17.871
Spring 2007
Problem Set #2

Handed out: March 15, 2007
Due: April 5, 2007

1. (10 points) Using that data set quartet.dta:
   a. Regress each y on its corresponding x (e.g., y1 on x1, y2 on x2). Present the results in a table with four columns, one for each regression. The table should contain the coefficients, the P values, and the standard error of regression (Root MSE).
   b. Interpret the coefficients and the standard error of regression.
   c. Do you believe these estimates? Explain.
   d. What should you conclude about the use of regression (and other fancy statistical procedures) from this example?

2: (20 Points) Death Sentence and Discrimination

In a 1987 Supreme Court case, McCleskey v. Zant, McCleskey's counsel argued that the race of both victim and defendant played a significant role in the decision to impose the death penalty, even after accounting for other facts of the case. On this basis they argued that capital punishment was being applied in an unconstitutionally discriminatory manner.

In a 5-4 decision, the Court rejected McCleskey's claim. Justice Powell, writing for the majority, argued that statistical evidence of discrimination was not sufficient to overturn McCleskey's conviction. Instead, Powell wrote, McCleskey would have to show strong evidence of racial bias in this particular case. Statistical patterns alone would not be sufficient. (Powell, now retired, is quoted in a recent book as saying he has now changed his mind on this issue and would vote differently if presented with the case today.) While McCleskey v. Zant has essentially settled the question for the courts, this issue has not disappeared. In an amendment to the 1994 "Crime Bill" members of the House of Representatives proposed a provision allowing statistical evidence to be used to overturn death penalty convictions. The Senate rejected the amendment and a hot debate between the two chambers ensued. The amendment was removed from the final bill, over the strong objections of many House members, especially the Black Caucus.

Nationally, public support for the death penalty has remained very strong, with over 60% of the public supporting executions.

With that background, reanalyze data from the McCleskey case. In the file death.dta you will find 100 cases used as part of McCleskey's evidence (this is a subset of the cases used by McCleskey).

The variables:

| | |
|---|---|
| ID | 1--100 ID number of each case |
| death | 1=Death sentence |
| | 0=Life sentence |
| bd | 1=Black defendant |
| | 0=White defendant |
| wv | 1=One or more white victims |
| | 0=No white victims |
| ac | Number (1-6) of aggravating circumstances |
| fv | 1=Female victim |
| | 0=Male victim |
| vs | 1=Victim was a stranger to defendant |
| | 0=Victim was known to defendant |
| v2 | 1=Two or more victims |
| | 0=One victim |
| ms | 1=Multiple stabs |
| | 0=No multiple stabs |
| yv | 1=Victim 12 years of age or younger |
| | 0=Victim over 12 years of age |

Analyze these data.

a. In sentencing people to death, does the Justice system discriminate against blacks? Use cross tabulations (the *tabulate* command) to check. Show and briefly summarize your findings. (Hint: consider more than one form of discrimination.)
b. Regress death on bd, that is, run a regression where death is the dependent variable and bd is the explanatory variable. Report and interpret the coefficient and the constant. (Hint: in regressions with a binary dependent variable, the coefficients can be interpreted as probabilities.)
c. Regress death on bd and wv. Report and interpret the coefficients and the constant. Why does the coefficient on bd change when controlling for wv? (Hint: look at the correlation matrix by running the command *corr*.)
d. Consider your findings in part a. What variables in the data set might provide alternative explanations for your finding? In other words, what variables in the data set might lead to the appearance (or absence) of discrimination in a. In answering this question, you might find it helpful to look at the results from the correlations in c., but you should primarily rely on your expectations and reasoning about the justice system.
e. Regress death on all the substantive variables. Before doing so, code ac so that it varies from zero to one. Do your bivariate findings from part a. hold? That is, does controlling for the other variables change the bivariate results? Report the regression results and discuss whether your findings from part a. change. In doing so, interpret the coefficient on wv.

f. From part e., which coefficient has the largest effect on the death sentence? Interpret its coefficient.
g. The O.J. Simpson case presented circumstances that can be applied to these data. For the benefit of those not familiar with the case, here is a brief summary. Simpson is a retired American football player. In 1994, he was charged with murder in the deaths of his former wife and a man she knew. The victims were killed by multiple stab wounds. Simpson obviously knew his exwife, but apparently did not know the male victim. Simpson is black. Both victims were white. No children where killed. We do not know how many aggravating circumstances there were, but it seems likely to be at least 1 and perhaps as many as 3 or 4. Use this information in your regression model from part e. to compute predicted probabilities that Simpson would receive the death penalty if convicted of murder based on these data from Georgia. (In fact, the Los Angeles District Attorney decided not to seek the death penalty, and Simpson was acquitted of the murder charges. That does not diminish the interest in asking if someone in a case similar to Simpson's could expect the death penalty.) Some of the explanatory variables are easily applicable. Two, vs and ac, are not so easily applicable. You decide how to handle these, and justify your decision. Discuss the results. (Hint: We've only briefly covered model-based predictions, but they are (usually) straightforward. You simply need to plug in the hypothetical values for the explanatory variables into the model produced by the regression. Don't forget to add the constant. You can probably do this most quickly by hand or in Excel.)