	.24	.24
$\alpha_3 - \alpha_5$	.24	.33
$\alpha_4 - \alpha_5$	.24	.24

Table 6. Monte-Carlo Standard Error Estimates for the Pair-Wise
Contrasts under the Alternative $(1.0, 1.0, 1.0, 1.0, 1.0)$

4.2 Educational Opportunities

A particularly nice aspect of this case study is that the client was interested in receiving advice on how to design future experiments. This was ideal for demonstrating to the CC that statistical consulting involves not only data analysis, but also experimental design as well. The cyclic design is a minimal design in the sense there is no design with fewer panels that can still estimate all 10 pair-wise contrasts. Exercise 4 in Appendix A asks the student to derive the ML estimates for all of the contrasts under the cyclic design (closed form expressions exist). The student is also asked to examine the consequences of the minimal nature of the design with respect to the goodness-of-fit of the Bradley-Terry link function.

Power and precision were introduced as criteria to compare the balanced and cyclic designs, and it was pointed out how a fair comparison between the two should have the same number of judges utilized for each case. The Director proposed the set of alternatives with consideration to their implied values for the

$\{\pi_{ij}\}_{i<j}^5$. Students in the CC conducted the power comparison by individually taking one of the alternatives $(\alpha_1, \dots, \alpha_5)$ and developing their own R program to obtain Monte Carlo estimates of the power for each design. The experience the students gained while working on the power study further reinforced their programming and simulation skills.

The CRA's presentation at a client meeting comparing the balanced and cyclic designs was very well received. In fact, the client ranked it as one of the most insightful aspects of the entire project, as it quantitatively justified the balanced design. Additionally, the power and precision metrics were shown to be a useful way to compare alternative designs if and when balanced designs become uneconomical (e.g., when t is large enough to make C_2^t unmanageable).

5. Discussion

5.1 Triangulation Analyses

The role of triangulation analyses (i.e., using two or more methods to verify results) in statistical consulting cannot be emphasized enough. The case study provided an opportunity to show the CC how to be creative in checking the validity of statistical analyses. In particular, a linear model approximation was developed in order to cross-check the ML analysis results, as well as the simulated power and precision estimates. The linear model approximation facilitates a nice link between the consulting problem and statistical methods that students should be very familiar with, and as such opened the door for a number of interesting homework assignments (see Exercises 5-8 in Appendix A).

5.2 Project Management Skills

Throughout the the project, the CRA was responsible for preparing slide presentations that were shared with the client at multiple client meetings. While the Director and/or Associate Director were always present at client meetings, the CRA gave the presentation and always had the first opportunity to answer client questions. This was valuable experience for the CRA, as was the process of preparing and rehearsing for the meetings.

Students in the CC typically work on 2-4 different projects at the same time, and the case study presented in this paper is illustrative of how they get involved in a class project. Other types of projects they work on are individual consulting and small group consulting. The motivation for project multiplexing is that it gives the students experience juggling projects and managing competing deadlines within the same course. While 2-4 simultaneous projects may be light by real world standards, it does provide the students a glimpse of what is to come if they were to take a job as a consulting statistician. The workload experience in the CC differs somewhat from the CRA experience, where for a CRA the simultaneous project load is usually capped at two. The rationale for the difference is that CRAs usually "own" client project whereas in the CC the responsibility for some of their projects is shared within a team environment.

5.3 Summary

The case study described in this paper illustrates how the Statistical Consulting Collaboratory at UC-Riverside not only functions to solve client problems, but also significantly enhances the ability to teach students statistical consulting skills. The exercises in Appendix A are illustrative of the intentional effort made in the CC to go beyond a pragmatic solution to the consulting project for the client and extract from it additional enriching technical material for the students.

