The program for the following sections follows.

```
dm'log;clear;output;clear';
***********************************************************;
*** EXST7034 Homework Example 1                         ***;
*** Problem from Neter, Wasserman & Kuttner 1989, #2.18   ***;
***********************************************************;
OPTIONS LS=132 PS=256 NOCENTER NODATE NONUMBER nolabel;
filename copier 'C:\Geaghan\Current\EXST7034\Fall2005\Datasets (KNNL)\CH08PR15.csv';
ODS HTML style=minimal rs=none body='C:\Geaghan\Current\EXST7034\Fall2005\SAS\CH01PR20A6.html' ;

Title1 'Assignment 6 : Copier maintenance example';

DATA copier; infile copier missover DSD dlm="," firstobs=2;
      LABEL machines = 'Number of machines serviced'
            minutes = 'Minutes to service machines'
            model = 'Machine model (small=1 or large=0)';
   INPUT minutes machines model;
      interaction = machines * model;
      small = model;
       large = abs(model - 1);
       IntSmall = machines * small;
       IntLarge = machines * large;
datalines; RUN;
;
proc print data=copier; run;

OPTIONS LS=99 PS=256;
PROC REG DATA=copier lineprinter; id interaction;
    title2 'Initial regression model (parallel lines)';
    MODEL  minutes = machines model / clb;
    output out=next1 p=yhat r=e;
run;

OPTIONS LS=99 PS=56;
proc plot data=next1; title2 'Plots from the initial regression model';
     plot e*interaction / vref=0;
     plot e*machines=model / vref=0;
run;

OPTIONS LS=99 PS=256;
PROC REG DATA=copier lineprinter; id interaction;
    title2 'Full Analysis of Covariance';
    MODEL  minutes = machines model interaction / clb;
    output out=next1 p=yhat r=e;
run;
```

8.15a KNNL) **–** The model fitted was $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$. The estimate of β_0 is the intercept for the group receiving the value of “0” for the indicator variable (large machined in this example). The value of β_1 is the common slope, since this model has only a single slope. The value β_2 is the intercept difference between the two categories.

8.15b KNNL) **–** The regression coefficients for the regression model fitted in the first part were estimated as $Y_i = \text{-}0.92247 + 15.04614X_{1i} + 0.75872X_{2i} + \varepsilon_i$. The SAS output for the fitted model is given below.

```
Assignment 6 : Copier maintenance example
Initial regression model

The REG Procedure
Model: MODEL1
Dependent Variable: minutes

Number of Observations Read          45
Number of Observations Used          45


Analysis of Variance
                                       Sum of           Mean
Source                    DF          Squares         Square    F Value    Pr > F
Model                      2            76966          38483     473.94    <.0001
Error                     42       3410.32825       81.19829
Corrected Total           44            80377

Root MSE               9.01101    R-Square     0.9576
Dependent Mean        76.26667    Adj R-Sq     0.9556
Coeff Var             11.81513

Parameter Estimates
                         Parameter       Standard
Variable     DF          Estimate          Error    t Value    Pr > |t|       95% Confidence Limits
Intercept     1          -0.92247        3.09969      -0.30      0.7675       -7.17789         5.33294
machines      1          15.04614        0.49000      30.71      <.0001       14.05728        16.03500
model         1           0.75872        2.77986       0.27      0.7862       -4.85125         6.36870
```

8.15c KNNL) **–** The estimated value was not significantly different from zero, so there does not appear to be a difference in the time required to repair the two machine types. The point estimate for this parameter was 0.75872, suggesting that the small machines took about 46 seconds longer to repair than the larger machines. Of course, this was not significantly different from zero and the 95% confidence interval for this term includes zero,

$P(\text{-}4.85125 \le \beta_2 \le 6.36870) = 0.95$.

