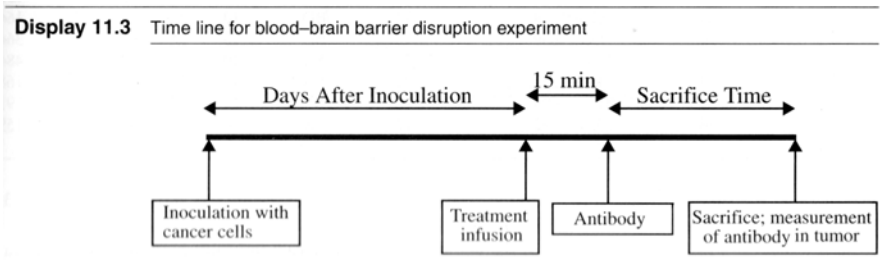
Chapter 11 : Model checking and refinement

An example: Blood-brain barrier study on rats

This study investigates the permeability of the blood-brain barrier to medication. Rats were given cells that would cause brain tumors then given a "barrier disruption" (BD) chemical or a saline solution placebo. Fifteen minutes later rats were administered a therapeutic antibody. After a set time rats were sacrificed and the brain examined for the antibody.



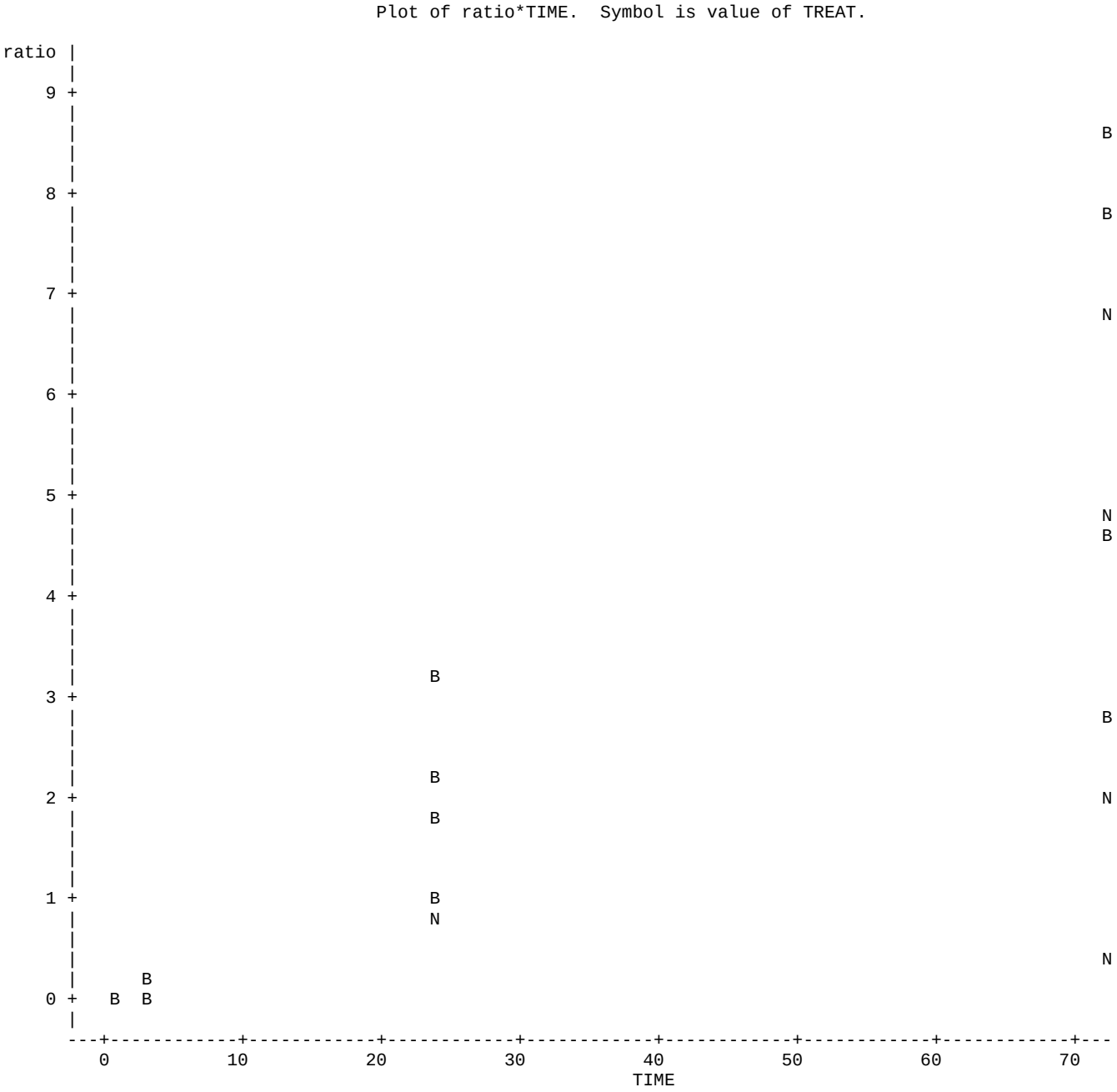

```
1          ******************************************************;
2          *** Blood-brain barrier study on rats               ***;
3          *** This study investigates the permeability of the ***;
4          *** blood-brain barrier to medication. Rats were    ***;
5          *** given cells that would cause brain tumors then  ***
6          *** given a "barrier disruption" (BD) chemical or a ***;
7          *** saline solution placebo.  Fifteen minutes later ***;
8          *** rats were administered a therapeutic antibody.  ***;
9          *** After a set time rats were sacrificed and the   ***;
10         *** brain examined for the antibody                 ***;
11         ******************************************************;
12
13         dm'log;clear;output;clear';
14         options nodate nocenter nonumber ps=512 ls=99 nolabel;
15         ODS HTML style=minimal rs=none
15       ! body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\Barrier01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\Barrier01.html
16
17         Title1 'Chapter 11 : Alcohol Barrier in men and women';
18         filename input1 'C:\Geaghan\Current\EXST3201\Datasets\ASCII\case1102.csv';
19
20         data Barrier; infile input1 missover DSD dlm="," firstobs=2;
21            input BRAIN LIVER TIME TREAT $ DAYS SEX $ WEIGHT LOSS TUMOR;
22              label brain = 'Brain tumor cell count (per gm)'
23                    liver = 'Liver cell count (per gm)'
24                    treat = 'barrier disruptor versus control'
25                    time = 'Sacrifice time (in hours)'
26                    days = 'Days post inoculation'
27                    sex = 'Sex of the rat'
28                    weight = 'Initial Weight'
29                    loss = 'Weight loss'
30                    tumor = 'Tumor weight';
31              ratio = brain / liver;
32         datalines;
NOTE: The infile INPUT1 is:
      File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\case1102.csv,
      RECFM=V,LRECL=256
NOTE: 34 records were read from the infile INPUT1.
      The minimum record length was 34.
      The maximum record length was 41.
NOTE: The data set WORK.BARRIER has 34 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds
33         run;
34
35         PROC PRINT DATA=Barrier; TITLE2 'Data Listing'; RUN;
NOTE: There were 34 observations read from the data set WORK.BARRIER.
NOTE: The PROCEDURE PRINT printed page 1.
```

```
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.13 seconds
      cpu time            0.06 seconds
36


NOTE: The PROCEDURE REG printed pages 13-22.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.15 seconds
      cpu time            0.07 seconds

Chapter 11 : Alcohol Barrier in men and women
Data Listing
```

