A note to the reader about the final project

Power and sample size analysis requires two skills. The first is the ability to understand a study design, and using the study design and a set of given inputs, to calculate power and sample size. The second skill is the ability to find inputs for power analysis, often from published literature.

In the first four exercises, the goal was to conduct power or sample size analysis using known inputs. Thus, each exercise included all the input information needed to a power or sample size analysis.

For the final exercise, the goal is to allow the student to practice both the skill of extracting input parameters, and the skill of conducting power and sample size analysis. Thus, we provide journal articles containing the sample size analysis inputs, and ask students to find some of the inputs.

****

1. Please answer all of the same questions as you answered for Exercises 1-4.

2. Additional requests (and tips for success)

2.1 In answering **Question 4.13** use the information from Table 1, p. 45 in Bullitt et al. (2005), and fill in the table in Exhibit 1 below.

2.2 In answering **Question 4.13**, report which estimate of the standard deviation you use and explain why you chose it.

2.3 In answering **Question 5.** (Provide the **results of power and sample size analysis**.), please make sure to adjust the power or sample size for dropout and loss-to-follow up. Consider how many research participants need to be assigned to each group, and remember that we can't have fractional study participants. Please give an estimate of how long it will take to recruit the sample. Describe recruitment feasibility.

Exercise 5: Sample Size Analysis for a Planned Subgroup Analysis

5.1 Short study description

A multivariate study with between-independent sampling unit factors.

5.2 Study vignette

The study described is a possible future extension of the study conducted by Bullitt et al. (2005). We have made strong efforts in this vignette to hew as much as possible to the science behind the study, and to include, as much as possible, reasonable values. However, we did create some values in order to set up a reasonable sample size analysis. We tried to indicate where we used real values, and where we used speculated values. In addition, we have made up some details of the proposed studies, such as the four major genotypes of VegF that are considered in the proposed study, and the chemotherapeutic regimen described that affects vessel tortuosity.

Before you read this study vignette, please read the accompanying article by Bullitt et al. (2005) carefully. The goal is not necessarily to understand the science. The goal is to look carefully at the article to find where the authors published means and standard deviations.

For the Muller et al. (2007) article, please look only at Equation 6 on page 3648, which displays a covariance matrix. In the class, we talked about correlation matrices. Recall that a correlation matrix describes the associations between two or more variables. A covariance matrix is like a correlation matrix in that it describes the associations between two or more variables. However, a correlation matrix contains scaled numbers, between -1 and 1. A covariance matrix contains unscaled numbers, still in the scale of the original variables.

In the class, we stressed the importance of publishing correlation or covariance values, or both, for the future use of researchers in the field who are conducting power or sample size analyses. We included the Muller et al. (2007) paper to show that some authors do publish these values. However, we note that the values are not published in the Bullitt et al. (2005) manuscript. Unfortunately, publication of correlation values, or covariance values, or both, is unusual.

In Bullitt et al. (2005) manuscript, the researchers described a measure of vessel tortuosity in the brain. Tortuosity is a measure of how twisted a vessel is. Tumors develop new blood vessels as they grow. Frequently, the new blood vessels have many small bends in them. Measuring tortuosity via MRA can help scientists differentiate between normal tissue and tumor tissue, and measure the effectiveness of treatments at stopping tumor growth and progression.

Bullitt et al. (2005) considered multiple ways to quantify tortuosity. This exercise will concentrate on a single way to quantify tortuosity, called SOAM1. In the Bullitt study, scientists quantified vessel tortuosity using SOAM1 in four regions of the brain. Two of the regions, the left and right middle cerebral groups, had similar summary SOAM1 measurements, as shown in Figure 2, on page 45, and in Table 4, page 46.

In the proposed study, the scientists planned to use SOAM1 as the outcome. They planned to measure this outcome in two regions of the brain for every patient, the left middle cerebral group, and the right middle cerebral group. This means that there are two repeated measurements of SOAM1 for each patient.

Using the information from Table 1, page 45 in Bullitt et al. (2005), please fill in the following table for use in your power analysis.

Exhibit 1: Statistics for summary SOAM1 measures by brain regions.		
	Mean	Standard Deviation
Left Middle Cerebral Group		
Right Middle Cerebral Group		

The Bullitt et al. (2005) paper does not provide correlations between measurements of SOAM1 on different brain regions. Converting the covariance matrix that appears in Equation 6 on page 3648 into a correlation matrix yields the correlation matrix shown in the following table.

Exhibit 2: Correlation between summary SOAM1 measures for different brain regions		
	Left Middle Cerebral	Right Middle Cerebral
Left Middle Cerebral Group	1	0.53
Right Middle Cerebral Group	0.53	1

Bullitt et al. (2005) studied the SOAM1 variable, and concluded that the distribution was appropriately normal, or Gaussian. You can look at their results by examining the p-values that appear in Table 1.

Bullitt et al. (2005) had theorized that abnormal tortuosity in vessels was perhaps caused by increases in nitrous oxide induced by VegF. There are four major genotypes of VegF, designated, for convenience, Genotype A, Genotype B, Genotype C and Genotype D.

A new group of investigators was interested in studying the tortuosity response of blood vessels in the left and right middle cerebral groups. The plan was to recruit study participants with glioblastoma multiforme, a brain tumor. It was expected that the researchers would be able to recruit equal numbers of those with Genotypes A, B and C. Because of the frequency of Genotype D in the population, there would be roughly twice as many study participants with Genotype D as with Genotype A.

