Statistics 104 — Fall, 2004 — Assignment 8
Due Wednesday, December 15th, 2004.

Readings (Moore and McCabe)

- Chapters 9, 10, 11.

Against All Odds videotape

The relevant tapes for this week are numbers 24 (Inference for two-way tables), 25 (Inference for relationships), and 26 (Case Study)..

Written Assignment (Moore and McCabe)

- **MM:** 8.6, 8.20, 8.36, 8.46, 9.28, 9.34, 9.35, 9.36, 9.46, 10.2, 10.9, 10.20, 10.32, 10.34, 10.40. For the chapter 8 problems, please show how you get your answers and only use Stata to verify your results. For the chapter 9 and 10 problems, you may use Stata to do all your calculations.

Stata Hints

Two way tables: As discussed in class, the analysis of two (and higher) way tables can be done easily in Stata. For a two way table, the data needs to be entered into 3 columns. The first column contains the cell counts, the second column contains the levels of one categorical factor, and the third column contains the levels for the other categorical factor. Assume that the factors levels are in variables `var1` and `var2` and the cell counts are in a variable `count`. Then the Chi-square analysis can be performed by

```
tabulate var1 var2 [fweight=count ], chi2
```

If Fisher's exact test is also desired (only do for 2x2 tables due computation time), in addition to `chi2` (or replacing it) give the option `exact`. For further information on the command and its additional options, check `help tabulate` and the overheads for the November 19th lecture.

One- and two-sample binomial tests: The large sample hypothesis tests based on the normal approximation to the binomial can be done with the `prtest` function. Usually the immediate form is more useful here. The forms for the one- and two-sample cases are

```
prtest n #successes, count
```

and

```
prtest n1 #successes1 n2 #successes2, count
```

Again for more information check `help prtest`.

Running a regression: To run a regression in Stata, a command of the form

```
regress y x1 x2 x3
```

needs to be run. Of course, the number of predictor variables given in the list depends on the problem of interest. After this command is run, the `predict` command can be used to get different diagnostic variables and summaries. The 4 that will be of the most use to you are

```
predict fits, xb
predict resid, residuals
predict semu, stdp
predict sepred, stdf
```

This 4 commands will store the fitted values for each observation in a variable `fits`, the residuals in `resid`, the standard errors of μ_y in `semu`, and the standard errors of prediction in `sepred`. Note that you can use any for these variables.

Getting confidence interval for μ_y and prediction intervals for new observations: Stata will not directly calculate these two types of intervals, however it is not difficult to do. First you need to add the levels of the predictor variables to your data set, while leaving the response variable empty for these rows. Then run the regression as described above. The rows with where the response variable is missing will not be included in the main regression command. However if the `predict` commands for getting the fits, the standard error of the mean response, and the standard error of prediction are run, the desired fits and standard errors will calculated and stored in the appropriate variables. Then the intervals can either be calculated by hand or with Stata using the calculated fits and standard errors.