Obs	BRAIN	LIVER	TIME	TREAT	DAYS	SEX	WEIGHT	LOSS	TUMOR	ratio
1	41081	1456164	0.5	BD	10	F	239	5.9	221	0.02821
2	44286	1602171	0.5	BD	10	F	225	4.0	246	0.02764
3	102926	1601936	0.5	BD	10	F	224	-4.9	61	0.06425
4	25927	1776411	0.5	BD	10	F	184	9.8	168	0.01460
5	42643	1351184	0.5	BD	10	F	250	6.0	164	0.03156
6	31342	1790863	0.5	NS	10	F	196	7.7	260	0.01750
7	22815	1633386	0.5	NS	10	F	200	0.5	27	0.01397
8	16629	1618757	0.5	NS	10	F	273	4.0	308	0.01027
9	22315	1567602	0.5	NS	10	F	216	2.8	93	0.01424
10	77961	1060057	3.0	BD	10	F	267	2.6	73	0.07354
11	73178	715581	3.0	BD	10	F	263	1.1	25	0.10226
12	76167	620145	3.0	BD	10	F	228	0.0	133	0.12282
13	123730	1068423	3.0	BD	9	F	261	3.4	203	0.11581
14	25569	721436	3.0	NS	9	F	253	5.9	159	0.03544
15	33803	1019352	3.0	NS	10	F	234	0.1	264	0.03316
16	24512	667785	3.0	NS	10	F	238	0.8	34	0.03671
17	50545	961097	3.0	NS	9	F	230	7.0	146	0.05259
18	50690	1220677	3.0	NS	10	F	207	1.5	212	0.04153
19	84616	48815	24.0	BD	10	F	254	3.9	155	1.73340
20	55153	16885	24.0	BD	10	M	256	-4.7	190	3.26639
21	48829	22395	24.0	BD	10	M	247	-2.8	101	2.18035
22	89454	83504	24.0	BD	11	F	198	4.2	214	1.07125
23	37928	20323	24.0	NS	10	F	237	2.5	224	1.86626
24	12816	15985	24.0	NS	10	M	293	3.1	151	0.80175
25	23734	25895	24.0	NS	10	M	288	9.7	285	0.91655
26	31097	33224	24.0	NS	11	F	236	5.9	380	0.93598
27	35395	4142	72.0	BD	11	F	251	4.1	39	8.54539
28	18270	2364	72.0	BD	10	F	223	4.0	153	7.72843
29	5625	1979	72.0	BD	10	M	298	12.8	164	2.84234
30	7497	1659	72.0	BD	10	M	260	7.3	364	4.51899
31	6250	928	72.0	NS	10	M	272	11.0	484	6.73491
32	11519	2423	72.0	NS	11	F	226	2.2	168	4.75402
33	3184	1608	72.0	NS	10	M	249	-4.4	191	1.98010
34	1334	3242	72.0	NS	10	F	240	6.7	159	0.41147

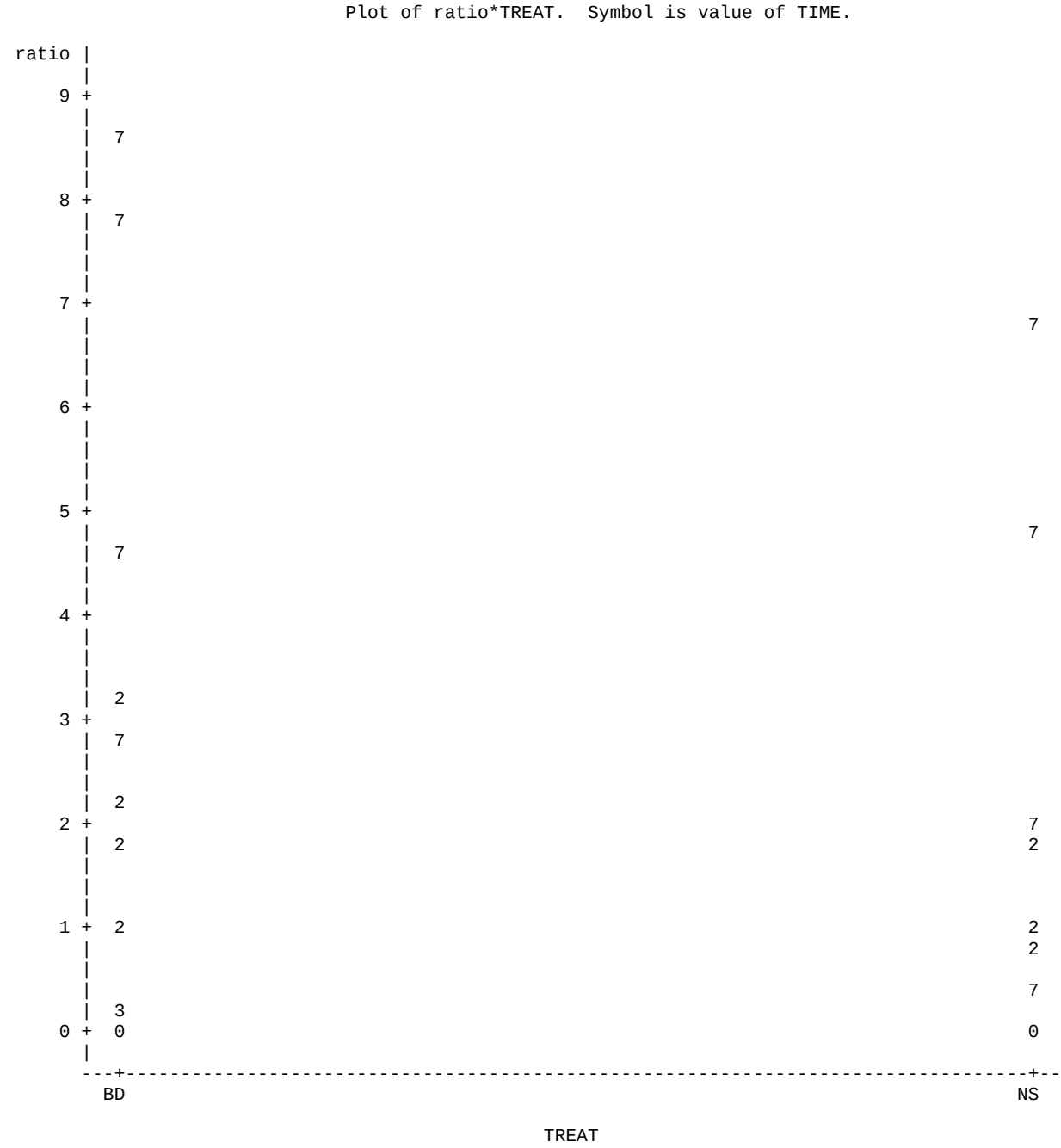
```
37         options ps=60 ls=132;
38         proc plot data=Barrier;
39            TITLE2 'Plot of the raw data with treatment variable';
40            plot ratio * time = treat;
41            plot ratio * treat = time;
42         RUN;
42       !      OPTIONS PS=256;
NOTE: There were 34 observations read from the data set WORK.BARRIER.
NOTE: The PROCEDURE PLOT printed pages 2-3.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.08 seconds
      cpu time            0.01 seconds
```

Chapter 11 : Alcohol Barrier in men and women
Plot of the raw data with treatment variable



NOTE: 18 obs hidden.

```
Chapter 11 : Alcohol Barrier in men and women
Plot of the raw data with treatment variable
```

