STAT436 **Homework 9** **Due Thursday 02/27 at 2.30pm**

Directions: Students are to answer both exercises below showing all relevant work. As always, conclusions and justifications are to be given in clear detailed English. Please type up your solutions or write very neatly.

1. Huet, Bouvier, et al (*Statistical Tools for Nonlinear Regression*, p.2) use the 'Pasture Regrowth' data from Ratkowsky (*Nonlinear Regression*, p.88) to fit a certain four-parameter sigmoidal growth model. In the dataset, Y = pasture regrowth (since last grazing), and X = time, and for our present purposes, we can assume that the data are independent measurements. The model function that these authors used to fit the data is rather complicated, and coming up with starting values for the model parameters is not easy, and can only come after we understand the roles they play.

(a) List all the needed assumptions for the following `proc nlin` analysis. Give an example of conditions where the above required independent-measurements assumption would ***not*** be met.

(b) After examining SAS Program B (`proc nlin`), write down the assumed 4-parameter model function that the researchers fit to the data – see the right-hand side of the `model` statement.

(c) Assuming that θ_4 is positive and using algebra and one 'limit', clearly list the roles of the parameters θ_1 and θ_2. (Hint: Which parameters are the upper and lower asymptotes for this model?) Upon examining the graph of the data below, what are your "eyeball estimates" of these two parameters?

(d) In order to obtain NLIN starting values for θ_3 and θ_4, we use the following approach: write down the expression with 'y' on the left-hand side and the above assumed nonlinear model function on the right-hand side (with no error term for now), substitute in our eyeball estimates for the upper and lower asymptotes, and solve so that the new right-hand expression is a linear model in 'log(x)'. Next, relate what you have found to the simple linear regression (`proc reg`) performed in SAS Program A, and use SAS Output A to write down the starting values for θ_3 and θ_4. Verify that these starting values (or approximations to these) are used in SAS Program B.

(e) Using SAS Output B (`proc nlin`), report the estimate of σ^2 here.

(f) Using SAS Output B (`proc nlin`), do a two-sided Wald test that θ_4 = 3 using **α = 1%**. Redo this ***Wald test*** using **α = 5%**. Clearly report your conclusions in each case.

(g) Repeat both of the tests done in part (f) but using ***Likelihood Ratio tests*** instead.

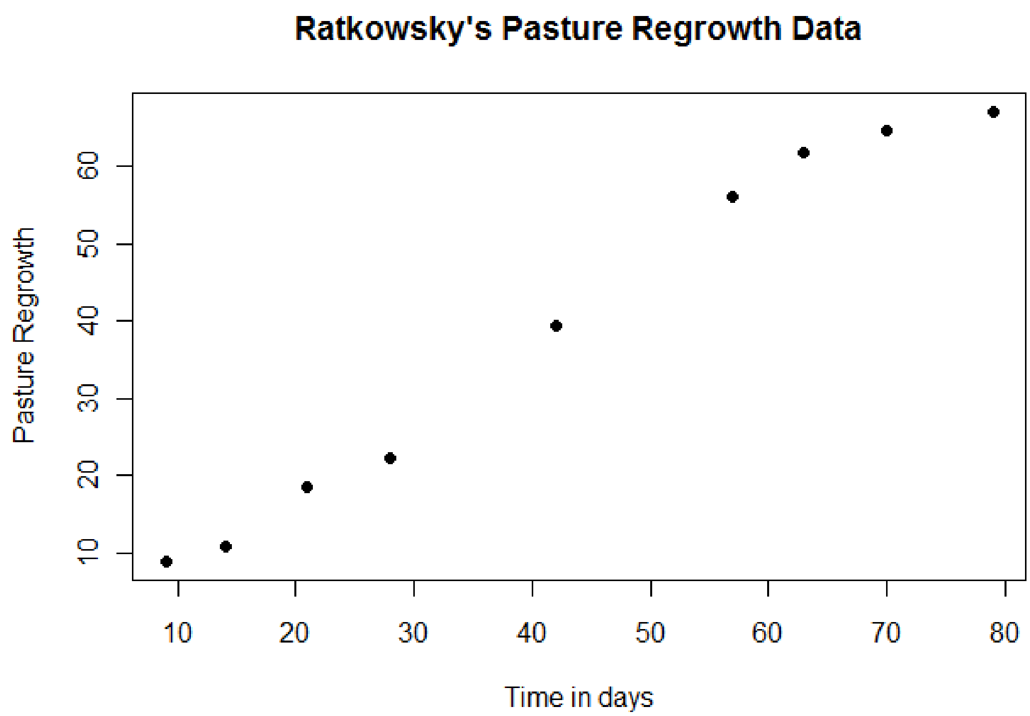
(h) In examining the listing of the residuals in Output C and the Residual Plot, it is apparent that one of the residuals (at x = 21) may be 'large'. If the `proc nlin` were to be rerun with this potential outlier removed, would the estimate of the lower asymptote increase or decrease?

