

Linear Regression Continued

1 More with PROC GLM

PROC GLM which we discussed in the last lab is a procedure that can be used to fit simple linear regression and many other type of models. For our purposes today we are going to use **PROC GLM** to do some inference (i.e. confidence intervals). Recall that the general statement for the models is:

```
PROC GLM < options > ;
  MODEL dependents=independents < / options > ;
  BY variables ;
  TEST < H=effects > E=effect < / options > ;
  OUTPUT < OUT=SAS-data-set > keyword=names < ... keyword=names > < / option > ;
```

Recall that the **MODEL** statement will consist of the variables being considered for a relationship, with the dependent being the y and the independent variable being x. The main output 'keyword's' we'll be using are **OUT**, **PREDICTED** or **P**, and **RESIDUAL** or **R**. Today we'll go over the options **CLPARM**, and **ALPHA** that go on the **MODEL** line after the backslash.

Here is an example using the hanes data set:

```
data new;
set tmp1.hanes;
sec=sex;
run;

proc sort data=new;
by sex;
run;

proc glm data=new;
by sex;
model sysbp=diabp\clparm alpha=.1;
output out=regdata r=resid p=yhat;
run;
quit;

proc gplot data=regdata;
plot yhat*resid;
run;
quit;
```

2 Using the Data Step, with the IF statement

The data step in SAS is powerful. Some companies (one's i've worked for) use the data step almost exclusively and the statistical options little. What the data step allows you to do is take a gigantic data set and manipulate it with ease. It also has allows the power of working with no data at all.

An example is turning a random variable that has 3 outcomes into one that has two. Take the hanes data set for example, there is a variable smoke that can be one of three levels: Not smoked more that 100 cigarettes (1), smoked more than 100 but does not smoke now(2), or smoked more than 100 and is still smoking now (3). Say we are interested in 1 vs (2 or 3).

Here is the code to do that;

```
data new;
set tmp1.nhanes;
if smode eq 1 then smoke2=1;
if smode eq 2 or smode eq 3 then smoke2=2;
run;
proc print;
run;
```

Another example is generating State's by random;

```
data new;
input SC $ NC $ ALB $ FL $ GA $ OTH $;
cards;
SouthCar NorthCar Alabama Florida Georgia Other
;
data test;
set new;
do i=1 to 50000;
st=ranuni(0);
if st le .2 then state=OTH;
if st gt .2 and st le .4 then state=SC;
if st gt .4 and st le .6 then state=NC;
if st gt .6 and st le .8 then state=ALB;
if st gt .8 and st le .9 then state=FL;
if st gt .9 and st le 1.0 then state=GA;
output;
end;
drop SC NC ALB FL GA OTH;
run;

proc freq;
tables state;
run;
```