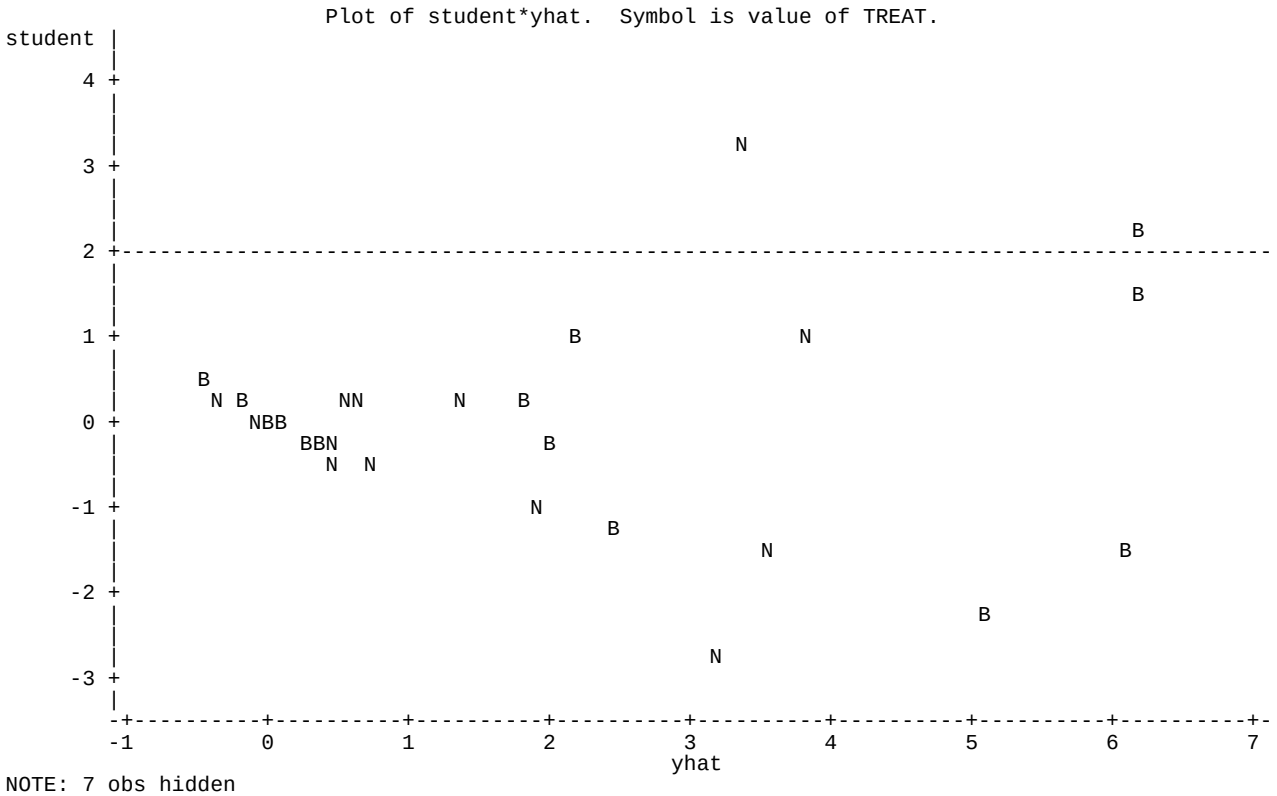



```
NOTE: 16 obs hidden.


43         PROC GLM DATA=Barrier; class treat sex;
44         Title2 'Fit of ratio on indicator variables with GLM';
45            MODEL ratio = time treat time*treat days sex weight loss tumor / solution;
46            output out=next1 r=resid p=yhat lclm=lclm uclm=uclm lcl=lcli ucl=ucli
47                   student=student rstudent=rstudent cookd=cookd h=leverage dffits=dffits;
48         RUN;
NOTE: The data set WORK.NEXT1 has 34 observations and 21 variables.
NOTE: The PROCEDURE GLM printed pages 4-5.
NOTE: PROCEDURE GLM used (Total process time):
      real time           0.25 seconds
      cpu time            0.13 seconds
```

```
Chapter 11 : Alcohol Barrier in men and women
Fit of ratio on indicator variables with GLM

The GLM Procedure

Class Level Information
Class         Levels    Values
TREAT              2    BD NS
SEX                2    F M

Number of Observations Read          34
Number of Observations Used          34

Dependent Variable: ratio
                                    Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F
Model                        8     137.5198271      17.1899784       9.68    <.0001
Error                       25      44.3885877       1.7755435
Corrected Total             33     181.9084147

R-Square     Coeff Var      Root MSE    ratio Mean
0.755984      88.61808      1.332495      1.503638

Source                      DF       Type I SS     Mean Square    F Value    Pr > F
TIME                         1     119.3889310     119.3889310      67.24    <.0001
TREAT                        1       5.7449314       5.7449314       3.24    0.0841
TIME*TREAT                   1       7.8813055       7.8813055       4.44    0.0453
DAYS                         1       1.5065660       1.5065660       0.85    0.3658
SEX                          1       0.2458716       0.2458716       0.14    0.7129
WEIGHT                       1       0.0371290       0.0371290       0.02    0.8862
LOSS                         1       0.6289360       0.6289360       0.35    0.5571
TUMOR                        1       2.0861565       2.0861565       1.17    0.2887

Source                      DF      Type III SS     Mean Square    F Value    Pr > F
TIME                         1     71.73825380     71.73825380      40.40    <.0001
TREAT                        1      0.00185100      0.00185100       0.00    0.9745
TIME*TREAT                   1     10.14595894     10.14595894       5.71    0.0247
DAYS                         1      0.45433115      0.45433115       0.26    0.6174
SEX                          1      0.86083198      0.86083198       0.48    0.4927
WEIGHT                       1      0.00406787      0.00406787       0.00    0.9622
LOSS                         1      1.98775141      1.98775141       1.12    0.3001
TUMOR                        1      2.08615646      2.08615646       1.17    0.2887

                                    Standard
Parameter             Estimate         Error     t Value    Pr > |t|
Intercept         -3.920378452 B     6.88309844       -0.57      0.5741
TIME               0.047642506 B     0.01335143        3.57      0.0015
TREAT      BD     -0.019592226 B     0.60680108       -0.03      0.9745
TREAT      NS      0.000000000 B              .           .           .
TIME*TREAT BD      0.039949673 B     0.01671215        2.39      0.0247
TIME*TREAT NS      0.000000000 B              .           .           .
DAYS               0.298264463       0.58963182        0.51      0.6174
SEX        F       0.546097231 B     0.78428957        0.70      0.4927
SEX        M       0.000000000 B              .           .           .
WEIGHT             0.000523527       0.01093757        0.05      0.9622
LOSS              -0.071348343       0.06743238       -1.06      0.3001
TUMOR              0.003112604       0.00287155        1.08      0.2887
NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to
solve the normal equations.  Terms whose estimates are followed by the letter 'B' are not
uniquely estimable.


50         options ps=60 ls=111;
51         proc plot data=next1;  TITLE2 'Various plot with group variable';
52            plot resid * yhat = treat / vref=0;
53         options ps=44 ls=99;
NOTE: There were 34 observations read from the data set WORK.NEXT1.
NOTE: The PROCEDURE PLOT printed page 6.
NOTE: PROCEDURE PLOT used (Total process time):
```