Exhibit 3: Relative sizes of genotype groups in study population.	
	Size Ratio
Genotype A	1
Genotype B	1
Genotype C	1
Genotype D	2

The researchers planned to randomize the study population either to a placebo, or to a new chemotherapeutic regimen. They wanted to measure vessel tortuosity in the left and right middle cerebral groups as the outcomes of the study. That is, for each study participant, they would have two repeated measurements of SOAM1, one on the left cerebral middle group, and the other on the right cerebral middle group. To ensure equal allocation within each group, they planned a block randomization scheme with study participants randomized one-to-one to treatment or placebo within each genotype group.

The researchers were confident that the response of each individual to treatment would be independent, even if the individuals went to the same clinic. They thought that the responses of the two brain regions for each study participant would be correlated.

The researchers wanted to test to see if there would be an interaction between treatment and genotype on the average response to treatment across the two brain regions. Recall that an interaction hypothesis describes how two factors (here, the two between ISU factors, treatment and genotype) interact to change the response. This is the same as asking whether the effect of treatment differs across the genotypes, where we are measuring the effect of treatment by looking at the average across the two brain regions.

In GLIMMPSE, when one requests the treatment by genotype interaction, GLIMMPSE automatically assumes that one wishes to average across the two brain regions, and calculates the power or sample size in that manner. This is exactly what the researchers wanted to do.

By the way, the variance of an average decreases with the number of observations that one averages. Remember that power goes up as variance goes down. Thus, you will get different answers for the power and sample size calculation if you assume that SOAM1 is measured once, twice, or four times. You can try this out empirically by changing the number of repeated measurements of SOAM1. Because of this, it is important to describe the study as having one outcome variable, with two repeated measurements.

They theorized that the pattern of mean responses would be as shown in the following table.

Exhibit 4: Predicted responses by treatment and genotype.			
Treatment	Genotype	Mean Left Middle Cerebral SOAM1	Mean Right Middle Cerebral SOAM1
Placebo	A	3.12	3.17
Placebo	B	3.12	3.17
Placebo	C	3.12	3.17
Placebo	D	3.12	3.17
Chemotherapy	A	2.1	2.15
Chemotherapy	B	2.3	2.35
Chemotherapy	C	2.5	2.55
Chemotherapy	D	2.7	3.25

Glioblastoma multiforme is an almost uniformly fatal disease with a rapid disease course. The chemotherapeutic regimen involved targeted mRNA, and previous studies had shown a low rate of side effects. The MRA imaging regimen used to measure study outcomes does have a low risk of anaphylaxis due to the use of a gadolinium contrast agent, but, weighing risks and benefits, the Institutional Review Board thought that the risk was acceptable, given the risks of the disease.

After an ethics consultation, the researchers decided that it would be preferable to have the study large enough to test the hypothesis, rather than adopting a conservative approach and limiting sample size to limit exposure to the risks of the study. Thus, they planned to use the larger of the two variance estimates they had. Notice that the researchers had a choice of two standard deviation values, as shown in Exhibit 1. To be conservative, round the standard deviation you choose to one value behind the decimal point.

Roughly 180 patients meeting the eligibility criteria are seen each month by the large glioblastoma clinic at the high volume tertiary care clinic at which the study will be done. Roughly 20% of the patients will consent to the study. The chance of consent is not associated with genotype, nor with response to treatment, nor with baseline tortuosity. The study investigators think that almost 30% of the participants will be lost to follow-up and not complete the study, for reasons independent of disease severity, genotype, or response. From previous experience, the investigators believe that if people complete the study, they are quite likely to be able to measure the vessel tortuosity in both brain regions. It is unlikely that they will be able to measure response in only one region, or in no regions at all.

Feasible sample sizes for the study would be 30, 40 and 50, with the smallest group size, respectively of 3, 4 and 5. The goal is to figure out what the power is for each of the sample sizes, and to choose the smallest sample size such that the power is at least 0.95.

5.3 Statistical analysis plan

We will fit a general linear mixed model. The outcome variables will be the two repeated measurements of SOAM1, one in the left cerebral middle group, and one in the right cerebral middle group. The predictors will be eight indicator variables for the genotype by treatment groups. Each indicator variable will take on the value 1 if the study participant is a member of the specific genotype-by-treatment group, and 0 otherwise. We will use a Wald statistic with Kenward-Roger degrees of freedom (which corresponds to a Hotelling-Lawley test for complete data) to assess the null hypothesis that there is no interaction between genotype and treatment on the average response over the two brain regions. We will use an unstructured covariance matrix, and assume that the variance-covariance matrix of the errors is the same for each person. We will use a Type I error rate of 0.05. The scale factor to be used for means is 1. The scale factor to be used for variability is 1.

This modeling technique assumes equal error variance, independence of the independent sampling units, finite second moments, and linearity, which means that the outcome could be described as a linear function of the predictors. We will use regression diagnostics and jackknifed studentized residuals to examine the assumptions.

References cited

Bullitt, E., Muller, K. E., Jung, I., Lin, W., & Aylward, S. (2005). Analyzing attributes of vessel populations. Medical Image Analysis, 9(1), 39–49. http://doi.org/10.1016/j.media.2004.06.024

Muller, K. E., Edwards, L. J., Simpson, S. L., & Taylor, D. J. (2007). Statistical tests with accurate size and power for balanced linear mixed models. Statistics in Medicine, 26(19), 3639–3660. http://doi.org/10.1002/sim.2827

Muller, K. E., & Stewart, P. W. (2006). Linear Model Theory: Univariate, Multivariate, and Mixed Models (1 edition). Hoboken, N.J: Wiley-Interscience.