

Statistics 104 — Fall, 2004 — Practice Problems

Not to be handed in.

Written Assignment (Moore and McCabe)

- **MM:** 11.2, 11.4, 11.32, 11.33, 11.34 11.52, 11.53, 12.6, 12.10, 12.32, 12.36, 12.38

Stata Hints

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 name 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 be 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.

Running an ANOVA: To run a one-way ANOVA you can either use the `oneway` command or the `anova` command. Unless you need the summary table of means and standard deviations for each group, `anova` is usually the preferable way to go. It allows for diagnostics to be calculated as with regression (the above commands in the regression section will work), multiple comparison procedures, and tests on particular parameters.

The general form of the `anova` command is

```
regress y f1 f2 f3
```

where f_1 , f_2 , f_3 are categorical factors you wish to examine. The number of different predictor variables used in the `anova` command is variable. It is allow possible to combine categorical factors with continuous predictors. For example, in interaction model for the fiber dataset discussed in class can be examined with

```
anova strength diameter machine diameter*machine, continuous(diameter)
regress anova
```

This approach is nice as Stata will calculate all the indicator variables for you. See `help anova` for more info.