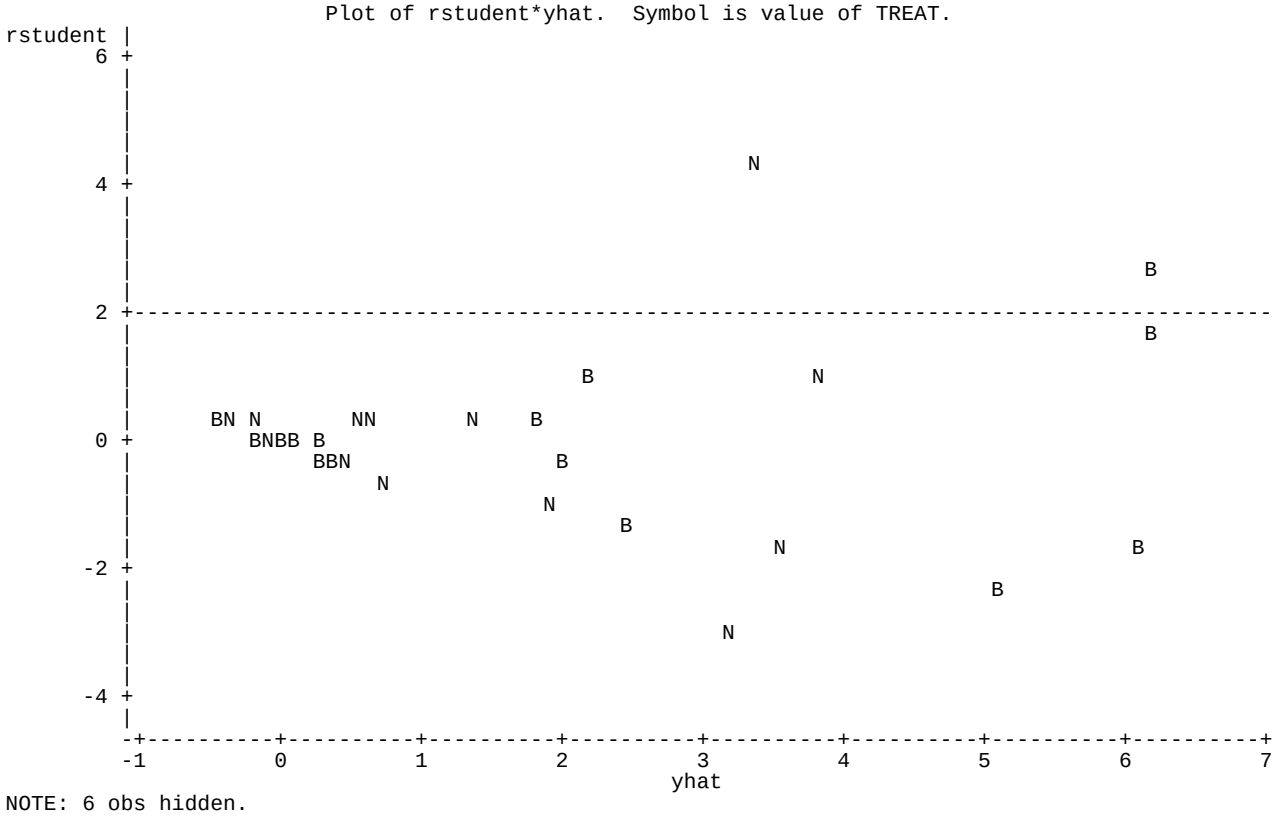
```
      real time           0.08 seconds
      cpu time            0.01 seconds


Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable

                        Plot of resid*yhat.  Symbol is value of TREAT.

resid |
      |
    4 +
      |
      |                                                N
      |
    3 +
      |
      |                                                                                B
      |
    2 +
      |
      |                                                                                B
      |
    1 +                                    B                N
      |
      |       BN                   N    B
      |          B       NN
    0 +-----------NBB-------------------------------------------------------------------------------------
      |               BBN                B
      |                 N
      |                    N
   -1 +                                   N
      |
      |                                        B           N                           B
      |
   -2 +
      |                                                                    B
      |
      |                                              N
   -3 +
      |
      --+----------+----------+----------+----------+----------+----------+----------+----------+----------+--
       -1          0          1          2          3          4          5          6          7
                                                   yhat
NOTE: 7 obs hidden.




54         proc plot data=next1;  TITLE2 'Various plot with group variable';
55            plot student * yhat = treat / vref=2;
56            plot rstudent * yhat = treat / vref=2;
57            plot leverage * yhat = treat / vref=0.55; *** 2p/n = 2*9/34 = 0.53 ***;
58            plot cookd * yhat = treat / vref=1;
59            plot dffits * yhat = treat / vref=1;
60         RUN;
60       !      OPTIONS PS=256 ls=132;
NOTE: There were 34 observations read from the data set WORK.NEXT1.
NOTE: The PROCEDURE PLOT printed pages 7-11.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.09 seconds
      cpu time            0.01 seconds
```

Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable



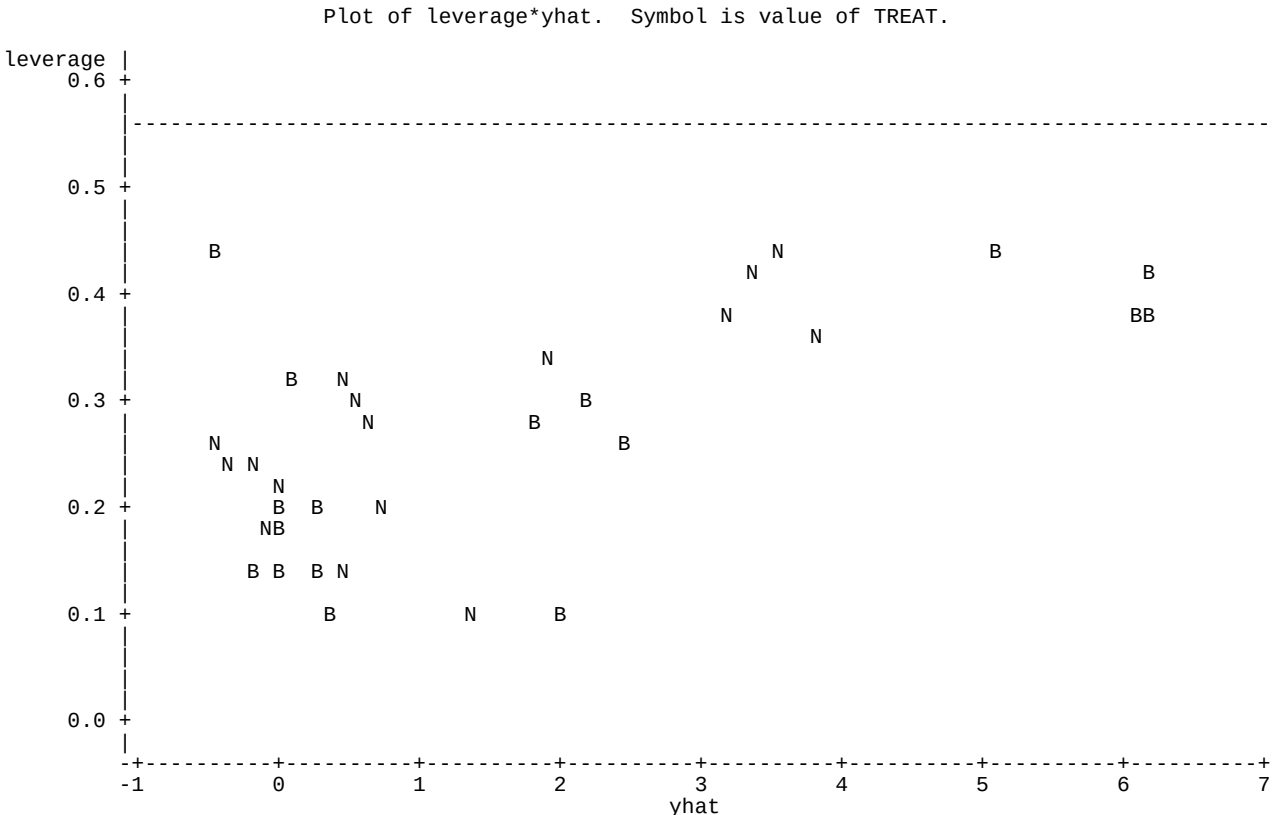
NOTE: 7 obs hidden

Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable



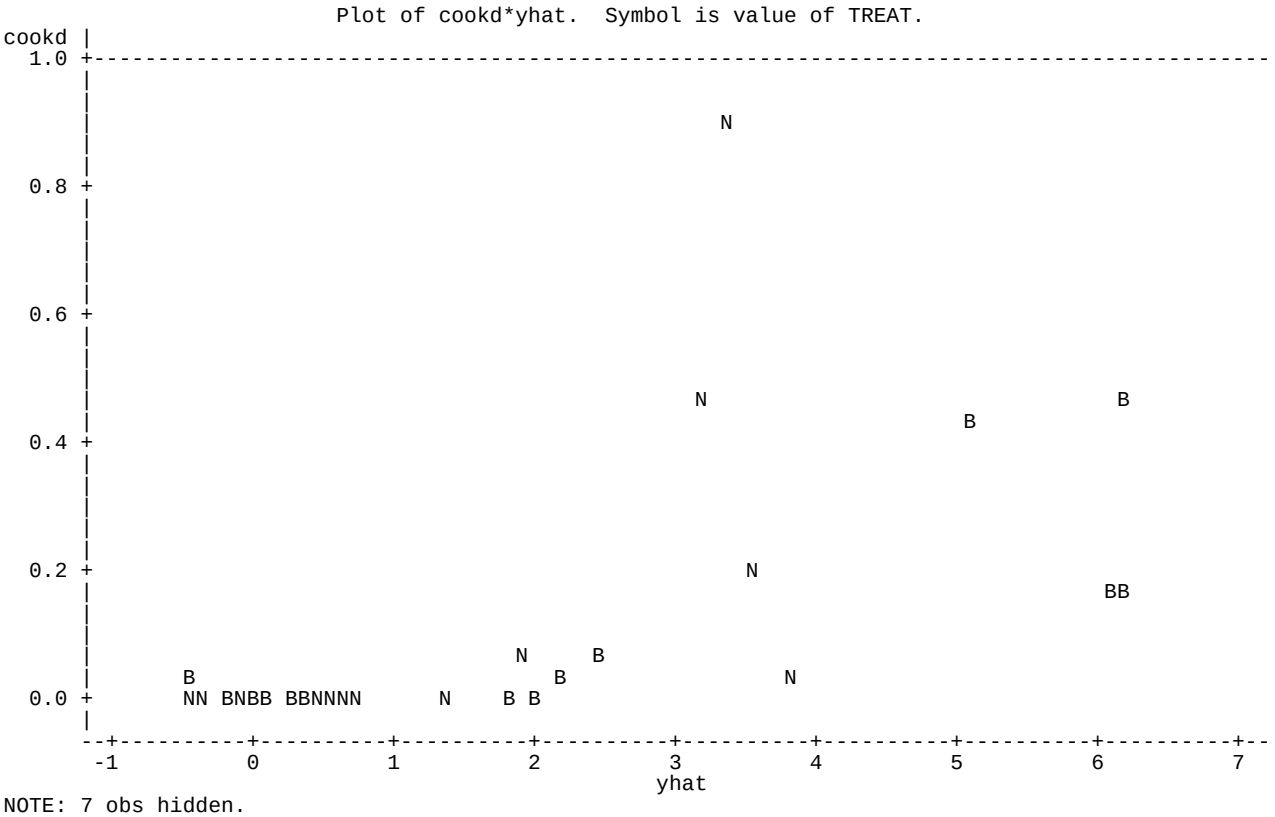
NOTE: 6 obs hidden.

Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable



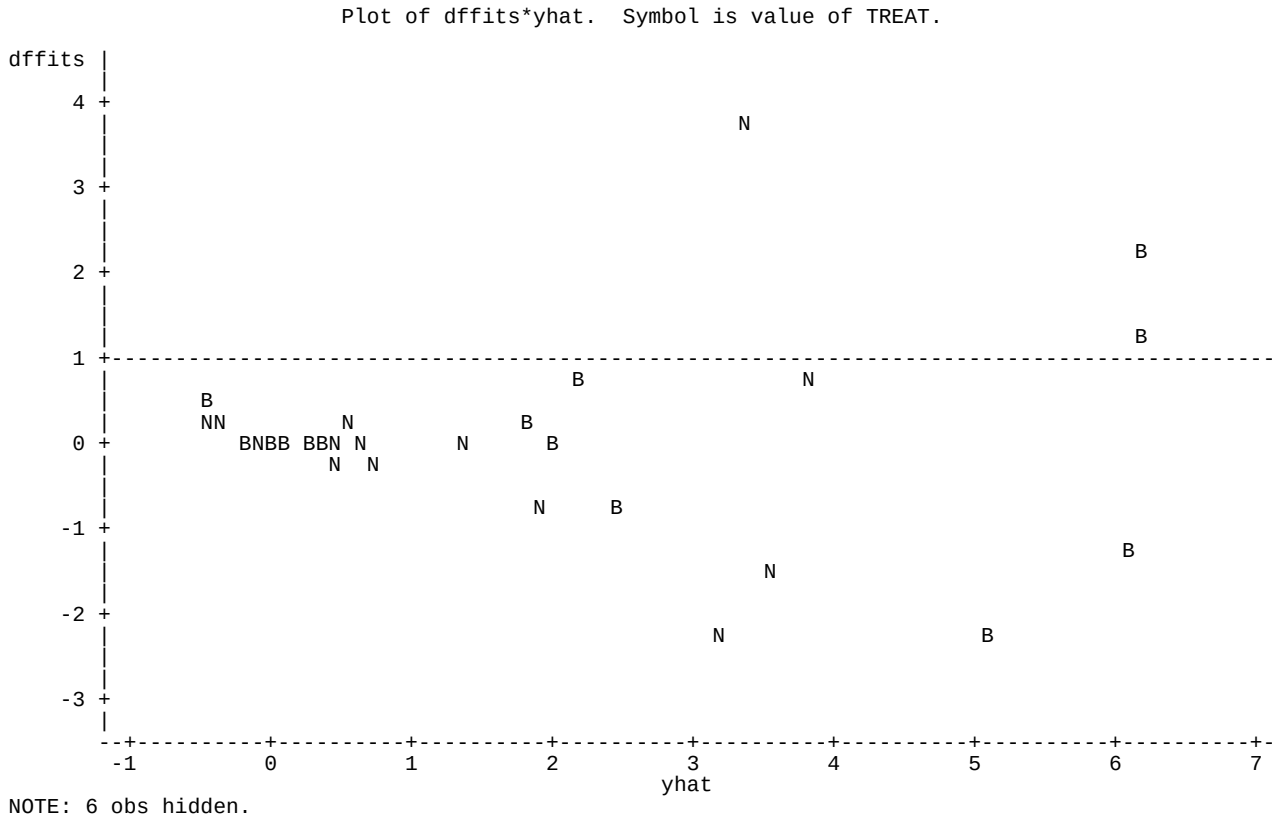
NOTE: 1 obs hidden.

Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable



NOTE: 7 obs hidden.

```
Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable
```




```
NOTE: 6 obs hidden.
```

```
61         PROC UNIVARIATE DATA=NEXT1 NORMAL PLOT; VAR resid; RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 12.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.11 seconds
      cpu time            0.03 seconds


Chapter 11 : Alcohol Barrier in men and women
Various plot with group variable

The UNIVARIATE Procedure
Variable:  resid

Moments
N                          34    Sum Weights                34
Mean                        0    Sum Observations            0
Std Deviation      1.15978822    Variance           1.34510872
Skewness           0.28318382    Kurtosis           2.18250796
Uncorrected SS     44.3885877    Corrected SS       44.3885877
Coeff Variation             .    Std Error Mean     0.19890204

Basic Statistical Measures
    Location                    Variability
Mean     0.000000     Std Deviation            1.15979
Median   0.076283     Variance                 1.34511
Mode      .           Range                    6.14799
                      Interquartile Range      0.76299

Tests for Location: Mu0=0
Test           -Statistic-    -----p Value------
Student's t    t         0    Pr > |t|    1.0000
Sign           M         4    Pr >= |M|   0.2295
Signed Rank    S      16.5    Pr >= |S|   0.7826
```

```
Tests for Normality
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.926874    Pr < W      0.0254
Kolmogorov-Smirnov    D     0.178598    Pr > D     <0.0100
Cramer-von Mises      W-Sq   0.22324    Pr > W-Sq  <0.0050
Anderson-Darling      A-Sq  1.153363    Pr > A-Sq  <0.0050

Quantiles (Definition 5)

Quantile        Estimate
100% Max       3.3783193
99%            3.3783193
95%            2.3914476
90%            1.0607492
75% Q3         0.3920070
50% Median     0.0762833
25% Q1        -0.3709844
10%           -1.5512254
5%            -2.2601855
1%            -2.7696660
0% Min        -2.7696660

Extreme Observations
------Lowest-----        ------Highest-----

   Value      Obs             Value      Obs
-2.76967       34          0.932867       32
-2.26019       29          1.060749       20
-1.57859       30          1.525437       28
-1.55123       33          2.391448       27
-1.38809       22          3.378319       31

   Stem Leaf  Boxplot                      Normal Probability Plot
      3 4                      1     *        3.25+                                                 *
      2                        1     |           |                                                 ++++
      2 4                      1     0           |                                              *++++
      1 5                      1     |           |                                          +*+++
      1 1                      1     |           |                                      +++*
      0 5559                   4     |           |                                 ++++ **
      0 0000112222444         13  +--+--+     0.25+                         *************
     -0 443322                 6  +-----+        |                    ******++
     -0 7                      1     |           |                 **++++
     -1 40                     2     |           |              +*++
     -1 66                     2     0           |         +++*+*
     -2 3                      1     0           |     ++++*
     -2 8                      1     *       -2.75+++++*
        ----+----+----+----+                       +----+----+----+----+----+----+----+----+----+----+----+
                                                      -2        -1         0        +1        +2

63         data Barrier; set Barrier;
64            treatment = 0; if treat eq 'BD' then treatment = 1;
65            sex2 = 0; if sex eq 'F' then sex2 = 1;
66            timeXtreat = treatment*time;
67         run;
NOTE: There were 34 observations read from the data set WORK.BARRIER.
NOTE: The data set WORK.BARRIER has 34 observations and 13 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.02 seconds
68
69         options ps=44 ls=99;
70         PROC REG DATA=Barrier; Title2 'Fit of ratio on indicator variables with REG';
71            MODEL ratio = time treatment timeXtreat days sex2 weight loss tumor / partial;
72         RUN;
73         quit;
NOTE: The PROCEDURE REG printed pages 13-22.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.15 seconds
      cpu time            0.07 seconds
```

