**Part 1: Power calculations for individual-level randomization**

1. Develop a data generating process for some Y that is normally disturbed around 0 with standard deviation of 1.
2. The average treatment effect should be 0.1 sd (with the effects being uniformly distributed between 0.0 – 0.2 sd)
3. The proportion of individuals receiving treatment should be 0.5 (i.e. half in control, and half in treatment) Calculate the number of individuals required to reach 80% power when you are trying to detect 0.1 sd treatment effect.
4. Now assume, 15% of the sample will attrite (assume similar attrition rates in control and treatment arms.) How does this change your sample size calculations from the previous part?
5. Now assume the intervention is very expensive and we can only afford to provide this specific treatment to 30% of the sample. How would this change the sample size needed for 80% power.

**Part 2: Power calculations for cluster randomization**

1. Develop data generating process for data for Y (assume math score of each individual student) in a school. We can only assign treatment at the school-level.
2. Your function should be able to change the number of clusters (i.e. schools) and the cluster size (i.e. number of students in each school)
3. Make sure the rho/icc is ~ 0.3 when generating these clusters. [Hint](https://www.statalist.org/forums/forum/general-stata-discussion/general/1460617-how-to-simulate-clustered-data-with-a-specific-intra-class-correlation).
4. Divide the schools evenly between treatment and control arms. And generate a treatment effect of 0.2 sd (with the effects being uniformly distributed between 0.15 – 0.25 sd)
5. Holding the number of clusters fixed at 200, what happens to the power when you increase the cluster size (use first 10 powers of 2) What cluster size would you recommend and why?
6. Now hold the cluster size fixed (15 students/school). How many schools do you need in your RCT to get 80% to detect 0.2 sd treatment effect.
7. Now assume that only 70% of the schools actually adopt your treatment. How many schools do you need now to get 80% power?

**Part 3: De-biasing a parameter estimate using controls**

1. Develop some data generating process for outcome Y, with some treatment variable and treatment effect.
2. This DGP should include strata groups and continuous covariates, as well as random noise. Make sure that the strata groups affect the outcome Y. You will want to create the strata groups first, then use a command like expand or merge to add them to an individual-level data set.
3. Make sure that at least one of the continuous covariates also affects both the outcome and the likelihood of receiving treatment (a "confounder"). Make sure that another one of the covariates affects the outcome but not the treatment. Make sure that another one affects the treatment but not the outcome. (What do these do?)
4. Construct at least five different regression models with combinations of these covariates and strata fixed effects. (Type h fvvarlist for information on using fixed effects in regression.) Run these regressions at different sample sizes, using a program like last week. Collect as many regression runs as you think you need for each, and produce figures and tables comparing the biasedness and convergence of the models as N grows. Can you produce a figure showing the mean and variance of beta for different regression models, as a function of N? Can you visually compare these to the "true" parameter value?
5. Fully describe your results in your README file, including figures and tables as appropriate.