



BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 8

DATE: April 12, 1988

SUBJECT: Standard errors for predicted values from Multiple Regression

There are two ways to get the standard errors for predicted values (as opposed to fitted values, see Note 2 below) from a multiple regression by using PROC REG in SAS:

```
1) PROC REG DATA = A;
    MODEL Y = X1 X2;
    OUTPUT OUT = PRED P = PY STDI = SEPREDD;
    PROC PRINT DATA = PRED;
    VAR X1 X2 Y PY SEPREDD;
```

The output data set contains the original data plus the predicted values and their standard errors. The standard errors are different depending on the values of X1 and X2. Values can be obtained from combinations of X1 and X2 not in the original data set by adding them in the DATA step before running PROC REG. For example:

```
DATA A;
  INFILE RAWIN EOF = ADDON;
  INPUT X1 X2 Y;
  ADDON: DO X1 = 1 TO 10 BY 1;
         DO X2 = 1 TO 15 BY 1;
           Y = . ;
           OUTPUT;
         END;
  END;
```

Although the values of Y will be missing the data set output from PROC REG will include the predicted values (PY) and their standard errors (SEPREDD).

```
2) PROC REG DATA = A;
    MODEL Y = X1 X2 / I;
```

This method provides the parameters for an equation of the standard error of predicted values. The general equation in matrix form is given in Weisberg (pg. 47) as:

$$\text{sepred}(\tilde{y}|x) = \hat{\sigma} \sqrt{1 + x^T (X^T X)^{-1} x}$$

where $\hat{\sigma}$ is the square root of the mean square error
 x is the vector of x-values, i.e. $[1 \ x_1 \ x_2]$

and $(X^T X)^{-1}$ is the inverse of the crossproducts matrix.

The inverse of the crossproducts matrix is output by PROC REG when I is an option on the MODEL statement. Suppose that the following matrix is output (see example printout below):

$$\begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}$$

Then the standard error is $\text{sepred}(\tilde{y}|x) = [\hat{\sigma}^2 (1 + a + 2dx_1 + 2ex_2 + 2fx_1x_2 + bx_1^2 + cx_2^2)]^{\frac{1}{2}}$

In both cases, confidence limits can be obtained by multiplying the standard error by the appropriate t-value and adding or subtracting from the predicted value (note the CLI option for the MODEL statement in PROC REG).

Example Printout:

X'X Inverse, Parameter Estimates, and SSE

INVERSE	INTERCEP	X1	X2	Y
INTERCEP	$a = 2.7787470343$	$d = -.0112419288$	$e = -.1060979041$	9.1268854149
X1	$d = -.0112419288$	$b = .00014620682$	$f = .00017546691$	-.0723929437
X2	$e = -.1060979041$	$f = .00017546691$	$c = .00478598541$	0.2028153865
----- -----				
Y	9.1268854149	-0.723929437	0.2028153865	9.6287038506

NOTE 1: The inverse of the crossproducts matrix for a simple regression would be:

$$\begin{bmatrix} a & d \\ d & b \end{bmatrix}$$

and $\text{sepred}(\tilde{y}|x) = [\hat{\sigma}^2 (1 + a + 2dx_1 + bx_1^2)]^{\frac{1}{2}}$

NOTE 2: Fitted and predicted values have the same estimate but the variability of the estimate is greater when extending the results to NEW data. The expected variability (measured by the standard error) of the predicted value includes not only the variability of the parameter estimates (as for the fitted value) but also the variation of the data about the fitted equation (as measured by the residual variation $\hat{\sigma}$).

Reference:

Weisberg, Sanford, 1980. *Applied linear regression*. John Wiley and Sons, New York, N.Y.

CONTACT: Wendy Bergerud
387-5676