The REG Procedure
Model: MODEL1
Dependent Variable: ratio

Number of Observations Read 34
Number of Observations Used 34

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	137.51983	17.18998	9.68	<.0001
Error	25	44.38859	1.77554		
Corrected Total	33	181.90841			

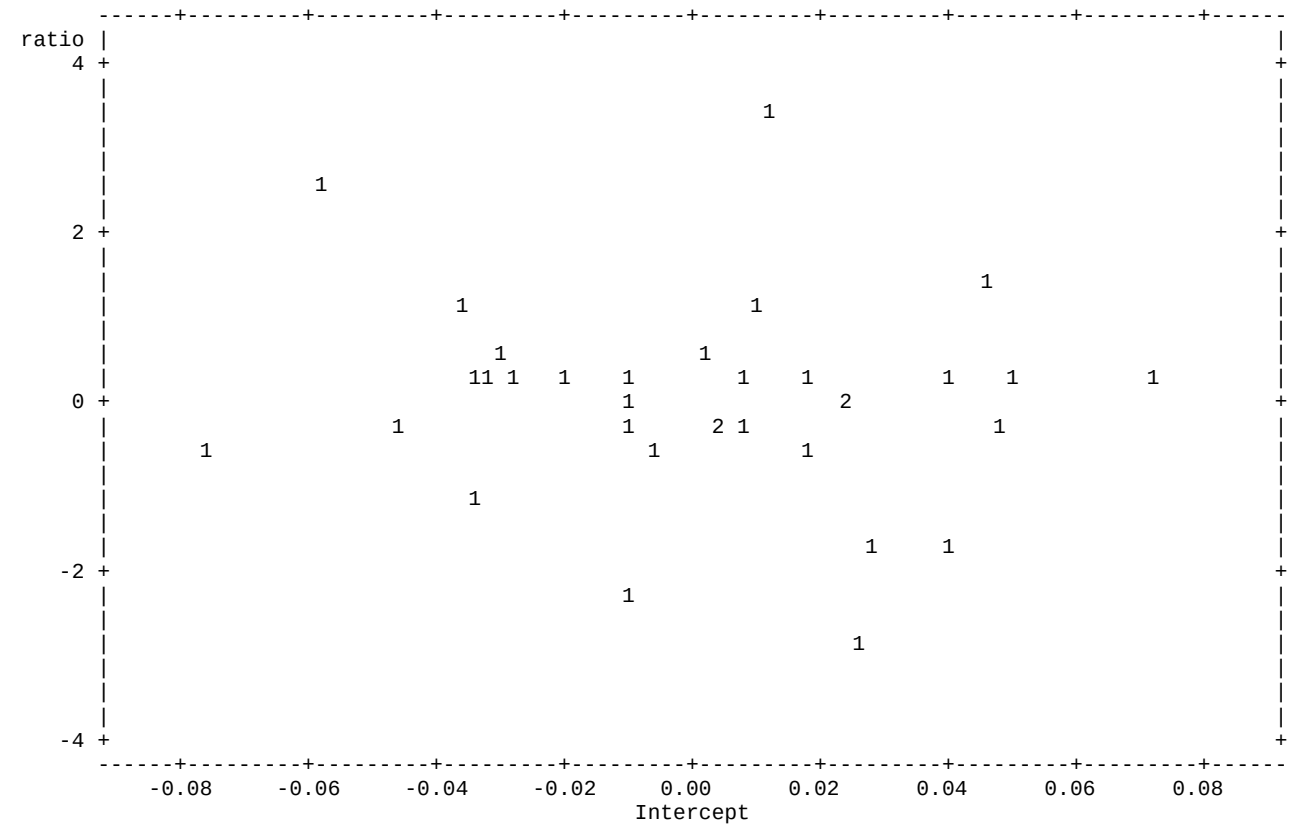
Root MSE	1.33250	R-Square	0.7560
Dependent Mean	1.50364	Adj R-Sq	0.6779
Coeff Var	88.61808		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.92038	6.88310	-0.57	0.5741
TIME	1	0.04764	0.01335	3.57	0.0015
treatment	1	-0.01959	0.60680	-0.03	0.9745
timeXtreat	1	0.03995	0.01671	2.39	0.0247
DAYS	1	0.29826	0.58963	0.51	0.6174
sex2	1	0.54610	0.78429	0.70	0.4927
WEIGHT	1	0.00052353	0.01094	0.05	0.9622
LOSS	1	-0.07135	0.06743	-1.06	0.3001
TUMOR	1	0.00311	0.00287	1.08	0.2887

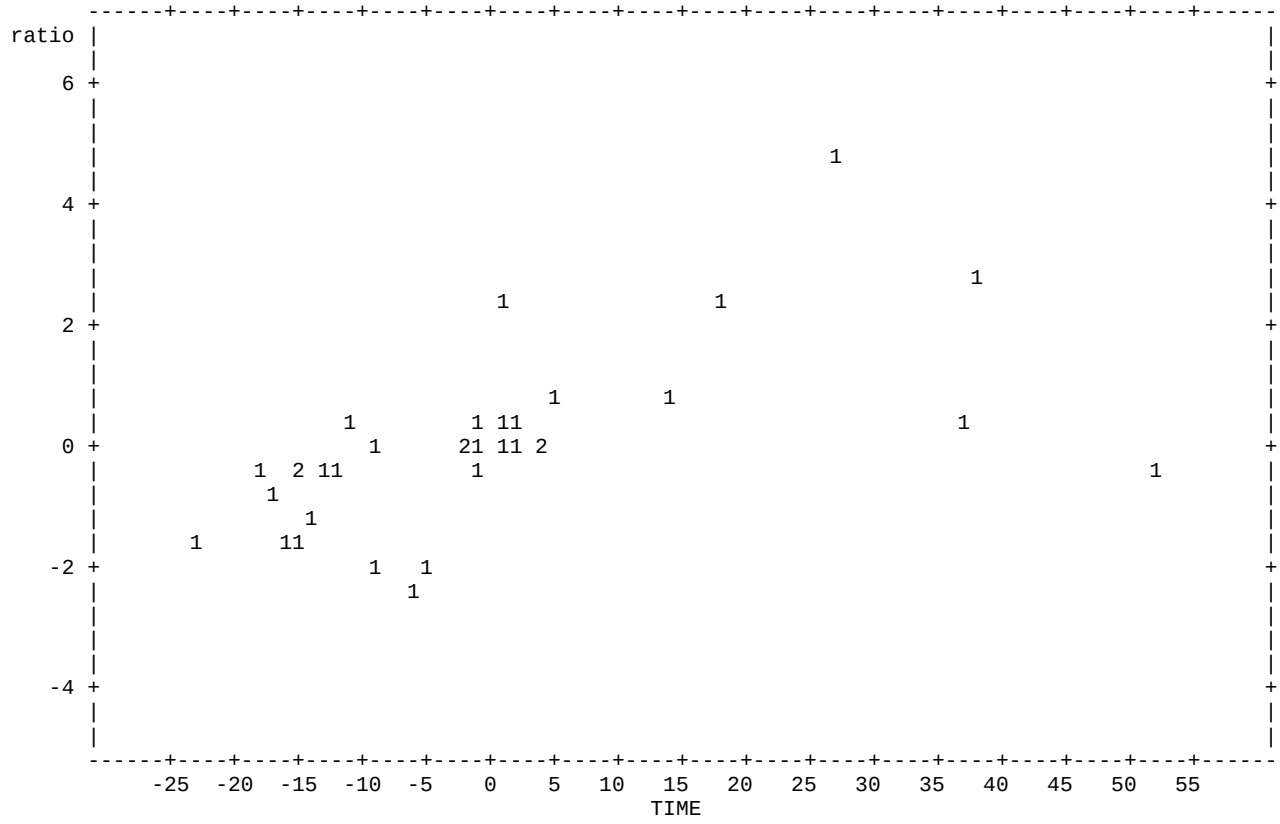
Chapter 11 : Alcohol Barrier in men and women
Fit of ratio on indicator variables with REG