The work tasks associated with the case study enhanced the training of students in three different quarters of the CC, and in addition provided a unique set of experiences for the CRA. Table 7 provides a timeline summary for the major activities. It can be seen that the technical part of the work all transpired over a 12-month period. While 12 months may seem a like a long time, the client understood the primary mission of the Collaboratory and was satisfied with incremental progress reports on the various facets of the analyses. Equally important, the timeframe of the project was also dictated by the client's own pace for being available to receive, digest and provide feedback on the reported progress.

Academic Quarter	CC Involved?	CRA Involved?	Principal Activities
Winter 2005	Yes	Yes	Proposal writing and submission
Spring 2005	No	No	Proposal reviewed by Organization X and project cost was negotiated. Proposal eventually approved.
Summer 2005	No	Yes	ML analysis of Bradley-Terry model with R program. Multiple client meetings with Organization X.
Fall 2005	Yes	Yes	Goodness-of-fit test, analysis of cyclic design, power and precision study, linear model analyses. Client meeting with Organization X.
Winter 2006	No	No	Organization X reviews results and uses R code to analyze new data sets on their own.
Spring 2006	Yes	No	Multiple comparisons analyses. Weak vs. Strong control for multiple comparison methods. Final client meeting with Organization X. Client is billed for the work.

Table 7. Timeline and Principal Activities Associated with Consulting Project

The case study that was presented here is a favorite example of projects handled by the Collaboratory due to the interest level that the Bradley-Terry model elicited from the students. The benefits to the CCs and CRAs reported here are based on first-hand experiences, as the last two authors are former students of the CC and former CRAs as well. The popularity of this project was substantiated by student course evaluations and an increase in the number of students wanting to participate in the Collaboratory as a CRA.

A number of other Collaboratory projects have similarly been amenable to the process of merging the consulting aspects of a Collaboratory project with the educational objectives of the CC. Examples include the application and development of a change-point algorithm for tracking reliability metrics in a data network, the use of Classification and Regression Tree Modeling (CART) methodology (and software) to predict success or failure in freshman chemistry classes, and the use of partial least squares analyses for performing chemical spectroscopy analysis. Not every project that comes to the Collaboratory can be integrated into the CC the way our case study was. For example, projects with very short timelines may need a more direct and efficient effort. However, many projects that cannot be worked with CC involvement in "real time" can still have one or more of their aspects incorporated retrospectively at a later date.

APPENDIX A – Synergistic Exercises

1. (*Bradley-Terry Model*) Literature associated with the Bradley-Terry model frequently expresses the likelihood shown in equation (3) in terms of quantities $\{r_{ijk}\}$, where r_{ijk} is the rank (1 or 2, with 1 corresponding to "preferred") of the *i*-th plan when compared to the *j*-th plan by judge *k*.

   a. Show $r_{ij\cdot} = S_{ij} + 2(n_{ij} - S_{ij})$ and infer from this that $S_{ij} = 2n_{ij} - r_{ij\cdot}$ and $n_{ij} - S_{ij} = r_{ij\cdot} - n_{ij}$.

   b. Show $r_{ij\cdot} + r_{ji\cdot} = 3n_{ij}$ and infer that $n_{ij} - S_{ij} = 2n_{ij} - r_{ji}$.

   c. Use (a) and (b) to show that the likelihood in equation (3) can be written as

   $$L(\alpha_1, \alpha_2, \ldots \alpha_t) \propto \prod_{i=1}^{t} \prod_{i<j}^{t} \frac{e^{\alpha_i(2n_{ij}-r_{ij\cdot})} e^{\alpha_j(2n_{ij}-r_{ji\cdot})}}{(e^{\alpha_i} + e^{\alpha_j})^{n_{ij}}}.$$

2. (*Goodness-of-Fit Test*) Let $\{S_{ij}\}_{i<j}^{t}$ be independently distributed as binomial random variables with parameters (n_{ij}, π_{ij}). Show that if the Bradley-Terry link function given by equation (1) is used for $\{\pi_{ij}\}_{i<j}$, then the null distribution of the goodness-of-fit test statistic X^2 will have $(t-1)(t-2)/2$ degrees of freedom.