2. In *Nonlinear Regression Analysis and its Applications* (1988, p.269), Bates and Watts report data from Treloar (1974) regarding the "velocity" of an enzymatic reaction. The number of counts per minute of radioactive product from the reaction was measured as a function of substrate concentration (ppm), and from these counts the initial rate, or "velocity," of the reaction was calculated (counts/min^2). The experiment was conducted once with the enzyme treated with puromycin (variable 'treat' = "yes") and once with the enzyme untreated ('treat' = "no"). The velocity is assumed to depend on the substrate concentration according to the Michaelis-Menton (MM_2) equation. In the word of the authors, **it has been hypothesized that the "ultimate velocity parameter" (θ_1) should be affected by introduction of the Puromycin, but not necessarily the "half-velocity parameter" (θ_2).**

(a) Clearly list all the needed assumptions for the following `proc nlin` analyses ***in the context of this situation/exercise***.
(b) Write down the model function fit in SAS Program A and clearly indicate the roles of θ_3 and θ_4 here.
(c) Give estimates for the MM_2 model parameters (upper asymptote and LD_{50}) for both the treated and untreated curves; report these in a 2x2 box.
(d) Test both of the hypotheses indicated by the authors above highlighted claim using **Wald** hypothesis tests (one at a time). In both cases and using the model in SAS Program A, report the hypotheses, test statistics, degrees of freedom, p-values and conclusions. Approximate the p-values here as best you can and also in the next two parts.
(e) Using the full-and-reduced (**likelihood-based**) F-test, test whether the **half-velocity parameters** are the same, reporting the calculated test statistic, degrees of freedom, p-value and your conclusion.
(f) Using the full-and-reduced (**likelihood-based**) F-test and assuming the half-velocity parameters are indeed the same, test whether the **ultimate velocity parameters** are the same, reporting the calculated test statistic, degrees of freedom, p-value and your conclusion.
(g) Finally, compare the model function in Program/Output A with that in Program/Output D – are they equivalent? Why/why not? In what way are the approaches different? In which situation(s) is Program/Output A preferred, and in which situation(s) is Program/Output D preferred? Be clear in your explanation.

Homework 5a Appendix

Exercise 1 Graph



Exercise 1 SAS Program A

```
data one;
  do x=9,14,21,28,42,57,63,70,79;
    input y @@; y=y/100; output;
```

```
  end; datalines;
893 1080 1859 2233 3935 5611 6173 6462 6708
;
data two;
  set one;
  ny=log(-log((70-y)/65)); nx=log(x);
proc reg data=two;
  model ny=nx;
run;
```

Exercise 1 SAS Output A

```
                          The REG Procedure
                            Model: MODEL1
                        Dependent Variable: ny

                Number of Observations Read          9
                Number of Observations Used          9

                        Analysis of Variance
                                 Sum of           Mean
Source                  DF      Squares         Square    F Value    Pr > F
Model                    1     16.87291       16.87291     550.90    <.0001
Error                    7      0.21439        0.03063
Corrected Total          8     17.08730

             Root MSE              0.17501    R-Square     0.9875
             Dependent Mean       -0.53872    Adj R-Sq     0.9857
             Coeff Var           -32.48571

                         Parameter Estimates
                      Parameter       Standard
     Variable     DF   Estimate          Error    t Value    Pr > |t|
     Intercept     1   -7.18999        0.28932     -24.85      <.0001
     nx            1    1.88511        0.08032      23.47      <.0001
```