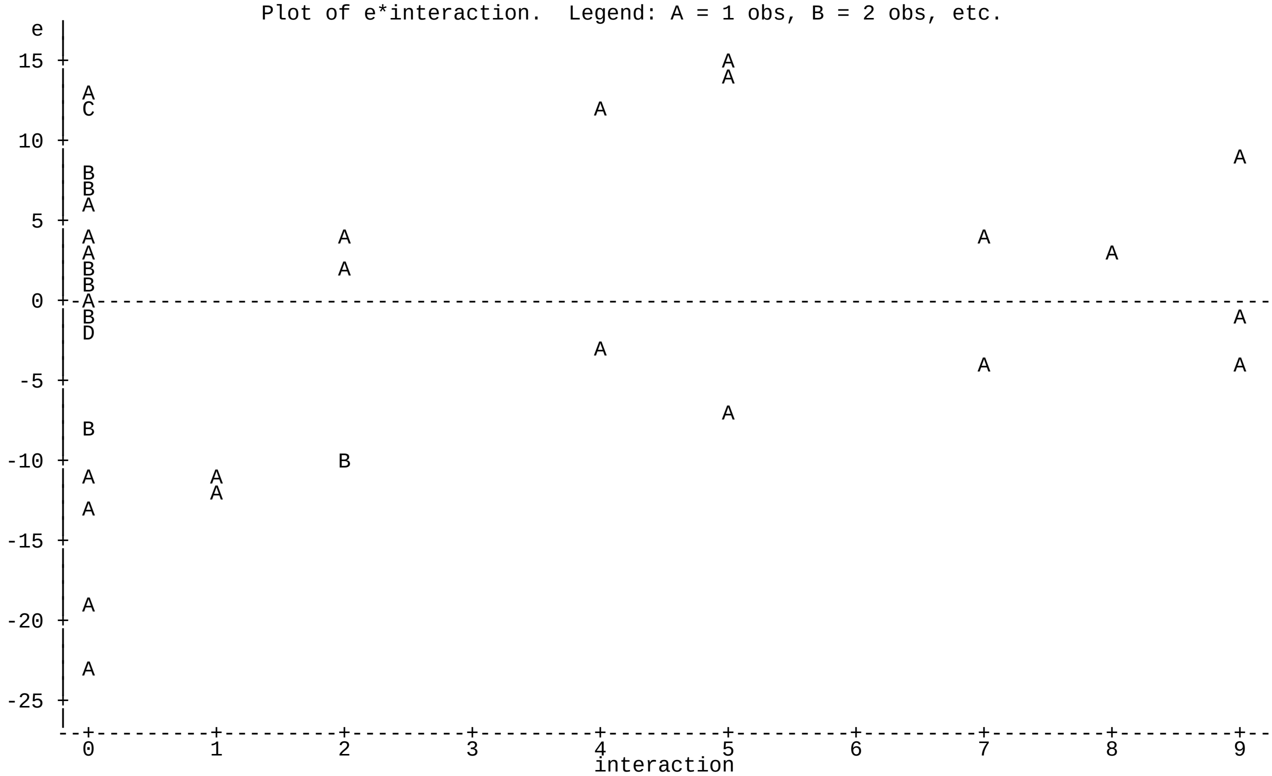
8.15d KNNL) **–** Clearly there is a relationship between the number of machines and the time needed to repair the machines. This test of interest for this model fits the extra SS for model after machine number (SSX2 | X0, X1). If the term X1 was omitted, and the extra SS (SSX2 | X0) was tested the large SS for X1 would be included in the error term and likely no significance could be detected even if a significant difference existed.

8.15e KNNL) **–** The requested residuals are plotted below. The first shows the group equal zero bunched at zero. The other points are for the groups assigned a one, which appears to show a slight increasing trend.