3. (*Weak vs. Strong Control*) Consider the set of hypotheses $\left\{H_0(i,j): \alpha_i = \alpha_j\right\}_{i<j\leq t}$ and the single-step Q-method that rejects $H_0(i,j)$ if and only if the event $E(i,j): \left|\hat{\alpha}_i - \hat{\alpha}_j\right| > q_\gamma$ occurs (refer to Section 3.2). It follows from the derivation of the Q-method that

   $$Pr\left\{\bigcup_{i<j} E(i,j) \;\Big|\; \bigcap_{i<j} H_0(i,j)\right\} \leq \gamma,$$ and this property is sufficient to say the Q-method has weak control of family-wise error rate (FWER). Now let K be an arbitrary subset of (i,j) pairs corresponding to specific $H_0(i,j)$ hypotheses. The Q-method is said to have strong control of FWER if and only if $Pr\left\{\bigcup_{(i,j)\in K} E(i,j) \;\Big|\; \bigcap_{(i,j)\in K} H_0(i,j)\right\} \leq \gamma$ for *any* subset K. Show the Q-method does *not* have strong control of FWER by demonstrating a counter-example for the case $t = 5$, $\{n_{ij} = 30\}_{i<j\leq 5}$ and $\gamma = 0.05$ using the following simulation experiment:

   a. Let the true state of nature to be $\alpha = (1,1,1,1,1)$, implying all of the hypotheses $\left\{H_0(i,j): \alpha_i = \alpha_j\right\}_{i<j\leq t}$ are true.

   b. Simulate $\{S_{ij}\}_{i<j\leq 5}$ from independent binomial distributions with parameters.

   c. Compute $Q = \underset{i<j\leq 5}{Max} |\hat{\alpha}_i - \hat{\alpha}_j|$

   d. Repeat (b) and (c) 10,000 times to show that the upper 5th percentile of the null distribution of Q is $q_{05} \approx 0.661$. (Note the Organization X design, three panels had $n_{ij} = 29$ which resulted in the value $q_{05} \approx 0.651$ that was reported in Section 3.2).

   e. Now let the true state of nature be $\alpha = (1,1,4,4,4)$. Show that this state of nature implies $H_0(1,2)$, $H_0(3,4)$, $H_0(3,5)$ and $H_0(4,5)$ are the only hypotheses amongst $\left\{H_0(i,j): \alpha_i = \alpha_j\right\}_{i<j\leq 5}$ that are true.

f. Compute $\pi_{ij} = e^{\alpha_i}/(e^{\alpha_i} + e^{\alpha_j})$ and simulate $\{S_{ij}\}_{i<j\le5}$ from independent binomial distributions with parameters $(30, \pi_{ij})$.

g. Compute ML estimates $\{\hat{\alpha}_i\}_{i=1}^5$ and identify if any of the events $E(1,2)$, $E(3,4)$, $E(3,5)$ and $E(4,5)$ occur.

h. Repeat steps (f) and (g) 1,000 times and verify that the fraction of cases where at least one of the events $E(1,2)$, $E(3,4)$, $E(3,5)$ and $E(4,5)$ occurred is about 0.12.

i. Explain why the result from (h) shows the Q-method does not exhibit strong control.