Exercise 1 SAS Program B

```
proc nlin data=one;
  parms th1=70 th2=65 th3=-7 th4=2;
  model y=th1-th2*exp(-exp(th3)*(x**th4));
  output out=three r=resids p=preds;
run;
```

Exercise 1 SAS Output B

```
                          The NLIN Procedure
                         Dependent Variable y
                         Method: Gauss-Newton
                           Iterative Phase
                                                              Sum of
        Iter          th1          th2          th3          th4      Squares
```

0	70.0000	65.0000	-7.0000	2.0000	832.1
1	68.9082	64.3067	-6.1978	1.5712	276.5
2	63.5647	52.7166	-9.3162	2.4354	85.3555
3	69.8974	61.6943	-9.0989	2.3440	9.7022
4	69.9211	61.6539	-9.2072	2.3777	8.3768
5	69.9575	61.6846	-9.2082	2.3776	8.3759
6	69.9552	61.6815	-9.2089	2.3778	8.3759
7	69.9552	61.6815	-9.2089	2.3778	8.3759

NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	7
Observations Read	9
Observations Used	9
Observations Missing	0

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	3	4639.7	1546.6	923.22	<.0001
Error	5	8.3759	1.6752		
Corrected Total	8	4648.1			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
th1	69.9552	2.3620	63.8835	76.0269
th2	61.6815	3.1927	53.4744	69.8885
th3	-9.2089	0.8173	-11.3098	-7.1080
th4	2.3778	0.2210	1.8098	2.9459

Approximate Correlation Matrix

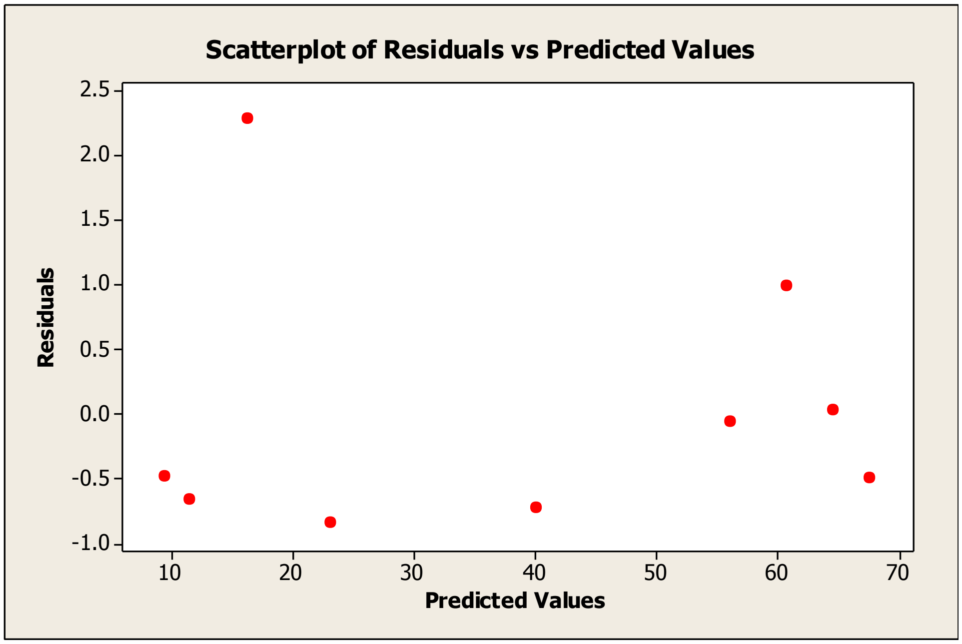
	th1	th2	th3	th4
th1	1.0000000	0.9251613	0.7095438	-0.7658736
th2	0.9251613	1.0000000	0.8615146	-0.8906628
th3	0.7095438	0.8615146	1.0000000	-0.9955752
th4	-0.7658736	-0.8906628	-0.9955752	1.0000000

