# FORMATTED DISPLAYS USING OUTPUT FROM PROC SYSREG

P.A.Reid and M.W.Harrell
Cambridge Systematics and Research Triangle Institute

Source statements and examples are presented to compute various statistics based on the regression output from PROC SYSREG. These statistics are on the printed output of PROC SYSREG ; however this code summarizes many pages of regression output into concise tables and graphical displays.

Using the regression coefficients from the OUTEST dataset of PROC SYSREG and the sums of squares from the OUTSSCP dataset; the following statistics are calculated:

- the t statistics associated with the regression coefficients,
- the associated standard errors of the regression coefficients,
- the associated p values of the t statistics,
- the R-square of the model, and
- the F value of the model.

Character variables of asterisks are created to indicate significance of the coefficients and the overall F test at the .05 and .01 significance levels ('*' and '**' respectively).

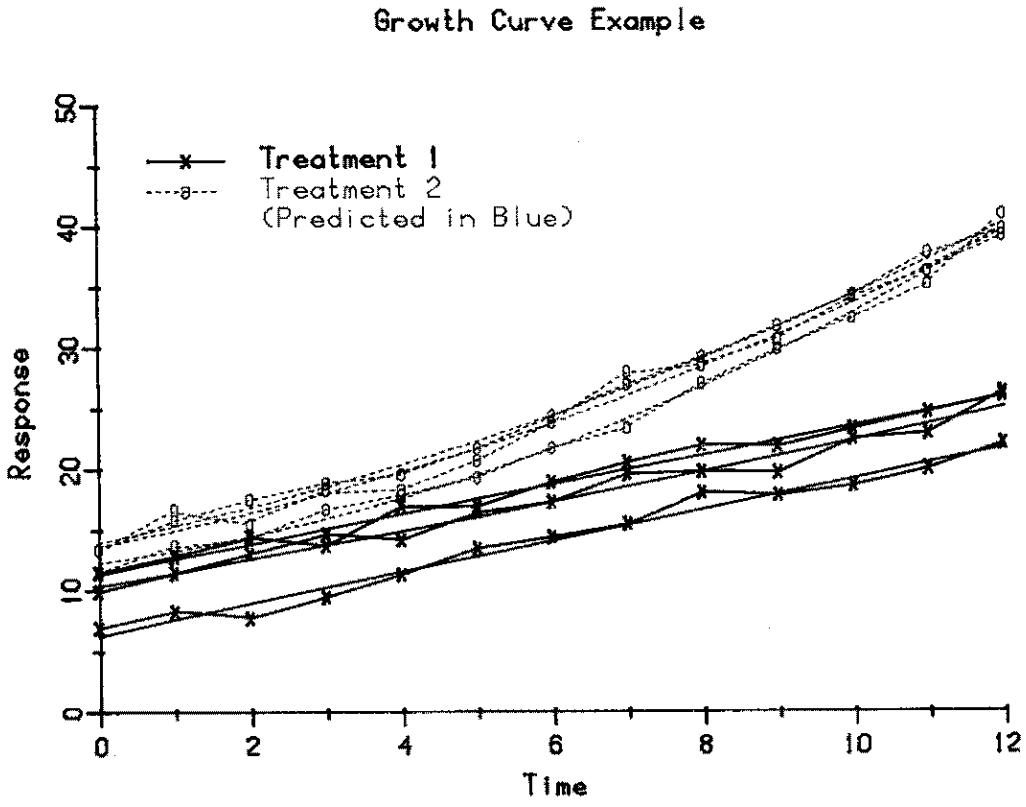
All calculations are performed in PROC MATRIX and the appropriate results are stored in a dataset for later use in printed tables and graphical displays.

Examples presented include a table and graph (using a user written SAS graphics procedure) summarizing growth curve analysis and a table summarizing many regressions based on electricity time-of-use data. SAS statements are given for each example.

