PROBLEM SET THREE

Antonio Pacifico

Federica Romei

Deadline : 29/03/2012

Use the dataset *bwght.dta*. You can download the dataset from Prof. Ragusa Luiss page on Didactic Material.

1. A problem of interest to health officials is to determine the effects of smoking during pregnancy on infant health. One measure of infant health is birth weight : a birth rate that is too low can put an infant at risk for contracting various illnesses. For example, higher income generally results in access to better prenatal care, as well as better nutrition for the mather. An equation that recognizes this is:

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 faminc + \upsilon$$

A. Compute this regression and explain the significance of coefficients. What is the most likely sign for β_2 ?

B. Do you think *cigs* and *faminc* are likely to be correlated? Explain why the correlation might be positive or negative.

C. Now estimate the equation with and without *faminc*. Interpret the coefficients. Discuss the result, focusing on whether adding *faminc* substantially changes the estimated effects of *cigs* on *bwght*.

D. Do you think that there is a causal effect between *cigs* and *bwght* ? Explain the answer.

Use the dataset *affairs.dta.* You can download the dataset from Prof. Ragusa Luiss page on Didactic Material.
This dataset contains 601 observations on 9 variables. The variables of interest are:

- **AFFAIRS : how often engaged in extramarital sexual intercourse during the past year.**

- **YEARSMARRIED : numeric variable coding number of years married. To be more precise: 0.125 = 3 months or less, 0.417 = 4–6 months, 0.75 = 6 months–1 year, 1.5 = 1–2 years, 4 = 3–5 years, 7 = 6-8 years, 10 = 9-11 years, 15 = 12 or more years.**
- **D_CHILDREN : dummy variable =1 if you have children in the marriage and =0 otherwise.**

2.

A. Construct a procedure statistic to test that number of betrayals is equal whether people are married for two years or for five years against the alternative hypothesis that married people for five years betray more. Use a significance level of 5%. Comment.

B. Regress *affairs* on *d_children* describing the model. Is it statistically significant? How can you interpret β_1 and β_0? Is there a causal effect between *d_children* and *affairs*? Explain the answer.

C. Test [using a significance level of 5%] that number of betrayals is equal whether you have a children in the marriage or not against the alternative hypothesis that it is different. Do you note some similarities with previous point? Why? Explain the answers.

Use the dataset *fatalities.dta*. You can download the dataset from Prof. Ragusa Luiss page on Didactic Material.
It is an annual data frame for 1982 through 1988 describing US traffic fatalities for the lower 48 US states (i.e. excluding Alaska and Hawaii). The dataset contains 336 observations on 34 variables:

- **STATE : Factor indicating state.**
- **YEAR : Factor indicating year.**
- **SPIRITS : Spirits consumption.**
- **UNEMP : Unemployment rate.**
- **INCOME : Per capita personal income in 1987 dollars.**
- **EMPPOP : Employment/population ratio.**
- **BEERTAX : Tax on case of beer.**
- **BAPTIST : Percent of southern baptist.**
- **MORMON : Percent of mormon.**
- **DRINKAGE : Minimum legal drinking age.**
- **DRY : Percent residing in dry countries.**

- **YOUNGDRIVERS : Percent of drivers aged 15-24.**
- **MILES : Average miles per driver.**
- **BREATH : Preliminary breath test law?**
- **JAIL : Mandatory jail sentence?**
- **SERVICE : Mandatory community service?**
- **FATAL : Number of vehicle fatalities.**
- **NFATAL : Number of night-time vehicle fatalities.**
- **SFATAL : Number of single vehicle fatalities.**
- **FATAL1517 : Number of vehicle fatalities, 15-17 year olds.**
- **NFATAL1517 : Number of night-time vehicle fatalities, 15-17 year olds.**
- **FATAL1820 : Number of vehicle fatalities, 18-20 year olds.**
- **NFATAL1820 : Number of night-time vehicle fatalities, 18-20 year olds.**
- **FATAL2124 : Number of vehicle fatalities, 21-24 year olds.**

3. From 1982 through 1989, the estimated number of fatalities in crashes in which at least one driver or non-occupant was intoxicated decreased 12%, from 20,356 to 17,849% in previous years. A fatal traffic crash is considered alcohol-related by NHTSA (National Highway Traffic Safety Administration's) if either a driver or non-occupant had a blood alcohol concentration (BAC) of greater than or equal to 0.01 g divided by L (the legal level of intoxication in most states) in a police-reported traffic crash[1].
Factors that may have contributed to the reduction in ARTFs (Alcohol Related Traffic Fatalities) include : changes in state laws and stricter enforcement of these laws ; increases in the minimum legal drinking age[2] in 35 states from 1982 through 1987 ; increased media attention resulting in increased public awareness ; increased number of programs emphasizing responsible behavior and alternatives to drinking and driving (I.E. education of persons who serve alcoholic beverages and designation of nondrinking drivers) et al.

A. Regress *nfatal* on *spirits*, *income*, *drinkage*, *youngdrivers* explaining the model and corresponding assumptions. Describe the coefficients and their significance. Are *spirits* and *income* correlated between them? Why? Explain the answer.

[1] To be more precise NHTSA defines a BAC of greater than or equal to 0.01 g divided by L but less than 0.10 g divided by L as indicating a low level of alcohol and a BAC of greater than or equal to 0.10 g divided by L as indicating intoxication.
[2] The minimum drinking age in US is 21 years old from 1984.

B. Do the same regression adding *beertax* and *unemp*. Rewrite the model with corresponding assumptions. How does model change? Is new model statistically more significant? Prove it and comment.

C. Now consider that β_3 and β_4 are not significant [$\beta_3 = \beta_4 = 0$] and make the regression. Is the model more profitable with respect to that of point **B.** in explaining *nfatal*? Why? Explain the answer.

D. Finally try making a new multiple regression and explain your choice. Comment carefully. {HINT : you have to consider a different dependent variable, hence you choice your regressors}