Exercise 1 SAS Program and Output C

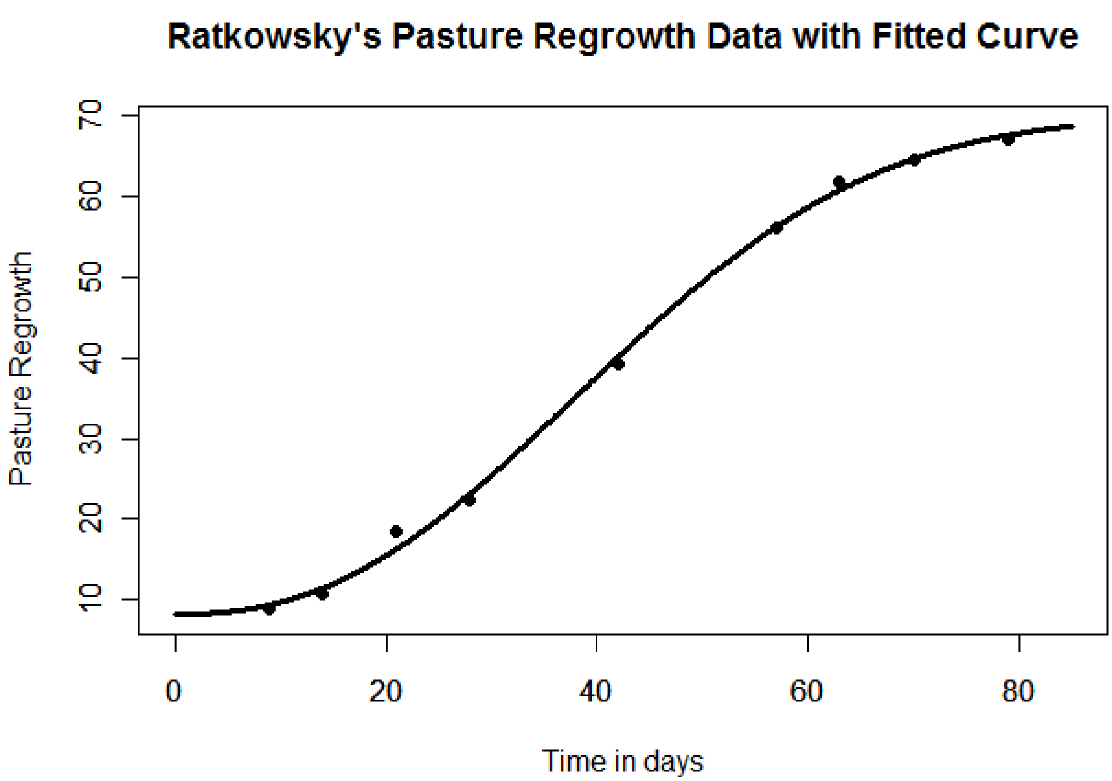
```
proc print noobs;
run;
```

x	y	preds	resids
9	8.93	9.4107	-0.48069
14	10.80	11.4693	-0.66931
21	18.59	16.3057	2.28432
28	22.33	23.1737	-0.84374
42	39.35	40.0846	-0.73458
57	56.11	56.1766	-0.06655
63	61.73	60.7442	0.98581
70	64.62	64.5949	0.02506
79	67.08	67.5803	-0.50032