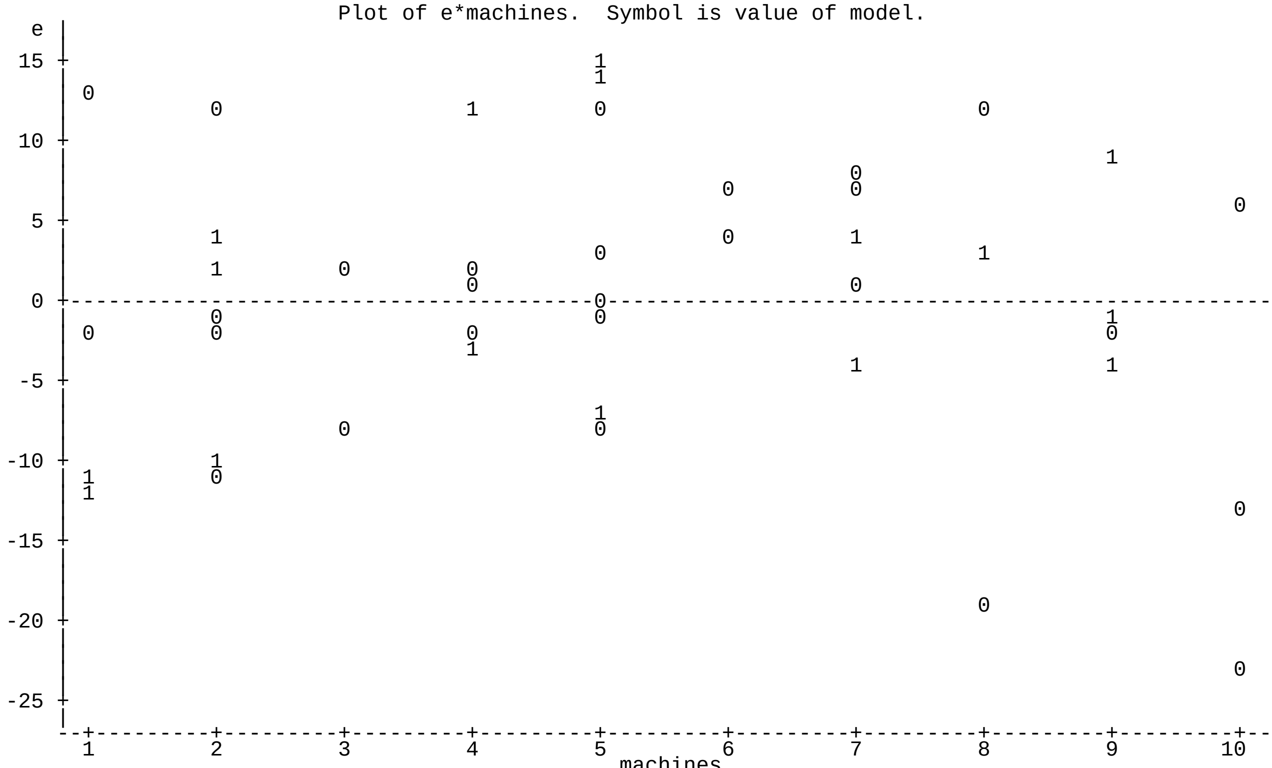
The plot that I suggested can potentially show differing trends for the two categories. In this case there are not obvious trends.

Assignment 6 : Copier maintenance example
Plots from the initial regression model

Plot of e*interaction. Legend: A = 1 obs, B = 2 obs, etc.



Plot of e*machines. Symbol is value of model.



NOTE: 2 obs hidden.