4. (*Estimation in Cyclic Design*) Let S_{12}, S_{23}, S_{34}, and S_{45} be independently distributed as binomial random variables with parameters (n_{ij}, π_{ij}). Suppose the Bradley-Terry link function given by equation (1) is used for $\{\pi_{ij}\}_{i<j}^5$. Impose the constraint $\hat{\alpha}_1 = 0$ on the solution to the likelihood and equations.

   a. Define $\tilde{\pi}_{ij} = S_{ij}/n_{ij}$ for $(i,j) \in \{(1,2),(2,3),(3,4),(4,5)\}$. Show that the remaining part of the solution to the likelihood equations is given by $\hat{\alpha}_2 = \log\{\tilde{\pi}_{12}/(1-\tilde{\pi}_{12})\}$, $\hat{\alpha}_3 = \log\{\tilde{\pi}_{12}\tilde{\pi}_{23}/(1-\tilde{\pi}_{12})(1-\tilde{\pi}_{23})\}$, $\hat{\alpha}_4 = \log\{\tilde{\pi}_{12}\tilde{\pi}_{23}\tilde{\pi}_{34}/(1-\tilde{\pi}_{12})(1-\tilde{\pi}_{23})(1-\tilde{\pi}_{34})\}$ and $\hat{\alpha}_5 = \log\{\tilde{\pi}_{12}\tilde{\pi}_{23}\tilde{\pi}_{34}\tilde{\pi}_{45}/(1-\tilde{\pi}_{12})(1-\tilde{\pi}_{23})(1-\tilde{\pi}_{34})(1-\tilde{\pi}_{45})\}$.

   b. Display the ML estimates for all of the contrasts $\alpha_i - \alpha_j$ and all of the $\{\pi_{ij}\}_{i<j}^5$.

   c. Is it possible to test the goodness-of-fit of the Bradley-Terry link function with this design? Why or why not?

5. (*Approximate Linear Model*) The independence of the $\{S_{ij}\}_{i<j\le t}$ observations combined with a standard delta method analysis suggests least squares methods could be used as an alternative to ML for the data analysis.

   a. Define $y_{ij} = \log\{\tilde{\pi}_{ij}/(1-\tilde{\pi}_{ij})\}$ and use the first-order delta method to show approximations to the mean and variance of $\alpha_i - \alpha_j$ are $y_{ij} = \log\{\tilde{\pi}_{ij}/(1-\tilde{\pi}_{ij})\}$ and $\sigma_{ij}^2 = [n_{ij}\pi_{ij}(1-\pi_{ij})]^{-1}$, respectively.

   b. Let y denote the $C_2^t \times 1$ vector of $\{y_{ij}\}_{i<j\le t}$ values. Display a $C_2^t \times t$ design matrix, X, and a $C_2^t \times C_2^t$ variance-covariance matrix Σ such that the first-order delta method motivates the approximate linear model $y \square X\alpha + e$, where $\mathrm{Var}(e) = \Sigma$.

6. (*Ordinary Least Squares Analysis*) Continue problem 5 as follows:

   a. Demonstrate that X does not have full column rank and thus $X'X$ is a singular matrix.

   b. Show a generalized inverse of $X'X$ is $(X'X)^- = (1/t)I_t - (1/t^2)J_t$, where I_t is $t \times t$ identity matrix and J_t is a $t \times t$ matrix of ones.

   c. Use the Organization X data in Table 1 to compute the ordinary least squares estimator of α and compare the corresponding ordinary least squares estimates of the contrasts $\alpha_i - \alpha_j$ with the ML estimates of the contrasts shown in Column 2 of Table 4.

7. (*Weighted Least Squares Analysis*) The weighted least squares estimator of α, $\hat{\alpha}_{WLS} = (X'\Sigma X)^- X'\Sigma^{-1}y$, cannot be computed without first estimating Σ.

a. Argue that the $C_2^t \times C_2^t$ diagonal matrix $\hat{\Sigma}$, defined to have diagonal elements equal to $\{\hat{\sigma}_{ij}^2\}_{i<j\le t}$, where $\hat{\sigma}_{ij}^2 = [n_{ij}\tilde{\pi}_{ij}(1-\tilde{\pi}_{ij})]^{-1}$, is a reasonable estimator of Σ.

b. Use the Organization X data in Table 1 to compute the two-stage weighted least squares estimator $\hat{\alpha}_{eWLS} = (X'\hat{\Sigma}X)^- X'\hat{\Sigma}^{-1}y$, and compare the corresponding estimates of the contrasts $\alpha_i - \alpha_j$ with the ML estimates of the contrasts shown in Column 2 of Table 4.

c. Compute the estimated standard errors of the two-stage weighted least squares estimates of the contrasts $\alpha_i - \alpha_j$ and compare these values to the standard errors of the ML estimates of the contrasts shown in Column 3 of Table 4.