Exercise 1 SAS Residual Plot



Exercise 1 Fitted Nonlinear Model and Data



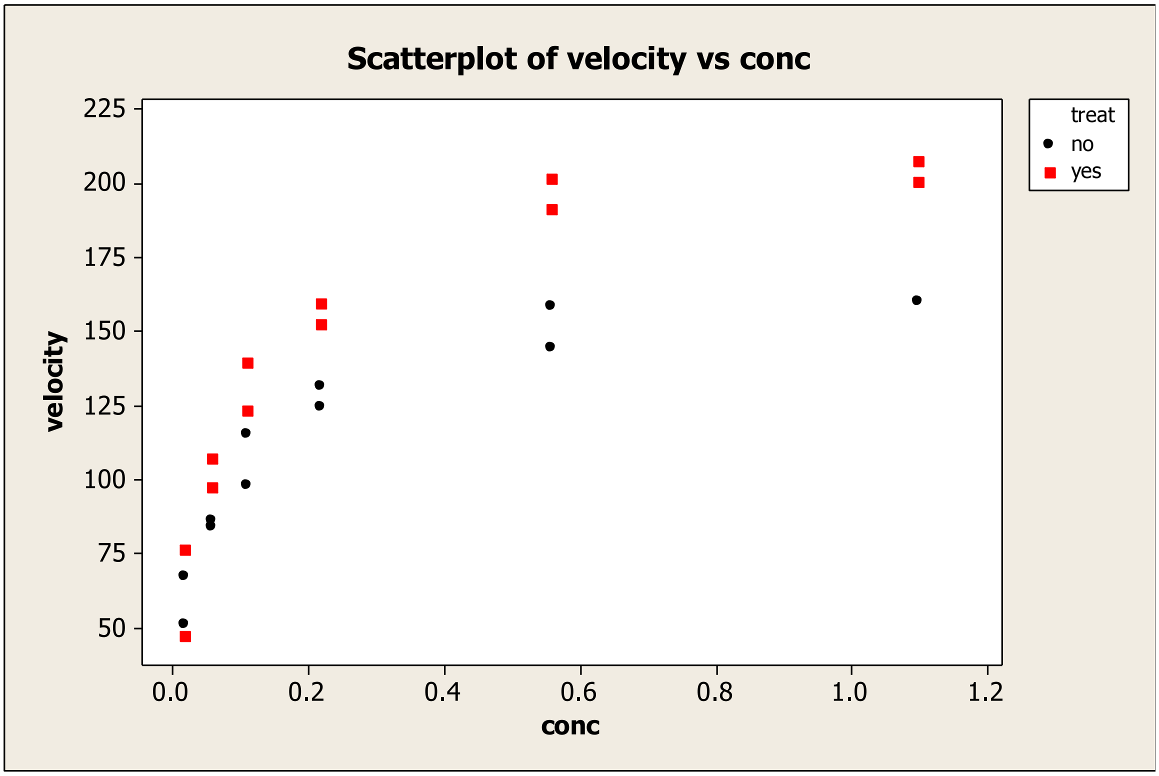
Exercise 1 SAS Program D

```
proc nlin data=one;
  parms th1=70 th2=55 th3=-12; th4=3;
  model y=th1-th2*exp(-exp(th3)*(x**th4));
run;
```

Exercise 1 SAS Output D

```
                              The NLIN Procedure
                             Dependent Variable y
                             Method: Gauss-Newton

                               Iterative Phase
                                                        Sum of
              Iter          th1         th2         th3     Squares
                 0      70.0000     55.0000    -12.0000       112.9
                 1      64.8344     53.7243    -11.5162     28.9029
                 2      66.7202     56.0710    -11.5451     19.2408
                 3      66.7573     56.0959    -11.5467     19.2384
                 4      66.7590     56.0969    -11.5469     19.2384
                 5      66.7592     56.0969    -11.5469     19.2384
        NOTE: Convergence criterion met.

                              Estimation Summary
                      Method                    Gauss-Newton
                      Iterations                           5
                      Observations Read                    9
                      Observations Used                    9
                      Observations Missing                 0

                                    Sum of        Mean               Approx
   Source                    DF    Squares      Square    F Value    Pr > F
   Model                      2     4628.8      2314.4     721.81    <.0001
   Error                      6    19.2384      3.2064
   Corrected Total            8     4648.1

                                    Approx
          Parameter      Estimate    Std Error    Approximate 95% Confidence Limits
          th1             66.7592       1.4695     63.1633     70.3550
          th2             56.0969       1.6005     52.1806     60.0133
          th3            -11.5469       0.0999    -11.7912    -11.3026

                       Approximate Correlation Matrix
                          th1             th2             th3
              th1     1.0000000       0.7565905      -0.7648290
              th2     0.7565905       1.0000000      -0.3861020
              th3    -0.7648290      -0.3861020       1.0000000
```

