

Introduction to Regression Analysis with SAS

The purpose of this exercise is to introduce you to the basics of doing regression analysis with SAS. Here we will use a ‘toy’ data set containing n=10 observations on the following variables:

Obs	NAME	SEX	PERSTEST	THERAPY	INTEXT	SX
1	John	M	26	32	3	0
2	Susan	F	24	40	4	1
3	Mary	F	22	44	8	1

Here, THERAPY is the outcome measure—improvement after therapy; PERSTEST is a personality test, INTEXT is a scale of internal-external locus of control and SX is a dummy variable coding SEX.

1. The data is stored in N:\data\therapy.sas. Read it into SAS using

```
%include data(therapy);
```
2. It is useful to get an overview of the data by plotting. One simple way is a scatterplot matrix,

```
%scatmat(data=therapy, var=therapy perstest intext sx, interp=rl);
```
3. Let’s start using PERSTEST as a single predictor. With PROC REG we can ask for more than the brief default output using options on the model statement.

```
proc reg data=therapy;
  m1: model therapy = perstest / P CLM;
  id name;
  plot residual. * pred. ;
  symbol1 v=dot;
run;
plot therapy * perstest / conf95 vref=50 ;
run;
```

The plot statements produce simple plots of residuals vs. predicted values and y vs. x, with confidence intervals for the mean predicted value. Use HELP REG in the command box to find out about other options for the MODEL and PLOT statements.

Q: Is there evidence that the PERSTEST, by itself, significantly predicts THERAPY? [Hint: Look at the Pr > F for the model and the Pr > |t| for the PERSTEST variable.]

4. We can fit additional predictors simply by adding them to the right on the MODEL statement

```
proc reg data=therapy;
  m2: model therapy = perstest intext/ P ;
  id name;
  plot residual. * pred.; run;
```

Examine the coefficients for PERSTEST and INTEXT, as well as the F test for the overall model. What has changed from the 1-predictor to the 2-predictor model?

5. Let’s add the dummy variable for SEX, here called SX – PROC REG can’t handle factors or character variables (PROC GLM can)

```
proc reg data=therapy;
  m3: model therapy = perstest intext sx / P ;
  id name;
  plot residual. * pred.; run;
```

6. Make a table containing the results of your analyses for these three models

Model	b: perstest	b: intext	b: sx	MSE	R ²
1		xxx	xxx		
2			xxx		
3					

In the 'b:' columns, enter the parameter estimates and their $\text{Pr} > |t|$ values.

Q: Why do you think the effect of the PERSTEST as a predictor of THERAPY changes?

We'll explore other ways of fitting models with multiple predictors starting next week.