8.19a KNNL) **–** The second model fitted was $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_1 X_{2i} + \varepsilon_i$. The estimate of β_3 is the difference in the slope between the two categories. The estimated regression function was $Y_i = 2.81311 + 14.33941 X_{1i} - 8.14120 X_{2i} + 1.77739 X_1 X_{2i} + \varepsilon_i$.

```
Assignment 6 : Copier maintenance example
Full Analysis of Covariance

The REG Procedure
Model: MODEL1
Dependent Variable: minutes

Number of Observations Read          45
Number of Observations Used          45

Analysis of Variance
                                    Sum of           Mean
Source                    DF       Squares         Square    F Value    Pr > F
Model                      3         77222          25741     334.57    <.0001
Error                     41    3154.43514       76.93744
Corrected Total           44         80377

Root MSE              8.77140    R-Square     0.9608
Dependent Mean       76.26667    Adj R-Sq     0.9579
Coeff Var            11.50096

Parameter Estimates
                     Parameter      Standard
Variable      DF      Estimate         Error   t Value   Pr > |t|      95% Confidence Limits
Intercept      1       2.81311       3.64685      0.77     0.4449      -4.55184      10.17807
machines       1      14.33941       0.61455     23.33     <.0001      13.09830      15.58052
model          1      -8.14120       5.58007     -1.46     0.1522     -19.41037       3.12797
interaction    1       1.77739       0.97459      1.82     0.0755      -0.19084       3.74561
```

8.19b KNNL) **–** From the table above it is clear that the requested test (H_0: $\beta_3 = 0$) would be rejected at the $\alpha = 0.10$ level. The p value for this test was 0.0755, less than the level specified by alpha. Since this test is significant it indicates that the two slopes are not the same, and the two lines will not be parallel.

Can you figure out how to estimate both slopes and both intercepts with standard errors in a single model with pooled variance? **YES!** (demonstrated below)**.**

There are two ways to do this. It requires a means model, so two columns are needed for the two levels of the indicator variable (large and small machines) and two for the interactions. One way to obtain these indicator variables is to use PROC GLM, put MODEL in the classes statement, and let SAS set up the indicator variables. This is probably easiest. The second way is to create the indicator variables in the data step. I did that by adding the following steps.

```
small = model;
large = abs(model - 1);
IntSmall = machines * small;
IntLarge = machines * large;
```

The programs run to obtain the results are given below.

```
PROC REG DATA=copier lineprinter; id interaction;
    title2 'Slopes and intercepts (Means model with REG)';
    MODEL  minutes = small large IntSmall IntLarge / clb noint;
run;

PROC glm DATA=copier; class model;
    title2 'Slopes and intercepts (Means model with GLM)';
    MODEL  minutes = model model*machines / noint solution;
run;
```

```
Assignment 6 : Copier maintenance example
Slopes and intercepts (Means model with REG)

The REG Procedure
Model: MODEL1
Dependent Variable: minutes

Number of Observations Read          45
Number of Observations Used          45

NOTE: No intercept in model. R-Square is redefined.

                             Analysis of Variance
                                   Sum of           Mean
Source                   DF       Squares         Square    F Value    Pr > F
Model                     4        338970          84742    1101.45    <.0001
Error                    41    3154.43514       76.93744
Uncorrected Total        45        342124

Root MSE              8.77140    R-Square     0.9908
Dependent Mean       76.26667    Adj R-Sq     0.9899
Coeff Var            11.50096
```

Parameter Estimates

Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
small	**1**	**-5.32808**	**4.22346**	**-1.26**	**0.2142**	**-13.85754**	**3.20137**
large	**1**	**2.81311**	**3.64685**	**0.77**	**0.4449**	**-4.55184**	**10.17807**
IntSmall	**1**	**16.11680**	**0.75641**	**21.31**	**<.0001**	**14.58920**	**17.64439**
IntLarge	**1**	**14.33941**	**0.61455**	**23.33**	**<.0001**	**13.09830**	**15.58052**

```
Assignment 6 : Copier maintenance example
Slopes and intercepts (Means model with GLM)

The GLM Procedure

   Class Level Information
Class         Levels    Values
model              2    0 1

Number of Observations Read          45
Number of Observations Used          45


                                     Sum of
Source                      DF      Squares     Mean Square    F Value    Pr > F
Model                        4   338969.5649     84742.3912    1101.45    <.0001
Error                       41     3154.4351        76.9374
Uncorrected Total           45   342124.0000

R-Square     Coeff Var      Root MSE    minutes Mean
0.960754      11.50096      8.771399        76.26667

Source                      DF     Type I SS    Mean Square    F Value    Pr > F
model                        2   262153.2038    131076.6019    1703.68    <.0001
machines*model               2    76816.3611     38408.1805     499.21    <.0001

Source                      DF   Type III SS    Mean Square    F Value    Pr > F
model                        2     168.22554       84.11277       1.09    0.3447
machines*model               2   76816.36108    38408.18054     499.21    <.0001
```