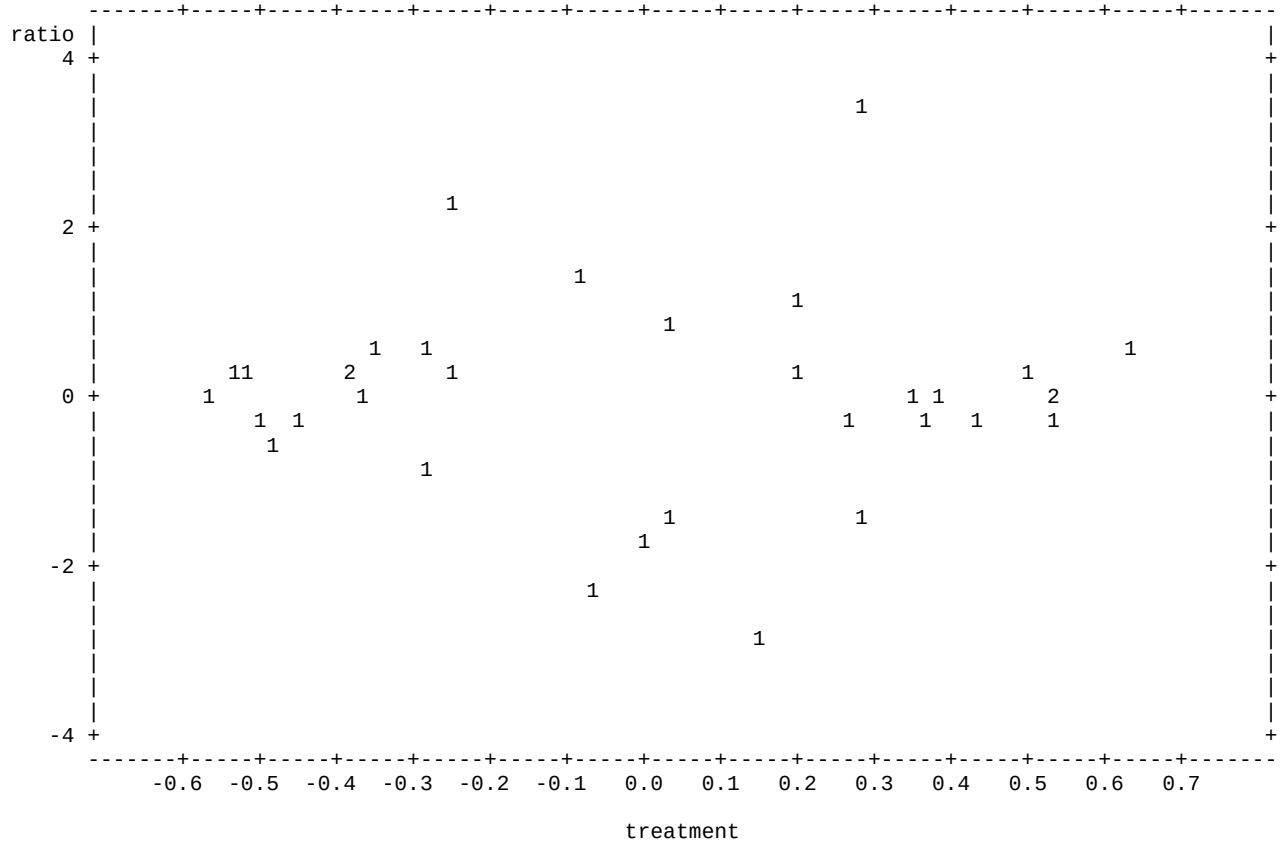
The REG Procedure
Model: MODEL1
Partial Regression Residual Plot



Chapter 11 : Alcohol Barrier in men and women
Fit of ratio on indicator variables with REG

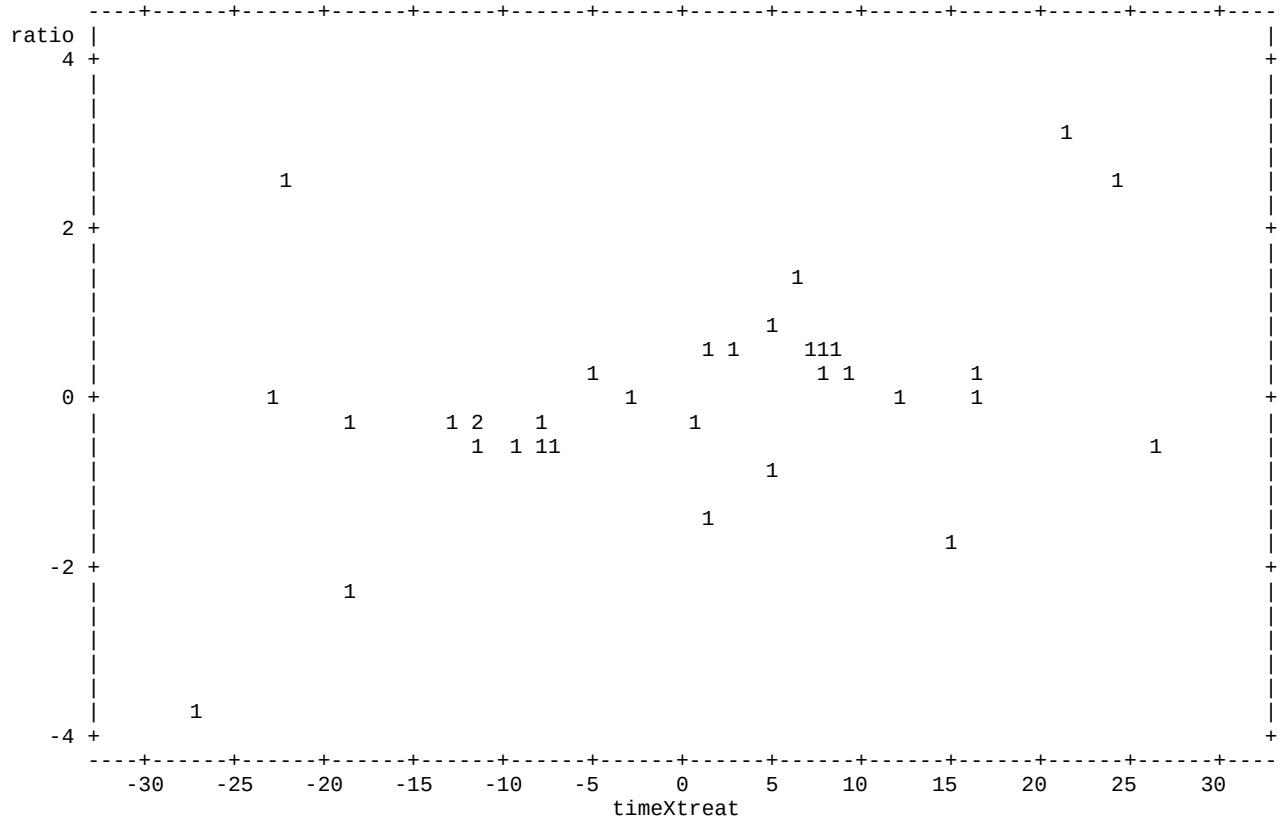
The REG Procedure
Model: MODEL1
Partial Regression Residual Plot

