Homework #8
RPAD 316
Professor Stephen Holt

**Instructions: You will be doing the problems in this assignment by hand and in Stata. For the problems by hand, show your work for each step. For any Stata questions, when asked for graphs, save the graph as a .png file and paste the image into the appropriate section of the word document. Paste your code that produced the graph below the graph. When asked for output, you can use a screen capture of the Stata output (saved as a .png and pasted into the homework document). Paste the code used to create the output under the screenshot.**

**Dataset 1. NBA team statistics, 2017-2018 season.**

| Eastern Conference | W | L | W/L% | GB | PS/G | PA/G | SRS |
|---|---|---|---|---|---|---|---|
| *Atlantic* | | | | | | | |
| Toronto Raptors* (1) | 59 | 23 | 0.72 | — | 111.7 | 103.9 | 7.29 |
| Boston Celtics* (2) | 55 | 27 | 0.671 | 4 | 104 | 100.4 | 3.23 |
| Philadelphia 76ers* (3) | 52 | 30 | 0.634 | 7 | 109.8 | 105.3 | 4.3 |
| New York Knicks (11) | 29 | 53 | 0.354 | 30 | 104.5 | 108 | -3.53 |
| Brooklyn Nets (12) | 28 | 54 | 0.341 | 31 | 106.6 | 110.3 | -3.67 |
| *Southeast* | | | | | | | |
| Miami Heat* (6) | 44 | 38 | 0.537 | 15 | 103.4 | 102.9 | 0.15 |
| Washington Wizards* (8) | 43 | 39 | 0.524 | 16 | 106.6 | 106 | 0.53 |
| Charlotte Hornets (10) | 36 | 46 | 0.439 | 23 | 108.2 | 108 | 0.07 |
| Orlando Magic (14) | 25 | 57 | 0.305 | 34 | 103.4 | 108.2 | -4.92 |
| Atlanta Hawks (15) | 24 | 58 | 0.293 | 35 | 103.4 | 108.8 | -5.3 |

Note: W = wins, L = losses, W/L% = proportion of games won, GB = Games behind leader, PS/G = Points scored per game, PA/G = Points against per game, SRS = Standardized rank score that accounts for win point differentials and strength of schedule.

Dataset 1 contains a variety of team statistics from the 2017-2018 NBA season, the last season that LeBron James played in the Eastern Conference. Treat each division in the Eastern Conference shown above as a sample to answer the following questions.

1. The PS/G represents a measure of the points scored by each team in an average game. What is the average PS/G for each division?

2. What is the standard deviation of PS/G for each division?

3. What is the standard error of PS/G for each division?

4. What is the 95% confidence interval of the PS/G for each division?

5. Bradley Beal (who played for the Washington Wizards) and Kyle Lowry (who played for the Toronto Raptors) get into a Twitter argument about who had the more difficult season in terms of facing high scoring opponents in their division. Using the PS/G as a measure of team scoring rates, you will settle the argument with a two-tailed t-test using the two samples.
   a. What is the null hypothesis for your test?

   b. What is the alternative hypothesis?

   c. Calculate the t-statistic for the two-sample hypothesis test.

   d. What degrees of freedom do you have for your test?

   e. What is the p-value?

   f. Are the results statistically significant? Who won the argument?

Part 2. Dataset 2 contains data from a randomized control trial run by Nobel Laureates Abhijit Banerjee, Esther Duflo, and co-authors. A long-running concern about social services in developing countries is that the lack of democratic participation and oversight leads to low quality services. One common theoretical proposition is that developing the capacity to independently assess service quality and advocate directly with leaders for improvements. The authors conducted a randomized control trial in rural India, providing villages with 3 different treatments: information about how officials run school programs and who to contact about them, information plus training parents to give a simple reading test at home so they can measure performance of their child themselves, and both other treatments plus a trained volunteer to run an extracurricular reading class.

1. First, using `any_treatment` as the indicator for participation in any of the three treatments relative to the control villages, run a balance test to compare the age, type of school, gender, and the proportion of students in the highest reading level before treatment. HINT: There are

separate indicators for each of the five school types.

2. Policy makers at the World Bank want to know if the treatments are effective at improving student reading levels. Use the indicator variable that a student is at the highest reading level after the treatments were implemented in the treatment villages (variable name: `reading_highest2`). What is the null hypothesis of a two sample test?

3. What is the alternative hypothesis?

4. Conduct a two sample t-test in Stata.