Exercise 2 Graph and Listing of Data



conc	treat	dummy	dumyes	dumno	velocity
0.02	no	0	0	1	67
0.02	no	0	0	1	51
0.06	no	0	0	1	84
0.06	no	0	0	1	86
0.11	no	0	0	1	98
0.11	no	0	0	1	115
0.22	no	0	0	1	131
0.22	no	0	0	1	124
0.56	no	0	0	1	144
0.56	no	0	0	1	158
1.10	no	0	0	1	160
0.02	yes	1	1	0	76
0.02	yes	1	1	0	47
0.06	yes	1	1	0	97
0.06	yes	1	1	0	107
0.11	yes	1	1	0	123
0.11	yes	1	1	0	139
0.22	yes	1	1	0	159
0.22	yes	1	1	0	152
0.56	yes	1	1	0	191
0.56	yes	1	1	0	201
1.10	yes	1	1	0	207
1.10	yes	1	1	0	200

Exercise 2 SAS Program and Output A

```
proc nlin data=one;
  parms th1=150 th2=0.10 th3=0 th4=0;
  model velocity=((th1+th3*dummy)*conc)/(th2+th4*dummy+conc);
run;
```

```
                         The NLIN Procedure
                    Dependent Variable velocity
                       Method: Gauss-Newton

                          Iterative Phase
                                                                Sum of
        Iter         th1         th2         th3         th4    Squares
           0       150.0      0.1000           0           0    45433.4
           1       158.5      0.0239     53.5635      0.0151     8794.1
           2       155.1      0.0376     53.7329      0.0182     2379.9
           3       158.9      0.0453     53.0719      0.0175     2066.1
           4       160.0      0.0473     52.5673      0.0167     2055.3
           5       160.2      0.0476     52.4332      0.0165     2055.1
           6       160.3      0.0477     52.4085      0.0164     2055.1
           7       160.3      0.0477     52.4044      0.0164     2055.1
           8       160.3      0.0477     52.4038      0.0164     2055.1
        NOTE: Convergence criterion met.

                         Estimation Summary
                   Method                   Gauss-Newton
                   Iterations                          8
                   Observations Read                  23
                   Observations Used                  23
                   Observations Missing                0

             NOTE: An intercept was not specified for this model.

                                       Sum of        Mean                 Approx
   Source                      DF     Squares      Square    F Value    Pr > F
   Model                        4      417562      104390     965.14    <.0001
   Error                       19      2055.1       108.2
   Uncorrected Total           23      419617

                                      Approx
   Parameter        Estimate       Std Error    Approximate 95% Confidence Limits
   th1                 160.3          6.8960         145.8         174.7
   th2                0.0477         0.00828        0.0304        0.0650
   th3               52.4038          9.5510       32.4135       72.3942
   th4                0.0164          0.0114      -0.00751        0.0403

                      Approximate Correlation Matrix
                  th1               th2               th3               th4
     th1     1.0000000         0.7768268        -0.7220184        -0.5628691
     th2     0.7768268         1.0000000        -0.5608833        -0.7245748
     th3    -0.7220184        -0.5608833         1.0000000         0.7712219
     th4    -0.5628691        -0.7245748         0.7712219         1.0000000
```

Exercise 2 SAS Program and Output B

```
proc nlin data=one;
  parms th1=150 th2=0.10 th3=0; th4=0;
  model velocity=((th1+th3*dummy)*conc)/(th2+th4+conc);
run;
```

```
                         The NLIN Procedure
                     Dependent Variable velocity
                        Method: Gauss-Newton

                           Iterative Phase
                                                          Sum of
           Iter          th1         th2         th3      Squares
              0        150.0      0.1000           0      45433.4
              1        161.6      0.0321     47.9824       9612.9
              2        162.3      0.0482     41.2622       2595.2
              3        165.7      0.0561     41.6938       2251.4
              4        166.5      0.0577     41.9696       2241.1
              5        166.6      0.0579     42.0189       2240.9
              6        166.6      0.0580     42.0251       2240.9
              7        166.6      0.0580     42.0259       2240.9
              8        166.6      0.0580     42.0260       2240.9
      NOTE: Convergence criterion met.

                          Estimation Summary
                    Method                    Gauss-Newton
                    Iterations                           8
                    Observations Read                   23
                    Observations Used                   23
                    Observations Missing                 0

           NOTE: An intercept was not specified for this model.

                                    Sum of        Mean               Approx
   Source                    DF    Squares      Square    F Value    Pr > F
   Model                      3     417376      139125    1241.70    <.0001
   Error                     20     2240.9       112.0
   Uncorrected Total         23     419617

                                    Approx
   Parameter      Estimate       Std Error    Approximate 95% Confidence Limits
   th1               166.6        5.8074        154.5        178.7
   th2              0.0580       0.00591       0.0456       0.0703
   th3             42.0260        6.2721      28.9426      55.1093

                       Approximate Correlation Matrix
                          th1               th2               th3
           th1      1.0000000         0.6112817        -0.5405580
           th2      0.6112817         1.0000000         0.0644066
           th3     -0.5405580         0.0644066         1.0000000
```

