

Simple Linear Regression

1 PROC GLM Overview

PROC GLM (generalized linear models) is a procedure that can be used to fit simple linear regression, with a single or multiple variables, quadratic regression, analysis of variance, along with many other types of models. For our purposes today **PROC GLM** will be used to obtain out least squares estimates for B_0 , B_1 , and R^2 . The general statement for the models we'll be fitting today is:

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

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.

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;
output out=regdata r=resid p=yhat;
run;
quit;

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

Notice that I create a new temporary data set in order to sort the variables by sex. **PROC GPLOT**, used here to plot the residual versus the predicted values, is a high resolution scatterplot procedure.

2 Creating Z-scores with PROC STANDARD

PROC STANDARD is a procedure that can be used to create z-scores in SAS. The syntax for STANDARD is similar to that of proc means, here is an example:

```
proc standard data=new out=ABC mean=0 std=1 REPLACE;  
var sysbp diabp;  
by sex;  
run;  
proc print data=ABC;  
run;
```

This procedure creates a temporary data set ABC where sysbp, and diabp are 'standardized'. This means that there values are equal to there z-scores.