SAS Work Shop
PROC REG
Handout #2

Statistical Programs
College of Agriculture

HTTP://WWW.UIDAHO.EDU/AG/STATPROG

Regression

Required Statements:

MODEL: The only required statement under PROC REG is the MODEL statement. This statement defines the model as well as the order in which variables are included in the model. The general form of the MODEL statement is: ***Dependent var's = Regressor var's / option$_1$ option$_2$ ... option$_n$.*** Here, ***dependent var's*** is a list of one or more numeric responses (variables to be modeled) and ***regressor var's*** is also a list of one or more numeric variables paired with the dependent var's. In most applications of PROC REG, both ***dependent var's*** and ***regressor var's*** are continuous, numeric values. ***Options$_i$*** are keywords which specify various details of the analysis, and as usual in SAS, they are listed at the end of a statement after a slash (/). More than one MODEL statement can be used in each call to PROC REG, if needed.

Unlike the MODEL statement of PROC GLM, PROC REG does not allow the use of crossed or nested terms. Thus, the symbols *, |, and () should not appear in the MODEL statement. If a crossed term is required in the model, it must be explicitly defined in a preceeding DATA STEP. An example would be a quadratic term for a second order polynomial:

SAS Work Shop
PROC REG
Handout #2

Statistical Programs
College of Agriculture

HTTP://WWW.UIDAHO.EDU/AG/STATPROG

```
DATA QUAD;

      infile ...;

      input ...;


      date_2 = date*date;  <------- make date²

            .

            .

PROC REG DATA=QUAD;

      MODEL y = date  date_2; <------- this is correct


      MODEL y = date  date*date;    <------- This will not work
```

MODEL Statement Options:

PROC REG provides a wide array of options for regression analysis. All of these, however, fall into three broad categories. **Printing** options determine what details of the analysis should be printed in the output. Examples would be P, CLM, or I for predicted values, prediction intervals, and the $(X\,X)^{-1}$ matrix, respectively. The second type of options are **"diagnostic"**. Examples include R, DW, INFLUENCE, for a preliminary residual analysis, the Durbin-Watson statistic, and influence statistics, respectively. The last category of options are used to specify **the model** itself. Examples of these are NOINT, SELECTION, INCLUDE. These, respectively, result in the fit of a model with no intercept, using various selection procedures when multiple regressor variables are present, and inclusion a specific list of regressors in all models tested.