It will make your life easier if you save the data files for this assignment in one folder (e.g. C:\important_assigment) . Then start your .do file with the command:

cd "C:\important_assigment"

This changes the default directory to C:\important_assigment. When you want to read in a data file (e.g. GPA3.DTA) you simply type:

use GPA3

1. Use the file GPA3.DTA for this exercise. The data is for 366 student athletes from a large university for the fall and spring semesters. The primary question of interest is: "In the semester of their sporting season, do athletes perform more poorly in school?"

    a. Use pooled OLS to estimate a model with term grade point average (*trmgpa*) as the dependant variable. The explanatory variables are *spring*, *sat**, *hsperc*, *female*, *black*, *white*, *frstsem*, *tothrs*, *crsgpa*, and *season*. Provide an interpretation of the coefficient on *season*. Is it statistically significant?

    b. Most of the athletes who play their sport only in the fall are football players. Suppose the ability levels of football players differ systematically from those of other athletes. If ability is not adequately captured by SAT scores and high school percentile, explain why the pooled OLS will be biased. Can you use the data to exam this hypothesis?

    c. Now use the data differenced across the fall and spring semesters. Which variables drop-out? Test for an in-season effect.

    Hint: use Stata's *tsset* command to assign *id* (a unique number for each student) as the panel variable and *spring* as the time variable. Once this is done, adding *D.* (e.g. *D.trmgpa*) in front of any variable will automatically first difference it.

    d. Can you think of one or more potentially important time-varying variables that have been omitted from the analysis?

    *Variable *sat* provides the students score on their SAT exam. These are standardized tests, taken prior to university application, used by American universities to assess student ability.