Parameter		**Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
model	**0**	**2.81311360**	**3.64684724**	**0.77**	**0.4449**
model	**1**	**-5.32808399**	**4.22346300**	**-1.26**	**0.2142**
machines*model	**0**	**14.33941094**	**0.61455128**	**23.33**	**<.0001**
machines*model	**1**	**16.11679790**	**0.75640644**	**21.31**	**<.0001**

The program for the following sections follows.

```
************************************************************;
*** EXST7034 Homework Example                            ***;
*** Problem from Kutner, Nachtsheim, Neter and Ku 2004,  ***;
*** Patient satisfaction example (9.17)                  ***;
************************************************************;

dm'log;clear;output;clear';
options nodate nocenter nonumber ps=512 ls=85 nolabel;
ODS HTML style=minimal rs=none
body='C:\Geaghan\Current\EXST7034\Fall2005\SAS\PatientSatisfaction02.html' ;

TITLE1 'EXST7034 - Assignment 7, KNNL 9.17 : Patient satisfaction';

DATA Satisfaction; INFILE CARDS MISSOVER;
     LABEL X1 = 'Patients age (years)';
     LABEL X2 = 'Severity of illness (an index)';
     LABEL X3 = 'Anxiety level (an index)';
     LABEL Y = 'Patient satisfaction level';
   INPUT Y X1 X2 X3; X0=1;
CARDS; RUN;
;

proc reg data=Satisfaction; TITLE2 'Stepwise selection';
   MODEL  Y=X2 X1 X3 / selection=stepwise slstay = 0.05 slentry = 0.10;
RUN;

proc reg data=Satisfaction; TITLE2 'Forward selection';
   MODEL  Y=X2 X1 X3 / selection=forward slentry = 0.10;
RUN;

proc reg data=Satisfaction; TITLE2 'Backwards elimination';
   MODEL  Y=X2 X1 X3 / selection=backward slstay = 0.05;
RUN;

PROC REG DATA=Satisfaction outest=parm1 ADJRSQ AIC BIC CP MSE RSQUARE SBC SSE;
    Title2 'RSquare procedure';
   MODEL Y=X2 X1 X3 / selection=rsquare start=1 stop=3 best=5 cp adjrsq aic sbc mse;
RUN;

proc print data=parm1; run;

QUIT;
```

9.17 KNNL preface) **–** This problem, as done in the book, requests that the tests be done at F values of 3.0 for entry to a model and 2.9 to stay in the model. This data set has 46 observations, so the likely regressions of interest will have 44 or fewer d.f. for the MSE. The probability for F1, 44 d.f. = 3.0 is 0.090270535 and for F1, 44 d.f. = 2.9 is 0.095634726. As the regression acquires more terms in the model, the error term will have fewer d.f. Four terms in the model reduce the probabilities to 0.090782826 and 0.096150223 respectively. Since the desired F value cannot be specified in SAS, and since these values correspond to values of approximately 0.10 and 0.05, slentry could be set to 0.10 and slstay to 0.05 if we wanted to simulate the solution from the book. Note that all calculations of probabilities were done in MicroSoft EXCEL.

