ECO 307: Introductory Econometrics **Your Name:** _____
Instructor: Dr. James Murray
In-class Exercise: Regression Analysis
Fall 2015

Learning Objective: (LO 2) Construct, estimate, and interpret regression models to identify relationships between explanatory and outcome variables.

Directions: The questions on the following page use the dataset `jtrain3.RData` on the class website. The dataset includes information relevant to estimate the impact of job training on earnings. Variables include workers' earnings for years before and after job training, years of education, demographic information, and a dummy variable for whether or not they received job training.

Group Work Terms and Conditions: Work in groups of up to four people and answer the following questions. All papers will be collected, but only one member's paper will be randomly selected and graded and all members of the group will receive the same grade.

By signing below, you agree that the following work represents the efforts of everyone in the group, and you are willing to accept as your own grade for the group project the grade earned from this representation of your group's work. Every member must agree to these terms to earn a non-zero grade for this assignment.

_____	_____	_____
Signature Group Member 1	Print Name	Date
_____	_____	_____
Signature Group Member 2	Print Name	Date
_____	_____	_____
Signature Group Member 3	Print Name	Date
_____	_____	_____
Signature Group Member 4	Print Name	Date

Some workers participated in a job training program in 1977. To estimate the effect of the program on subsequent earnings, consider a regression model that predicts earnings in 1978 (`ret78`) (in thousands of dollars) based on the following explanatory variables:

- Earnings in 1974 (in thousands of dollars) (`ret74`)
- Earnings in 1975 (in thousands of dollars) (`ret75`)
- Years of education (`educ`)
- Worker's age (`age`)
- Dummy variable for Black / African American (`black`)
- Whether or not the person participated in the training program (`train=1` for participants)

1. Using ordinary least squares, is there evidence that the training program positively influenced earnings?

2. Plot the residuals (vertical axis) against the predicted values (horizontal axis). Is there evidence of heteroskedasticity? If so, describe the behavior of the variance of the residuals.

3. Is there statistical evidence for heteroskedasticity? Conduct the appropriate hypothesis test.

4. Using heteroskedastic consistent standard errors, is there evidence that the training program positive influenced earnings?

It has been argued that the training program is more effective for poor workers than higher income earners. Create a new dummy variable in the dataset called `poor` that is equal to 1 for workers that had an income less than $10,000 in 1974 (`ret74 < 10`). This can be accomplished with the following command:

```
data$poor <- as.numeric( data$ret74 < 10 )
```

Re-estimate the regression above, interacting the training variable with your new `poor` variable. To use `poor` *only in the interaction term* include the interaction with `train` as `train:poor`. Notice the use of ':' instead of '*'.

5. Is there evidence that the impact of the training on earnings is different for people with income below $10,000 versus people at this level or above. Test the appropriate hypothesis.

6. Is there evidence of heteroskedasticity in this model. Test the appropriate hypothesis.

7. What is the estimated impact from participating in the training on annual earnings in 1978 for a person with income below $10,000?

8. Is there statistical evidence that there is a positive impact of the job training on annual earnings for people with income below $10,000? Test the appropriate hypothesis.

   Hint: This requires a hypothesis test on a linear combination of coefficients

   Another hint: The function `glht()` *can take a parameter* `vcov` *to use the appropriate heteroskedasticity robust standard errors.*