## EXAMPLE 1 - GROWTH CURVE DATA

```
*GROWTH CURVE PROBLEM;
*SIMULATE SOME DATA;
DATA T;
RETAIN A0 10 A1 1 A2 .025  B0 12.5 B1 1.2 B2 .085;
ARRAY C0(TREAT) A0 B0; ARRAY C1(TREAT) A1 B1; ARRAY C2(TREAT) A2 B2;
DO TREAT=1 TO 2;
    DO ID=1 TO 3;
    COF0=C0+1.25*NORMAL(13); COF1=C1+.15*NORMAL(0);
    COF2=C2+.009*NORMAL(0);
        DO T=0 TO 12;  T2=T*T;
        Y=COF0+COF1*T+COF2*T2+.75*NORMAL(0);   OUTPUT;
        END;
    END;
END;
KEEP TREAT ID COF0 COF1 COF2 T T2 Y;LABEL Y=Response T=Time;
PROC SYSREG NOPRINT OUTEST=ESTS;BY TREAT ID;
MODEL Y=T T2;
DATA PRED;SET T END=EOF;
OUTPUT; KEEP TREAT ID T Y;  *GET ORIGINAL DATA;
IF ^EOF THEN RETURN;
GETEST:SET ESTS END=EOF2; COF0=INTERCEP; COF1=T; COF2=T2;
TREAT=TREAT+2;
    DO T=0 TO 12 BY .2; *GENERATE PREDICTED CURVES;
    Y=COF0+COF1*T+COF2*T*T;
    OUTPUT;
    END;
IF ^EOF2 THEN GO TO GETEST;
PROC GRAPH NOCLOSE NX=600; GRAPH T Y; CLASS TREAT; SUBCLASS ID;
PUT 'Growth Curve Example' Y=4.5;
KEY 1 TEXT='Treatment 1' CHAR=X LINE=0 COLOR=RED X=1.1 Y=3.6;
KEY 3 LINE=0 COLOR=BLUE;
KEY 2 TEXT='Treatment 2' CHAR=O LINE=6 LSW=.026 COLOR=RED X=1.1;
KEY 4 LINE=6 LSW=.026 COLOR=BLUE;
PUT '(Predicted in Blue)' X=1.1;
```

Figure 1.1.Sysreg and Graphical Presentation of Growth Curve Data




```
* FORMAT TABLE OF COEFFICIENTS OF GROWTH CURVE;
DATA _NULL_;SET ESTS END=EOF;BY TREAT;
FILE PRINT HEADER=H;
IF FIRST.TREAT THEN PUT / @8 TREAT 1. @;
  PUT @18 ID 1. @26 (INTERCEP T T2) (6.3 +9);
IF EOF THEN PUT @3 63*'-';
RETURN;
H : PUT ////@21 'GROWTH CURVE ANALYSIS' /
          @15 'COEFFICIENTS OF FITTED POLYNOMIALS'//
          @39 'COEFFICIENT' @54 'COEFFICIENT' /
          @4 'TREATMENT' @18 'ID' @25 'INTERCEPT' @42 'OF T'
          @56 'OF T**2';
   PUT @3 63*'-';
```

GROWTH CURVE ANALYSIS
COEFFICIENTS OF FITTED POLYNOMIALS

| TREATMENT | ID | INTERCEPT | COEFFICIENT OF T | COEFFICIENT OF T**2 |
|---|---|---|---|---|
| 1 | 1 | 10.433 | 1.066 | 0.012 |
| | 2 | 11.415 | 1.252 | -0.003 |
| | 3 | 6.362 | 1.324 | -0.003 |
| 2 | 1 | 12.339 | 0.834 | 0.122 |
| | 2 | 13.709 | 1.215 | 0.079 |
| | 3 | 13.996 | 1.308 | 0.073 |

