Econometrics assignment 5a

*Joost Bouten, SNR: 1265889  
Twan Vissers, SNR: 1266283  
Fons Strik, SNR: 1257943*

The commands and output of all questions can be found in the log-file included below. (Many lines and thus not compatible to include within some questions)

1. 1. We find that the mean of the main outcome variable is equal to 10.29

Command + output:



* 1. We find there are 65 garbage routes (unique values).

Command + output:



* 1. In the dataset there are 52 calendar weeks of data (unique values).

Command + output:



* 1. Scatter graph of the outcome variable per calendar week.

Command:   
graph twoway (scatter residual\_weight calendar\_week), by( route)

Output:  


* 1. Scatter graph of the outcome variable per calendar week, pre- and post-treatment.

Command:   
graph twoway (scatter residual\_weight calendar\_week if TreatmentOngoing==0&TreatmentCompleted==0) (scatter residual\_weight calendar\_week if TreatmentOngoing==1| TreatmentCompleted==1) if route<200, by( route)

Output:   
If the blue and red scattered dots are compared, we cannot clearly observe a difference. So, we do not clearly see a clear change in the weight of residual waste pre- and post-treatment. Thus, the raw data do not indicate that something is going on as of the date that the announcement letter was sent out.

1. 1. Commands + output:



* 1. See log-file below



1. As we ignored the time variation in the treatment, we are likely to overestimate the treatment effect since the short run treatment effect is likely greater than the long run treatment effect as can be concluded from the graph above. This can be seen from the upward trend of the scatter plot after treatment towards the null hypothesis of no change. When we do not cluster by route we can see that the standard error LetterReceived is greater. The standard errors of other variables slightly change as well. All the coefficients remain the same. For instance, the clustering gives a better representation of the LetterReceived standard errors as not every unit of the sample contributes information. Only every route contributes new information, therefore clustering by routes gives a more precise value for the standard errors.  
   The degree of serial correlation seems to be fairly large for the standard errors of the LetterReceived variable and other variables.
2. To find the number of weeks, we divide the absolute value of the initial treatment effect (LetterReceived coefficient) by the linear decay. We find that the number of weeks until the initial treatment effect has completely disappeared is approximately 37 weeks. This seems in line with the graph shown under III(d).
3. The long term effect is approximately 13.5 percent smaller than the short term effect.

Based on the F test, we are not able to reject the null hypothesis of no difference at conventional levels of statistical significance (1%, 5% and 10%).