

# Biost/Stat 571: Homework # 4

Due 5pm, Thu Mar 9 via Canvas

**Note:** Homework should be submitted in as a PDF document with problems clearly labeled and in order. Use of Latex is preferred, but hand-written math is acceptable. In the case of the latter, handwriting must be in clearly legible print. Illegible (unreadable by the TAs and instructor) and unclear work will not receive credit. Clarity in derivations and exposition count as these are essential in any professional setting.

**Problem 1.** Empirically (via simulation), assess the performance of complete-case and available-case analysis under Mixed Models and GEE under the three distinct situations in which dropouts that are (1) MCAR; (2) MAR and (3) NMAR.

**Problem 2.** For many applied statisticians, we spend much more time designing studies than we do in carrying out analysis. Suppose that a researcher approaches you with the following proposal. Please help them in developing (1) the Statistical Analysis Plan section and (2) the power and sample size section. Note that since the sample size is fixed (often due to cost or availability of subjects), we would want to calculate the minimum detectable effect sizes. You will need to make assumptions, which is fine, but be explicit in noting the assumptions made. Typically, the power analysis also involves picking a primary endpoint.

Here is the study proposal:

**Title:** "The Influence of Taylor Swift on Graduate Students' Academic Performance and Mental Health: A Longitudinal Study"

**Objective:** To investigate the impact of Taylor Swift on the academic performance and mental health of graduate students over a period of two years.

## Summary of Methodology:

Participants: The study will recruit 200 graduate students (100 male and 100 female) from two universities in the United States. Participants will be selected based on their willingness to participate, and they should be Taylor Swift fans.

Procedure: The study will be conducted over two years, with data collection taking place at four time points (baseline, 6 months, 12 months, and 24 months). At baseline, participants will complete a survey that collects information on their demographics, academic performance, and mental health. They will also complete a Taylor Swift fandom questionnaire, which assesses their level of interest in Taylor Swift and their knowledge of her music.

At each time point, participants will complete follow-up surveys that measure their academic performance (e.g., GPA, academic progress) and mental health (e.g., stress, anxiety, depression). They will also complete the Taylor Swift fandom questionnaire at each time point to assess any changes in their level of interest and knowledge of Taylor Swift.

Note that this is very typical of the types of project proposals that you may see in practice. The investigator hasn't specifically identified a particular endpoint and that is often left to the statistician. It is up to you to "simplify" the study and abstract it sufficiently to be able to write a section. There are no "wrong" answers to this aspect as long as you are stating what you're treating as the endpoints.

**Problem 3.** (BONUS) Conduct a simulation study to evaluate the impact of non-normality of random effects.