9.17a KNNL) **–** The program below employed the "selection = stepwise" option in SAS. The selection used two steps and is shown below.

```
EXST7034 - Assignment 7, KNNL 9.17 : Patient satisfaction
Stepwise selection

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read                        47
Number of Observations Used                        46
Number of Observations with Missing Values          1
```

```
Stepwise Selection: Step 1
Variable X1 Entered: R-Square = 0.6190 and C(p) = 8.3536

Analysis of Variance
                                        Sum of           Mean
Source                    DF           Squares         Square    F Value    Pr > F
Model                      1       8275.38885     8275.38885      71.48    <.0001
Error                     44       5093.91550      115.77081
Corrected Total           45            13369

             Parameter     Standard
Variable      Estimate        Error   Type II SS  F Value  Pr > F
Intercept    119.94317      7.08475        33182   286.62  <.0001
X1            -1.52060      0.17985   8275.38885    71.48  <.0001


Stepwise Selection: Step 2
Variable X3 Entered: R-Square = 0.6761 and C(p) = 2.8072

Analysis of Variance
                                        Sum of           Mean
Source                    DF           Squares         Square    F Value    Pr > F
Model                      2       9038.80461     4519.40231      44.88    <.0001
Error                     43       4330.49973      100.70930
Corrected Total           45            13369

             Parameter     Standard
Variable      Estimate        Error   Type II SS  F Value  Pr > F
Intercept    145.94123     11.52509        16149   160.35  <.0001
X1            -1.20047      0.20411   3483.89147    34.59  <.0001
X3           -16.74205      6.08083    763.41576     7.58  0.0086

All variables left in the model are significant at the 0.0500 level.
No other variable met the 0.1000 significance level for entry into the model.

Summary of Stepwise Selection
      Variable   Variable   Number   Partial    Model
Step  Entered    Removed    Vars In  R-Square  R-Square   C(p)    F Value  Pr > F
  1   X1                         1     0.6190    0.6190    8.3536    71.48  <.0001
  2   X3                         2     0.0571    0.6761    2.8072     7.58  0.0086
```

9.17b KNNL) **–** This was not requested, but the issue is addressed in the preface to this example above.

9.17c KNNL) **–** This was not requested.

9.17d KNNL) **–** The program below employed the “selection = backward” option in SAS. The selection used two steps and is shown below.

```
EXST7034 - Assignment 7, KNNL 9.17 : Patient satisfaction
Backwards elimination

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read                          47
Number of Observations Used                          46
Number of Observations with Missing Values            1

Backward Elimination: Step 0
All Variables Entered: R-Square = 0.6822 and C(p) = 4.0000

Analysis of Variance
                                        Sum of           Mean
Source                    DF           Squares         Square    F Value    Pr > F
Model                      3       9120.46367     3040.15456      30.05    <.0001
Error                     42       4248.84068      101.16287
Corrected Total           45            13369

             Parameter     Standard
Variable      Estimate        Error   Type II SS  F Value  Pr > F
Intercept    158.49125     18.12589   7734.51573    76.46  <.0001
X2            -0.44200      0.49197     81.65905     0.81  0.3741
X1            -1.14161      0.21480   2857.55338    28.25  <.0001
X3           -13.47016      7.09966    364.15952     3.60  0.0647
```

```
Backward Elimination: Step 1
Variable X2 Removed: R-Square = 0.6761 and C(p) = 2.8072

Analysis of Variance
                                      Sum of           Mean
Source                   DF          Squares         Square    F Value    Pr > F
Model                     2       9038.80461     4519.40231      44.88    <.0001
Error                    43       4330.49973      100.70930
Corrected Total          45            13369

             Parameter     Standard
Variable      Estimate        Error   Type II SS  F Value  Pr > F
Intercept    145.94123     11.52509        16149   160.35  <.0001
X1            -1.20047      0.20411   3483.89147    34.59  <.0001
X3           -16.74205      6.08083    763.41576     7.58  0.0086

All variables left in the model are significant at the 0.0500 level.

Summary of Backward Elimination
        Variable     Number     Partial      Model
Step    Removed      Vars In    R-Square     R-Square      C(p)      F Value    Pr > F
  1     X2               2       0.0061       0.6761      2.8072       0.81    0.3741
```

9.17 KNNL) – You were asked to compare the results of the two models above (stepwise selection and backward elimination). In this case both approaches finished with the same model.