8. (*Power Approximation*) Using the approximate linear model proposed in problem 5:
   a. Display a $(t-1)\times t$ matrix Λ to write the hypothesis $H_0 : \alpha_1 = \cdots = \alpha_t$ as $H_0 : \Lambda\alpha = 0$.
   b. Motivate $W = (\Lambda\hat{\alpha}_{eWLS})'[\Lambda(X'\hat{\Sigma}^{-1}X)^- \Lambda']^{-1}(\Lambda\hat{\alpha}_{eWLS})$ as a test statistic for H_0 and argue that its null distribution is approximately chi-square with $t-1$ degrees of freedom.
   c. Extend the motivation in (b) to assert the approximate distribution of W under arbitrary alternatives is non-central chi-square with $t-1$ degrees of freedom and non-centrality parameter equal to $\delta = (\Lambda\alpha)'\left[\Lambda(X'\Sigma^{-1}X)^- \Lambda'\right]^{-1}\Lambda\alpha$.
   d. Take $t=5$ and use (c) to repeat the construction of Table 5, but this time using the test of H_0 based on W. Comment on the power of the test based on W relative to the power of the LRT.

APPENDIX B

The R code (Version 2.1.1) shown below provides the ML estimates and the variance-covariance matrix of $\{\hat{\alpha}_i\}_{i=1}^5$ based on the Organization X data shown in Table 1. The key function in the R program is BTm, which fits Bradley-Terry models using the identifiability constraint $\hat{\alpha}_1 = 0$. Lines 4-8 of the R program create a data matrix of the $\{S_{ij}\}_{i\ne j}^5$ values. The R structure 'plan.data', created in line 9, formats the $\{S_{ij}\}_{i\ne j}^5$ according to the requirements of the BTm function. Since $\hat{\alpha}_1 = 0$, the BTm function only returns $\{\hat{\alpha}_i\}_{i=2}^5$ and therefore line 14 is necessary to insert a zero for the first coordinate of $\hat{\alpha}$. Lines 18-19 similarly append a row and column of zeros to the variance-covariance matrix of $\{\hat{\alpha}_i\}_{i=2}^5$, which is returned in Line 17.

The BTm function is contained in a library named 'BradleyTerry' that needs to be invoked as show in line 2. Prior to invoking the 'BradleyTerry' library, two packages need to be downloaded from the local R CRAN (refer to http://www.r-project.org/) and installed into the local R environment. The two packages are the bias-reduced logistic regression (brlr) package and the Bradley-Terry models (BradleyTerry) package. After downloading the .zip files of these packages to a local hard drive, they can be installed from the 'packages' menu in the R window, selecting the option to "Install packages from local zip files." The brlr package should be installed first, and then the BardleyTerry package.

```
1.  # Invoke Bradley-Terry Library
2.      library(BradleyTerry)
3.  # Read in $S_{ij}$ Values
4.      obs=matrix(c(NA,20,22,20,1,
5.                  9,NA,6,7,1,
6.                  8,24,NA,8,2,
7.                  10,23,21,NA,3,
8.                  29,29,27,27,NA),5,5,byrow=T)
9.      plan.data=as.data.frame.table(obs)
10. # Call BTm function
11.     plan.model=BTm(plan.data ~ ..)
12.     output=summary.glm(plan.model)
13. # Retrieve the MLEs
14.     alpha.hat=c(0,output$coeff[,1])
15.     print(alpha.hat)
16. # Retrieve the variance-covariance matrix of MLEs
17.     varcov.0=output$cov.unscaled
18.     varcov.1=rbind(c(0,0,0,0),varcov.0)
19.     varcov=cbind(c(0,0,0,0,0),varcov.1)
20.     print(varcov)
```

Acknowledgement

References

Agresti, A. (2002), Categorical Data Analysis, 2nd edition, New York, Wiley-Interscience.