Figure 1.2 Table of Growth Curve Analysis

EXAMPLE 2-ELECTRICITY TIME-OF-USE
DATA

```
OPTIONS BLKSIZE=2048 NONUMBER NODATE;
PROC SYSREG NOPRINT OUTEST=ESTOUT OUTSSCP=SSCPOUT;
MODEL Y1-Y32=TAU LPHTLS HWH DRY WASH RANGE LREF LFREZ
            LSQFEET LINCOME;
MACRO KEEPVARS INTERCEP TAU LPHTLS HWH DRY WASH RANGE LREF LFREZ
      LSQFEET LINCOME % *MACRO TO DEFINE VARIABLES IN KEEP  ;
DATA VNAME(KEEP=PCOL TCOL SCOL); *VARIABLE NAME DATASET;
INPUT COL $ @@;
IF COL='INTERCEP' THEN COL='INT'; *VARIABLE NAME ONLY 8 CHARACTERS LONG;
PCOL='P'||COL;TCOL='T'||COL;SCOL='S'||COL;
CARDS;
INTERCEP TAU LPHTLS HWH DRY WASH RANGE LREF LFREZ LSQFEET LINCOME
PROC MATRIX ;

*DEFINE BETA COEFFICIENTS,SUMS OF SQUARES,DEGREES OF FREEDOM,T VALUES,
 F VALUES,P VALUES AND STANDARD ERRORS;
FETCH EST DATA=ESTOUT(KEEP=KEEPVARS _SIGMA_);*-SYSREG ESTIMATES;
NDEP=NROW(EST);                              *-# DEPENDENT VARIABLES;
NIND=NCOL(EST)-1;                            *-# INDEPENDENT VARIABLES;
NTOT=NIND+NDEP;
FETCH SSCP   DATA=SSCPOUT COLNAME=VNAME;     *-SUMS OF SQUARES
                                               AND CROSS PRODUCTS;
BETA=EST(1:NDEP,1:NIND);                     *-REGRESSION COEFFICIENTS;
SIGMA=EST(1:NDEP,NIND+1);                    *-SQRT(MEAN SQUARE ERROR);
EDF=SSCP(1,1) - NIND;                        *-ERROR DEGREES OF FREEDOM;
RDF=NIND - 1;                                *-REGRESSION DEGREES OF FREEDOM;
XSS=SSCP(*,1) || SSCP(*,NDEP+2:NTOT);        *-SUMS OF SQUARES
                                               AND CROSS PRODUCTS OF
                                               INDEPENDENT VARS;

SST=VECDIAG(SSCP(2:NDEP+1,2:NDEP+1))-
    ((SSCP(2:NDEP+1,1)##2)#/SSCP(1,1));      *-TOTAL SUM OF SQUARES;
FREE SSCP EST;
DXPX=SQRT(VECDIAG(INV(XSS(1,*)//XSS(NDEP+2:NTOT,*))));
                                             *-SQRT(DIAGONAL OF X'X);
FREE XSS;
SE=SIGMA#*DXPX';                             *-STANDARD ERROR OF BETAS;
T=BETA #/SE;                                 *-T STATISTIC FOR BETAS;
FREE BETA DXPX;
PVALT=2#(1-PROBT(ABS(T),EDF));               *-P VALUE OF T STATS(2-TAILED);
SSE=(SIGMA##2)#EDF;                          *-ERROR SUM OF SQUARES;
RSQUARE=(SST-SSE)#/SST;                      *-RSQUARE OF MODEL;
F=((SST-SSE)#/RDF)#/(SSE#/EDF);              *-F VALUE OF MODEL;
FREE SST SSE SE;
PVALF=1-PROBF(F,RDF,EDF);                    *-PVALUE OF MODEL;
TRFP=T||PVALT||RSQUARE||F||PVALF;            *-TVALUES,RSQUARE,
                                               FVALUE, PVALUES;

*DEFINE ASTERISKS TO INDICATE SIGNIFICANCE OF T VALUES AND F VALUES;
TSTAR=' '*(PVALT>.05) +'*'*(PVALT<=.05&PVALT>=.01) +
      '**'*(PVALT<=.01);
FSTAR=' '*(PVALF>.05) +'*'*(PVALF<=.05&PVALF>=.01) +
      '**'*(PVALF<=.01);
STAR=TSTAR||FSTAR;                           *-CHARACTER ASTERISKS MATRIX;

*NAME VARIABLES FOR OUTPUTTING MATRICES;
FETCH VARNAME DATA=VNAME TYPE=CHAR;
PNAME=VARNAME(*,1);                          *-NAMES OF PVALUES OF T STATS;
TNAME=VARNAME(*,2);                          *-NAMES OF T VALUES;
FNAME='FSTAR';                               *-ASTERISKS FOR F;
SNAME=VARNAME(*,3)'||FNAME;                  *-NAMES OF ASTERISK VARS;
RFP='RSQUARE' 'FVALUE' 'PROBF' ;
CNAME=TNAME'||PNAME'||RFP ;                  *-NAMES OF NUMERIC VARIBLES;

*OUTPUT MATRICES FOR FORMATTED TABLES;
OUTPUT TRFP OUT=TRFP COLNAME=CNAME;          *-NUMERIC VARIABLES DATASET;
OUTPUT STAR OUT=STAR TYPE=CHAR COLNAME=SNAME;*-CHARACTER VARIABLES DATA SET;
```