9.17e KNNL) – This was not requested, and we did not do the “all possible models”. In this case there are only seven possible models. I fitted the RSquare selection model with requests for a number of diagnostics.

```
EXST7034 - Assignment 7, KNNL 9.17 : Patient satisfaction
RSquare procedure

The REG Procedure
Model: MODEL1
Dependent Variable: Y

R-Square Selection Method

Number of Observations Read                          47
Number of Observations Used                          46
Number of Observations with Missing Values            1


Number in                Adjusted
  Model      R-Square    R-Square       C(p)         AIC         BIC         MSE
      1        0.6190      0.6103     8.3536    220.5294    222.1686   115.77081
      1        0.4155      0.4022    35.2456    240.2137    240.2685   177.59980
      1        0.3635      0.3491    42.1123    244.1312    243.9081   193.38737
---------------------------------------------------------------------------------
      2        0.6761      0.6610     2.8072    215.0607    217.4970   100.70930
      2        0.6550      0.6389     5.5997    217.9676    220.0202   107.27907
      2        0.4685      0.4437    30.2471    237.8450    237.5357   165.26498
---------------------------------------------------------------------------------
      3        0.6822      0.6595     4.0000    216.1850    218.9287   101.16287

Number in
  Model      R-Square           SBC           SSE    Variables in Model
      1        0.6190     224.18667    5093.91550    X1
      1        0.4155     243.87101    7814.39120    X3
      1        0.3635     247.78848    8509.04435    X2
---------------------------------------------------------------------
      2        0.6761     220.54658    4330.49973    X1 X3
      2        0.6550     223.45357    4613.00020    X2 X1
      2        0.4685     243.33093    7106.39406    X2 X3
```

```
-----------------------------------------------------------------------
       3      0.6822      223.49953      4248.84068     X2 X1 X3

EXST7034 - Assignment 7, KNNL 9.17 : Patient satisfaction
RSquare procedure

Obs _MODEL_ _TYPE_ _DEPVAR_  _RMSE_ Intercept    X2        X1        X3      Y _IN_
1   MODEL1   PARMS     Y     10.7597  119.943     .      -1.52060     .      -1   1
 2  MODEL1   PARMS     Y      13.3267  146.449    .         .      -37.1167 -1   1
 3  MODEL1   PARMS     Y      13.9064  183.077  -2.40928   .          .      -1   1
 4  MODEL1   PARMS     Y      10.0354  145.941    .      -1.20047 -16.7421 -1   2
 5  MODEL1   PARMS     Y      10.3576  156.672  -0.92079 -1.26765    .      -1   2
 6  MODEL1   PARMS     Y      12.8555  181.573  -1.23948   .      -25.1402 -1   2
 7  MODEL1   PARMS     Y      10.0580  158.491  -0.44200 -1.14161 -13.4702 -1   3

Obs _P_ _EDF_  _SSE_   _MSE_   _RSQ_   _ADJRSQ_   _CP_   _AIC_   _BIC_   _SBC_
1   2    44  5093.92 115.771 0.61898  0.61032  8.3536 220.529 222.169 224.187
 2   2    44  7814.39 177.600 0.41550  0.40221 35.2456 240.214 240.268 243.871
 3   2    44  8509.04 193.387 0.36354  0.34907 42.1123 244.131 243.908 247.788
 4   3    43  4330.50 100.709 0.67609  0.66102  2.8072 215.061 217.497 220.547
 5   3    43  4613.00 107.279 0.65496  0.63891  5.5997 217.968 220.020 223.454
 6   3    43  7106.39 165.265 0.46845  0.44373 30.2471 237.845 237.536 243.331
 7   4    42  4248.84 101.163 0.68219  0.65949  4.0000 216.185 218.929 223.500
```