Exercise 3: Power for a Multilevel Study

3.1 Study name

A multilevel study with a hypothesis test of a between-independent sampling unit factor.

3.2 Study vignette

The study described in this homework exercise is a strongly modified version of the one described in Piquette et al. (2014).

Researchers plan to conduct a randomized controlled clinical trial of an intervention designed to help young children learn fundamental early literacy skills. The intervention is named ABRACADABRA, which is an acronym for "A Balanced Reading Approach for Canadians." Alternative approaches include two standardized Canadian training options for teaching literacy. The first is an English Language Arts program, and the second is a bi-lingual program for native English speakers designed to build literacy in both French and English.

Researchers plan to randomize 45 schools, with 15 in each treatment arm. That is, 15 schools will get ABRACADABRA, 15 schools will get the English Language Arts Program, and 15 schools will get a bi-lingual French/English program.

All of the schools are very far apart from each other. In fact, they are so far apart that the students, parents and teachers in each school have no contact with the students, parents and teachers in any other schools. Thus, schools can be considered to be independent.

Initial study planning will begin by assuming each school has exactly 4 kindergarten classrooms, each with 5 students, all of whom will take part in the trial. Thus, each school has 20 total students, calculated as $20 = (5 \text{ students/classroom}) \times (4 \text{ classrooms/school})$. In most real scenarios, there will be a different number of students in each classroom. There will be a different number of classrooms in each school. There will be a different number of schools in each neighborhood. For scenarios with different numbers in each cluster, we will discuss how to handle power and sample size analysis later.

The extent that gender, age and ethnicity of the students is related to the outcomes is not going to be considered for the initial study planning.

The intraclass correlation coefficient for classrooms within schools was assumed to be 0.11. The intraclass correlation coefficient for students within classrooms was assumed to be 0.04.

Students will be evaluated on three composite outcome scores that are considered to have a multivariate normal distribution. The three composite scores are letter-sound knowledge (LSK), the Comprehensive Test of Phonological Processing (CTOPP) blending words subtest age equivalent scores, and the Group Reading Assessment and Diagnostic Evaluation (GRADE) listening comprehension subtest stanine score.

Students will be measured before and after the intervention, and for each component, a difference score will be calculated as $\text{post}-\text{pre}$. The outcomes of interest are the difference scores for the three composite scores. Each student will contribute three difference scores: one for LSK, one for CTOPP and one for GRADE. The common standard deviation (across the three treatment arms) for GRADE is 4.4, for LSK is 4.2, and for CTOPP is 0.6.

Exhibit 1: Predicted mean difference scores for three literacy scale scores, stratified by treatment arm.			
	GRADE	LSK	CTOPP
ABRA	0.3	0.3	0.3
English Language Arts	0.1	0.1	0.1
bi-lingual literacy Program	0.1	0.1	0.1

Initial study planning will begin by assuming no missing data, which corresponds to requiring everyone present for either the pre or post test is present for both. Subsequent refinement of the sample size analysis could include an allowance for missing data, if it is deemed appropriate.

Based on knowledge of previous studies, scientists have a good guess as to what the correlation matrix looks like for the difference scores of the three outcome measures.

Exhibit 2: Correlation Matrix for difference scores			
	GRADE	LSK	CTOPP
GRADE	1	0.9	0.2
LSK	0.9	1	0.4
CTOPP	0.2	0.4	1

The three difference variables define a multivariate response profile. Scientists hypothesize that the three literacy training programs do not differ in any combination of the outcome differences. The scientists wonder what their power will be for the proposed trial.

3.3 Statistical analysis plan

Scientists plan to fit a general linear multivariate model with the 20 difference scores in LSK, CTOPP and GRADE, respectively, observed per school as outcomes. There will be 45 schools contributing data to the model, with 15 assigned to each of three treatment arms. As predictors, the scientists plan to use indicator variables for the three treatments. The scientists plan to test a multivariate analysis of variance (MANOVA) hypothesis. Scientists will use the Hotelling-Lawley trace statistic at a Type I error rate of 0.05 to evaluate the null hypothesis of no differences in response difference profiles among the three literacy programs. The scale factor to be used for means is 1. The scale factor to be used for variability is 1.

References cited

Piquette, Noella A., Robert S. Savage, and Philip C. Abrami. "A Cluster Randomized Control Field Trial of the ABRACADABRA Web-Based Reading Technology: Replication and Extension of Basic Findings." (2014) *Frontiers in Psychology*, **5**. doi:10.3389/fpsyg.2014.01413.