You are encouraged to work with your classmates on this homework. However, you each must turn in your own answers. I will not accept identical answers from multiple students nor do I want to see the identical answers from the back of the text.

Part I – Questions from the text

1. Chapter 6, number 2
2. Chapter 6, number 3
3. Chapter 6, number 11
4. Chapter 6, number 12
5. Chapter 7, number 4
6. Chapter 7, number 7
7. Chapter 7, number 15
8. Chapter 7, number 16
9. Chapter 8, number 6
10. Chapter 8, number 11
11. Chapter 9, number 11
12. Chapter 9, number 13
13. Chapter 9, number 16
14. Chapter 10, number 8
15. Chapter 13, number 2
16. Chapter 13, number 5
17. Chapter 13, number 6
18. Chapter 13, number 11
19. Chapter 14, number 7
20. Chapter 14, number 13
Part II – SAT and Expenditures Data

This dataset is called ‘sat.dta’. It is already in Stata format. You must use Stata for the following questions. Here is a list of the variables in the dataset and their descriptions:

- state: Name of state (in quotation marks)
- expend: Current expenditure per pupil in average daily attendance in public elementary and secondary schools, 1994-95 (in thousands of dollars)
- pt_ratio: Average pupil/teacher ratio in public elementary and secondary schools, Fall 1994
- salary: Estimated average annual salary of teachers in public elementary and secondary schools, 1994-95 (in thousands of dollars)
- percent: Percentage of all eligible students taking the SAT, 1994-95
- verb_sat: Average verbal SAT score, 1994-95
- math_sat: Average math SAT score, 1994-95
- tot_sat: Average total score on the SAT, 1994-95

1. Make sure you create a print out the stata log file. Edit it where appropriate.

2. Examine the data. Briefly discuss how the data is set up.

3. Create dummy variables for regions of the United States. [NORTHEAST (CT, ME, MA, NH, RI, VT, DE, MD, NJ, NY, PA), SOUTH (AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV), MIDWEST (IL, IN, MI, OH, WI, IA, MN, MO), WESTISH (KS, NE, ND, SD, CO, ID, MT, UT, WY, OK, TX), FARWEST (AZ, NM, AK, CA, HI, NV, OR, WA)]

4. Report the means and standard deviations for all variables. Compare the means and standard deviations by region. Discuss.

5. Report the correlation coefficients. Discuss.

6. Run an OLS regression with total sat as your dependent variable and expenditures and salary as your independent variables. Report and discuss these results. Do you suspect any of the problems that we discussed in chapters 6-10? Which one(s) and why? How should this (these) problem(s) be dealt with?

7. Run an OLS regression with total sat as your dependent variable and expenditures as your only independent variable. Interpret the results. Is anything omitted? Assume that ‘percent’ is an omitted variable. Why might this be considered omitted? What do you expect the bias to be and why? Run another OLS regression and discuss the bias.

8. Discuss and interpret the results. What specification do you feel most comfortable with and why?
9. Run an OLS regression using the variables that you suggest in question 8 but changing total sat and expenditures to logs. Discuss the results.

10. Run an OLS regression using the variables that you suggest in 8 but add an expenditures-squared variable to the equation. Why might you do this? Report and discuss the results.

**Part III – Car Data (again)**

*Use the car data from homework 1.*

1. Make sure you create a print out the stata log file. Edit it where appropriate.

2. Run an OLS regression on price with midsize, airbag, horsepower, and U_turn as independent variables. Report and discuss.

3. What classical assumptions are you concerned with here and why?

4. Run another OLS regression correcting for the potential problem above. Report and discuss. What conclusion can be made about the uncorrected regression? How have your results changed?

**Part IV – Impeachment Data**

*This dataset is called ‘impeach.dta’. It is already in Stata format. You must use Stata for the following questions. Here is a list of the variables in the dataset and their descriptions:*

- **name**: Name of senator
- **state**: State (postal code)
- **perjury**: Vote on Article I, Perjury: 0 = Not Guilty, 1 = Guilty
- **obstruct**: Vote on Article II, Obstruction of Justice: 0 = NG, 1 = G
- **n_guilty**: Number of votes for guilt
- **repub**: Party: 0 = Democrat, 1 = Republican
- **conserva**: Senator's degree of ideological conservativism (0-100)
- **vote_wjc**: Percent of the vote Clinton received in the 1996 election in each state
- **election**: The year each Senator's seat is up and he/she must run for re-election
- **frosh**: First-term senator? 0 = no, 1 = yes

1. Make sure you create a print out the stata log file. Edit it where appropriate.

2. Examine the data. Briefly discuss how the data is set up.
3. Report and discuss the means of the variables.

4. Report and discuss the correlation coefficients.

5. Use a linear probability model with perjury as your dependent variable and repub and vote_wjc as your independent variables. Report and discuss results. Use a linear probability model with obstruct as your dependent variable and repub and vote_wjc as your independent variables. Report and discuss results.

6. Repeat replacing repub with conserva.

7. Use a logit model with perjury as your dependent variable and repub and vote_wjc as your independent variables. Report and discuss results. Use a logit model with obstruct as your dependent variable and repub and vote_wjc as your independent variables. Report and discuss results.

8. Repeat replacing repub with conserva.

9. Discuss the differences in the results between the linear probability model and the logit model. Which results do you prefer and why?