Exercise 2 SAS Program and Output C

```
proc nlin data=one;
  parms th1=150 th2=0.10; th3=0;
  model velocity=((th1+th3*dummy)*conc)/(th2+conc);
run;
```

```
                         The NLIN Procedure
                    Dependent Variable velocity
                       Method: Gauss-Newton

                          Iterative Phase
                                                      Sum of
               Iter          th1          th2       Squares
                  0        150.0       0.1000       45433.4
                  1        190.7       0.0398       11454.1
                  2        187.5       0.0536        7424.5
                  3        190.1       0.0591        7280.7
                  4        190.7       0.0602        7276.6
                  5        190.8       0.0604        7276.5
                  6        190.8       0.0604        7276.5
                  7        190.8       0.0604        7276.5
      NOTE: Convergence criterion met.

                                   Sum of        Mean               Approx
  Source                    DF    Squares      Square    F Value    Pr > F
  Model                      2     412340      206170     595.00    <.0001
  Error                     21     7276.5       346.5
  Uncorrected Total         23     419617

                                                 Approx
  Parameter      Estimate     Std Error    Approximate 95% Confidence Limits
  th1               190.8        8.7646       172.6        209.0
  th2              0.0604        0.0108      0.0380       0.0828

                     Approximate Correlation Matrix
                                  th1              th2
                  th1       1.0000000        0.7757154
                  th2       0.7757154        1.0000000
```

Exercise 2 SAS Program and Output D (after introducing TWO dummy variables in the dataset for puromycin YES called 'dumyes' and for puromycin NO called 'dumno')

```
proc nlin data=one;
  parms th1yes=150 th1no=150 th2yes=0.10 th2no=0.10;
  th1=th1yes*dumyes+th1no*dumno;
  th2=th2yes*dumyes+th2no*dumno;
  model velocity=(th1*conc)/(th2+conc);
run;
```

The NLIN Procedure
Dependent Variable velocity
Method: Gauss-Newton

Iterative Phase

Iter	th1yes	th1no	th2yes	th2no	Sum of Squares
0	150.0	150.0	0.1000	0.1000	45433.4
1	212.0	158.5	0.0390	0.0239	8794.1
2	208.8	155.1	0.0558	0.0376	2379.9
3	212.0	158.9	0.0628	0.0453	2066.1
4	212.6	160.0	0.0640	0.0473	2055.3
5	212.7	160.2	0.0641	0.0476	2055.1
6	212.7	160.3	0.0641	0.0477	2055.1
7	212.7	160.3	0.0641	0.0477	2055.1
8	212.7	160.3	0.0641	0.0477	2055.1

NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	8
Observations Read	23
Observations Used	23
Observations Missing	0

NOTE: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	4	417562	104390	965.14	<.0001
Error	19	2055.1	108.2		
Uncorrected Total	23	419617			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
th1yes	212.7	6.6081	198.9	226.5
th1no	160.3	6.8960	145.8	174.7
th2yes	0.0641	0.00788	0.0476	0.0806
th2no	0.0477	0.00828	0.0304	0.0650

Approximate Correlation Matrix

	th1yes	th1no	th2yes	th2no
th1yes	1.0000000	0.0000000	0.7650837	0.0000000
th1no	0.0000000	1.0000000	0.0000000	0.7768268
th2yes	0.7650837	0.0000000	1.0000000	0.0000000
th2no	0.0000000	0.7768268	0.0000000	1.0000000