PS205 Final, Fall 2013
(Due Monday, December 16th)

You will have unlimited time to work on this final. This final is worth 100 points. I will deduct 10 points per day the final is late starting on Tuesday, March 26th, unless prior arrangements are made.

The final will be open book and open notes, but you cannot work with anyone else. You may ask me clarification questions (such as “If I estimated model X, would you expect me to report Y?”), but I cannot give you any substantial help (such as telling you how to interpret model results).

I have posted two new datasets to our class page for the final exam, named CES2011.txt (tab delimited) and USCounties.csv (comma delimited). The codebooks for each of these datasets are on the following pages. You may use one or both of these datasets to answer the questions below. You may also use any of the data we’ve used in class in the past, as long as you examine a new hypothesis we didn’t cover in class or on a homework.

For the final you must form three different hypotheses based on the variables in whichever datasets you choose to use. At least one of your hypotheses must be a conditional hypothesis.

Then test each of the three hypotheses with a linear regression. For each model I’m expecting you to (1) estimate an appropriate regression, (2) interpret it completely, (3) perform all necessary diagnostics to ensure there are no problems, (4) fix any problems that you find (or at least try to fix them), and (5) reach the appropriate conclusion about your hypothesis.

For each regression model, include at least two control variables in addition to the independent variables necessary to test your hypothesis. You must do a joint hypothesis test for at least one of your models.

To recap:

- Examine the available data and form three hypotheses.
- At least one of those hypotheses will be a conditional hypothesis.
- Estimate appropriate regressions to test those hypotheses.
- Each regression should include at least two control variables.
- Perform all necessary diagnostics and fix any problems you find.
- Perform at least one joint hypothesis test.

Write your results up nicely, as if you were going to submit them to a journal. Put your results into tables, rather than just pasting in your computer output. Write things up carefully, making sure to justify your decisions. It might help to look at the methods sections of a couple of journal
articles (either ones I’ve handed out or others) to get a feel for what these sections should look like. You should write about 3-4 pages per problem.

For each hypothesis you test give me a 3-4 page write-up and your R script file. I should be able to replicate your results by running your script file.
Codebook for the 2011 Canadian Election Survey (CES 2011)

- **id**: Respondent ID number.
- **vote**: Vote choice among the four major parties in the election. 1 = Conservatives, 2 = Liberals, 3 = NDP, 5 = Green.
- **feel_Cons, feel_Lib, feel_NDP, feel_Green**: 101 point thermometer ratings for the four parties (0 = greatly dislike, 50 = neutral, 100 = greatly like).
- **natecon**: National economy in the last year (1 = better, 2 = stayed the same, 3 = worse).
- **persecon**: Personal finances in the last year (1 = better, 2 = stayed the same, 3 = worse).
- **education**: An 11 point scale for the respondent’s education (higher = more education).
- **daysnews**: Number of days in the week the respondent spent reading news on the internet.
- **ownhome**: 0 = renter, 1 = home owner.
- **health_spend, ed_spend, unemp_spend, def_spend, old_spend, bus_spend**: A five point scale measuring the respondent’s opinion on how much the government should spend on health, education, unemployment benefits, defense, old age pensions, and supporting business, respectively. (1 = much more, 2 = more, 3 = same, 4 = less, 5 = much less).
- **LR_self, LR_Cons, LR_Lib, LR_NDP, LR_Green**: 11 point scale for respondent’s self-rated ideology and ideological rating of each party. (0 = far left, 5 = moderate, 10 = far right).
- **yearborn**: The year in which the respondent was born.
- **gender**: 0 = male, 1 = female.
Codebook for the 2004 US County Voting Data

- **name_state**: State name.
- **county**: County name.
- **totalvote**: Total votes cast in the county.
- **dem**: % of votes for the Democratic Presidential candidate (Kerry).
- **rep**: % of votes for the Republican Presidential candidate (Bush).
- **tot_property_damage**: Total property damage in the county from natural disasters.
- **tot_crop_damage**: Total crop damage in the county from natural disasters.
- **tot_fatalities**: Total fatalities in the county from natural disasters.
- **tot_injuries**: Total injuries in the county from natural disasters.
- **cpi**: The Consumer Price Index (relative inflation) in the county.
- **population_density**: The population density of the county (people per square mile).
- **median_hhold_income**: The median household income of the county.
- **pct_farm**: The percentage of land in the county devoted to farming.
- **pcFedGovExp**: Federal expenditures in the county, in dollars per capita.
- **pct_poor**: Percentage of county residents below the poverty line.
- **college_or_more**: Percentage of county residents with a college education or greater.
- **percent_nonwhite**: Percentage of nonwhite county residents.
- **population**: Total county population.