Figure 2.1.MATRIX statements

```
*FORMAT TABLES HAVING REGRESSION COEFFICIENTS,RSQUARE,F-VALUES,
 AND ASTERISKS TO INDICATE SIGNIFICANT COEFFICIENTS;
MACRO SMODEL1 SINT STAU SLPHTLS SHWH SDRY SWASH SRANGE SLREF SLFREZ
      SLSQFEET SLINCOME  % *MACRO FOR ASTERISK NAMES;
DATA _NULL_;MERGE EST.MOD1(KEEP=KEEPVARS)
                  TRFP(KEEP=RSQUARE FVALUE )
                  STAR(KEEP=SMODEL1 FSTAR) END=EOF;
FILE PRINT HEADER=H;
MACRO REST @1 TIME $CHAR32. @29
      (LFREZ SLFREZ LSQFEET SLSQFEET LINCOME SLINCOME RSQUARE FVALUE
       FSTAR)
      (8.3 $2. 7.3 $2. 7.3 $2. 7.3 11.3 $2.) %
IF _N_=1 OR _N_=6 OR _N_=13 OR _N_=18 OR _N_=21 OR _N_=24 OR _N_=25
  THEN PUT @1 74*'-';
LENGTH TIME $ 30;
IF _N_=1  THEN TIME='    12pm -  1am   Base     ';
IF _N_=2  THEN TIME='     1am -  2am            ';
IF _N_=3  THEN TIME='     2am -  3am            ';
IF _N_=4  THEN TIME='     3am -  4am            ';
IF _N_=5  THEN TIME='     4am -  5am            ';
IF _N_=6  THEN TIME='     5am -  6am            ';
IF _N_=7  THEN TIME='     6am -  7am            ';
IF _N_=8  THEN TIME='     7am -  8am   Peak1    ';
IF _N_=9  THEN TIME='     8am -  9am            ';
IF _N_=10 THEN TIME='     9am - 10am            ';
IF _N_=11 THEN TIME='    10am - 11am            ';
IF _N_=12 THEN TIME='    11am - 12am            ';
IF _N_=13 THEN TIME='    12am -  1pm   Intermed1';
IF _N_=14 THEN TIME='     1pm -  2pm            ';
IF _N_=15 THEN TIME='     2pm -  3pm            ';
IF _N_=16 THEN TIME='     3pm -  4pm            ';
IF _N_=17 THEN TIME='     4pm -  5pm            ';
IF _N_=18 THEN TIME='     5pm -  6pm   Peak2    ';
IF _N_=19 THEN TIME='     6pm -  7pm            ';
IF _N_=20 THEN TIME='     7pm -  8pm            ';
IF _N_=21 THEN TIME='     8pm -  9pm   Intermed2';
IF _N_=22 THEN TIME='     9pm - 10pm            ';
IF _N_=23 THEN TIME='    10pm - 11pm            ';
IF _N_=24 THEN TIME='    11pm - 12pm   Base     ';
IF _N_=25 THEN TIME='     7am - 12am   Peak1    ';
IF _N_=26 THEN TIME='     5pm -  8pm   Peak2    ';
IF _N_=27 THEN TIME='    12am -  5pm   Intermed1';
IF _N_=28 THEN TIME='     8pm - 11pm   Intermed2';
IF _N_=29 THEN TIME='     Peak Hours            ';
IF _N_=30 THEN TIME='  Intermed Hours           ';
IF _N_=31 THEN TIME='    Base Hours            ';
IF _N_=32 THEN TIME='    All Hours             ';
IF _N_<25 OR _N_=26 OR _N_=28 OR _N_=30 OR _N_=31 THEN PUT REST;
IF _N_=25 OR _N_=27 OR _N_=29 THEN PUT / REST;
IF _N_=32 THEN PUT // REST;
IF EOF THEN PUT @1 74*'-';
RETURN;
 H:PUT //// @11 'ESTIMATED REGRESSION COEFFICIENTS FOR';
   PUT @8  'ALL CUSTOMERS ON AVERAGE FEBRUARY DAY:MODEL I';
   PUT / @1 74*'-';
   PUT
   @5 'TIME OF USE'
   @19 'PERIOD'
   @30 'LFREZ' @39 'LSQFEET' @48 'LINCOME' @57 'R-SQUARE'
   @66 'F-VALUE';RETURN;
?
```

