ESM 206A Problem Set 3

The two datasets for this homework are posted on the course webpage: ExoticSpecies.xls and cars_1993.xls. The following table defines the variables from ExoticSpecies.

Variable	Definition
Country	The name of the country
Area_sqkm	The area in square kilometers
Natives	Estimate of the number of native plant species
Exotics	Estimate of the number of exotic plant species
Island	=1 if the country is an island, otherwise = 0
GDP	Per-capita GDP (income)
Pasture	Percent of land used as pasture
Agr	Percent of land used for agriculture
Pop_dens	Number of people per sq kilometer
Prop_exotic	Estimate of the proportion of plant species that are exotic
M_imports	Percentage of GDP from imports

1) Examine some alternative predictive models for the exotic species data.
    a. Separately regress the proportion of exotics on the per capita GDP and on population density. Which is a better model? Explain (in words) the quantitative impact of GDP on the fraction of exotic species. Discuss whether you think this might be a causal relationship.
    b. A measure of economic intensity might be GDP per unit area, instead of GDP per person. Figure out how to calculate such a quantity, and use it as the independent variable in the regression. How do the results of this model differ from those of the per capita GDP? Why do you think this is so?
    c. Many exotic species are introduced through trade, and thus the amount of imports (measured on a per-capita, per-area, or total basis) might be a good predictor of the extent to which a country has been exposed to invasion. Construct one or more measures of the amount of imports each country has, and use it to predict the proportion of exotic species. How does this compare, statistically and from a causality perspective, with the other factors you have looked at? *Note that one country is missing import data; if you want to compare the fit of this model to the previous ones, you should re-run the previous models with that country deleted, so that they are strictly comparable.*

For this section use the cars_1993.xls data on the course website

The following table defines the variables.

Variable	Definition
Manufacturer	Manufacturer name
Model	Car model
Type	Car type
Price	Price in $000s
CityMPG	MPG in the city
HighwayMPG	MPG on highways
EngineSize	In liters
Horsepower	Horsepower
Fuel Tank	Capacity of tank in gallons
Passengers	Number of passengers car fits
Weight	In pounds
Origin	Whether or not the car is an American brand

2) a. Use single-variable regression to examine the relationship between Weight and CityMPG. Can we reject the null hypothesis that increasing car weight does not decrease fuel efficiency? What is the predicted effect of increasing car weight by one standard deviation on CityMPG?

   b. Now examine separately the relationship between weight and fuel efficiency by origin (e.g., run regressions of Weight on CityMPG for US and non-US cars). Is there evidence that the effect of Weight differs by origin?

   c. Perform a log transformation of Weight and CityMPG, and run a regression of the natural log of Weight on the natural log of CityMPG (don't do this separately by origin). How does the linear fit of these data compare with those in (a)? What is the predicted effect of decreasing car weight by 15% on CityMPG?

   d. Based on our analysis, do you think we have tightly estimated the casual relationship of weight on city fuel efficiency? How could we improve our analysis?