Birch, J. B. and Morgan, J. P. (2005), "TA Training at Virginia Tech: A Stepwise Progression," *The American Statistician*, Vol. 59, pp. 14-18.

Boen, J. R. and Zahn, D. A. (1982), The Human Side of Statistical Consulting, Lifelong Learning, Belmont, CA.

Bradley, R.A. and Terry, M.E. (1952), "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, Vol. 39, pp. 324-345.

Cabrera, J. and McDougall, A. (2002), Statistical Consulting, Springer, New York.

Carter, R. L, Scheaffer, R. L. and Marks, R. G. (1986), "The Role of Consulting Units in Statistics Departments, *The American Statistician*, Vol. 40, pp. 260-264.

Derr, J. (2000), Statistical Consulting: A Guide to Effective Communication, Duxbury, Pacific Grove, CA.

Dobson, A. J. (2002), An Introduction to Generalized Linear Models, 2nd Edition, Chapman and Hall, Boca Raton, Florida.

Does, R. J. M. M. and Zempleni, A. (2001), "Establishing a Statistical Consulting Unit at Universities," *Kwantitatieve Methoden*, Vol. 67, pp. 51-63.

Hertzberg, V. S., Clark, W. S., and Brogan, D. J. (2000), "Developing Pedagogical and Communications Skills in Graduate Students: The Emory University Biostatistics TATTO Program, *Journal of Statistics Education*, Vol. 8, No. 3.

Holm, S. (1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, Vol. 6, pp. 65-70.

Johnson, H. D. and Warner, D. A. (2004), "Factors Relating to the Degree to Which Statistical Consulting Clients Deem Their Consulting Experiences to be a Success," *The American Statistician*, Vol. 58, pp. 280-286.

Kirk, R. E. (1991), "Statistical Consulting in a University: Dealing With People and Other Challenges," *The American Statistician*, Vol. 45, pp. 28-33.

Mendenhall, W., Beaver, R. J. and Beaver B. M. (2006), Introduction to Probability and Statistics, Thompson/Brooks/Cole, Belmont, CA.

Romano, J. P. and Wolf, M. (2005), "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing," *Journal of the American Statistical Association*, Vol. 100, pp. 94-108.

Russell, K. G. (2001), "The Teaching of Statistical Consulting," in Probability, Statistics and Seismology: A Festschrift for David Vere-Jones, pp. 20-26, edited by D. J. Dayley, Applied Probability Trust, Sheffield, UK

Strauss, D. (1992), "The Many Faces of Logistic Regression," *The American Statistician*, Vol. 46, pp. 321-327.

Taplin, R. H. (2003), "Teaching Statistical Consulting Before Statistical Methodology," *Australian and New Zealand Journal of Statistics*, Vol. 45, pp. 141-152.

Tweedie, R. (1998), "Consulting: Real Problems, Real Interactions, Real Outcomes," *Statistical Science*, Vol. 13, pp. 1-29.

Westfall, P. H. and Young, S. S. (1993), Resampling-Based Multiple Testing, John Wiley & Sons, New York, New York.

Daniel R. Jeske
Department of Statistics
University of California
Riverside, CA 92521
U.S.A.
daniel.jeske@ucr.edu

Scott M. Lesch
Department of Statistics
University of California
Riverside, CA 92521
U.S.A.
slesch@ussl.ars.usda.gov

Hongjie Deng
Department of Statistics
University of California
Riverside, CA 92521
U.S.A.
hdeng001@student.ucr.edu