ESTIMATED REGRESSION COEFFICIENTS FOR
ALL CUSTOMERS ON AVERAGE FEBRUARY DAY:MODEL I

| TIME OF USE | PERIOD | LFREZ | LSQFEET | LINCOME | R-SQUARE | F-VALUE |
|---|---|---|---|---|---|---|
| 12pm - 1am | Base | 0.052 | 0.130 | 0.214* | 0.651 | 28.197** |
| 1am - 2am | | 0.034 | 0.107 | 0.173 | 0.665 | 30.018** |
| 2am - 3am | | 0.020 | 0.053 | 0.254* | 0.659 | 29.199** |
| 3am - 4am | | -0.071 | 0.059 | 0.248* | 0.673 | 31.007** |
| 4am - 5am | | -0.084 | 0.152 | 0.101 | 0.624 | 25.009** |
| 5am - 6am | | 0.120 | -0.010 | 0.218* | 0.633 | 26.024** |
| 6am - 7am | | 0.225 | -0.017 | 0.274** | 0.625 | 25.114** |
| 7am - 8am | Peak1 | -0.067 | 0.187 | 0.245* | 0.621 | 24.717** |
| 8am - 9am | | -0.220 | 0.370** | 0.202 | 0.577 | 20.584** |
| 9am - 10am | | -0.203 | 0.349* | 0.107 | 0.532 | 17.163** |
| 10am - 11am | | -0.161 | 0.363* | 0.132 | 0.564 | 19.526** |
| 11am - 12am | | -0.094 | 0.523** | 0.015 | 0.586 | 21.350** |
| 12am - 1pm | Intermed1 | 0.005 | 0.403** | -0.047 | 0.628 | 25.509** |
| 1pm - 2pm | | 0.008 | 0.306* | 0.013 | 0.623 | 24.933** |
| 2pm - 3pm | | 0.008 | 0.317** | 0.034 | 0.620 | 24.627** |
| 3pm - 4pm | | -0.033 | 0.221 | 0.077 | 0.575 | 20.388** |
| 4pm - 5pm | | -0.067 | 0.126 | 0.059 | 0.509 | 15.645** |
| 5pm - 6pm | Peak2 | 0.002 | 0.030 | 0.125 | 0.581 | 19.307** |
| 6pm - 7pm | | -0.059 | 0.065 | 0.199* | 0.596 | 22.253** |
| 7pm - 8pm | | -0.011 | -0.028 | 0.255** | 0.659 | 29.125** |
| 8pm - 9pm | Intermed2 | -0.081 | 0.032 | 0.253** | 0.695 | 34.329** |
| 9pm - 10pm | | -0.041 | 0.005 | 0.221** | 0.692 | 33.959** |
| 10pm - 11pm | | -0.073 | 0.055 | 0.201* | 0.677 | 31.591** |
| 11pm - 12pm | Base | -0.051 | 0.129 | 0.161 | 0.675 | 31.335** |
| 7am - 12am | Peak1 | -0.142 | 0.357** | 0.141 | 0.607 | 23.313** |
| 5pm - 8pm | Peak2 | -0.021 | 0.021 | 0.194** | 0.626 | 25.232** |
| 12am - 5pm | Intermed1 | -0.029 | 0.253* | 0.023 | 0.627 | 25.378** |
| 8pm - 11pm | Intermed2 | -0.060 | 0.040 | 0.230** | 0.719 | 38.559** |
| Peak Hours | | -0.100 | 0.161 | 0.154* | 0.652 | 28.296** |
| Intermed Hours | | -0.037 | 0.156 | 0.110 | 0.695 | 34.375** |
| Base Hours | | 0.075 | 0.056 | 0.195* | 0.715 | 37.883** |
| All Hours | | -0.038 | 0.137 | 0.147* | 0.713 | 37.494** |

Figure 2.2.Table of Regression
Results of Electricity