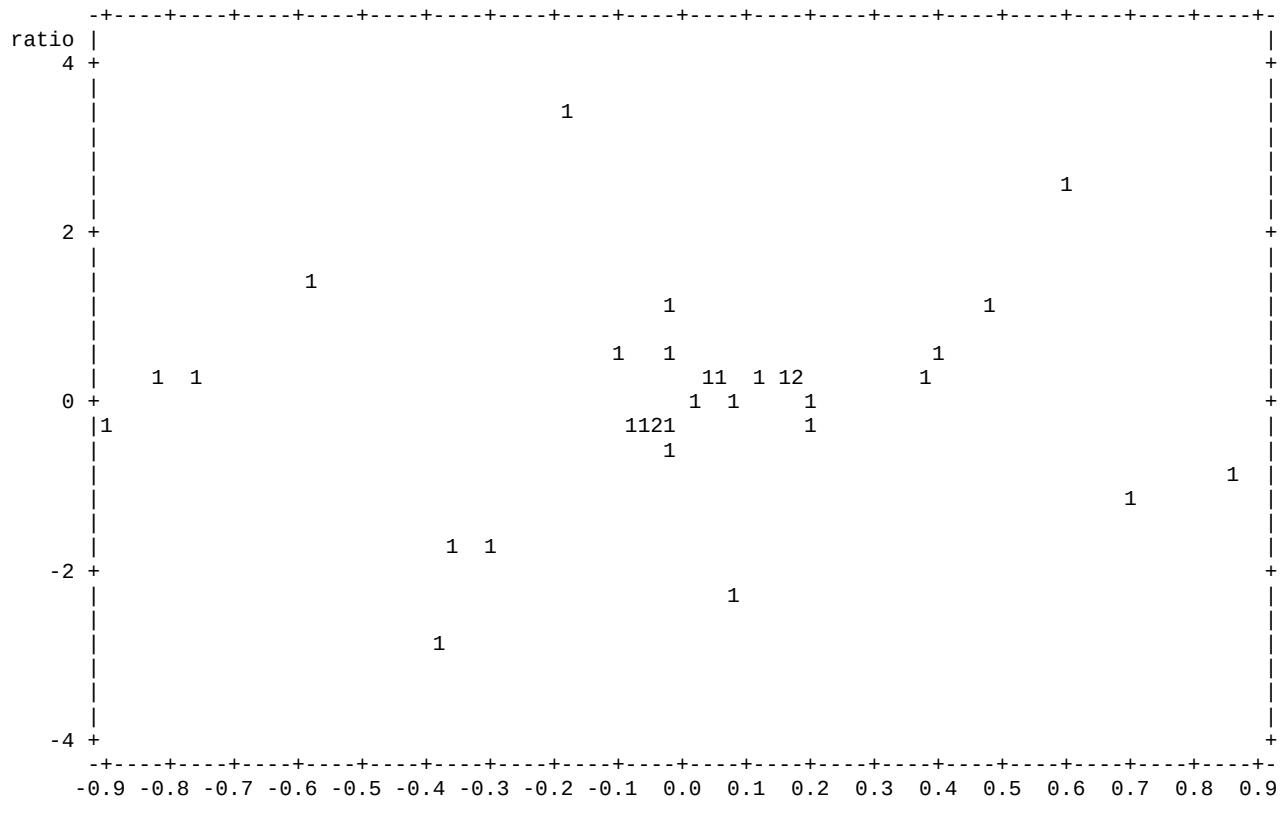




Chapter 11 : Alcohol Barrier in men and women
Fit of ratio on indicator variables with REG

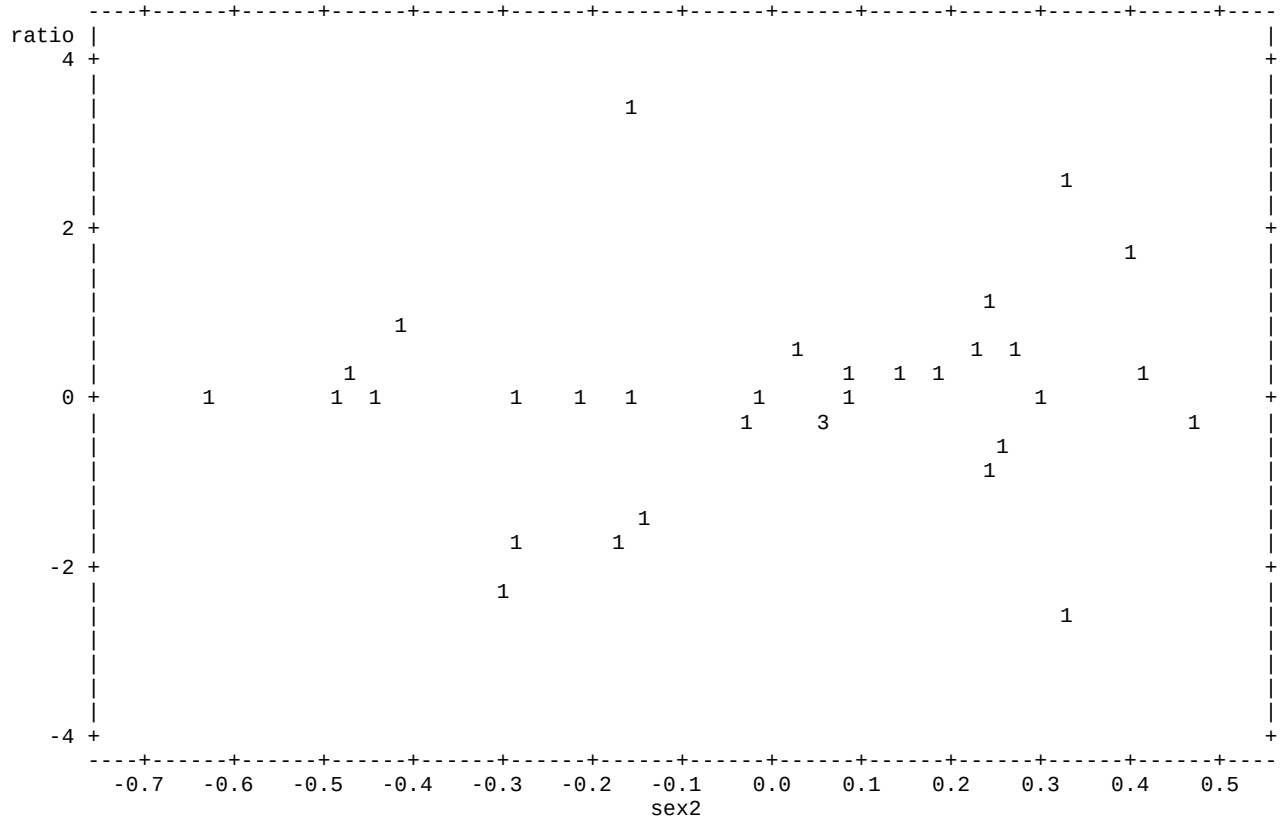
The REG Procedure
Model: MODEL1
Partial Regression Residual Plot

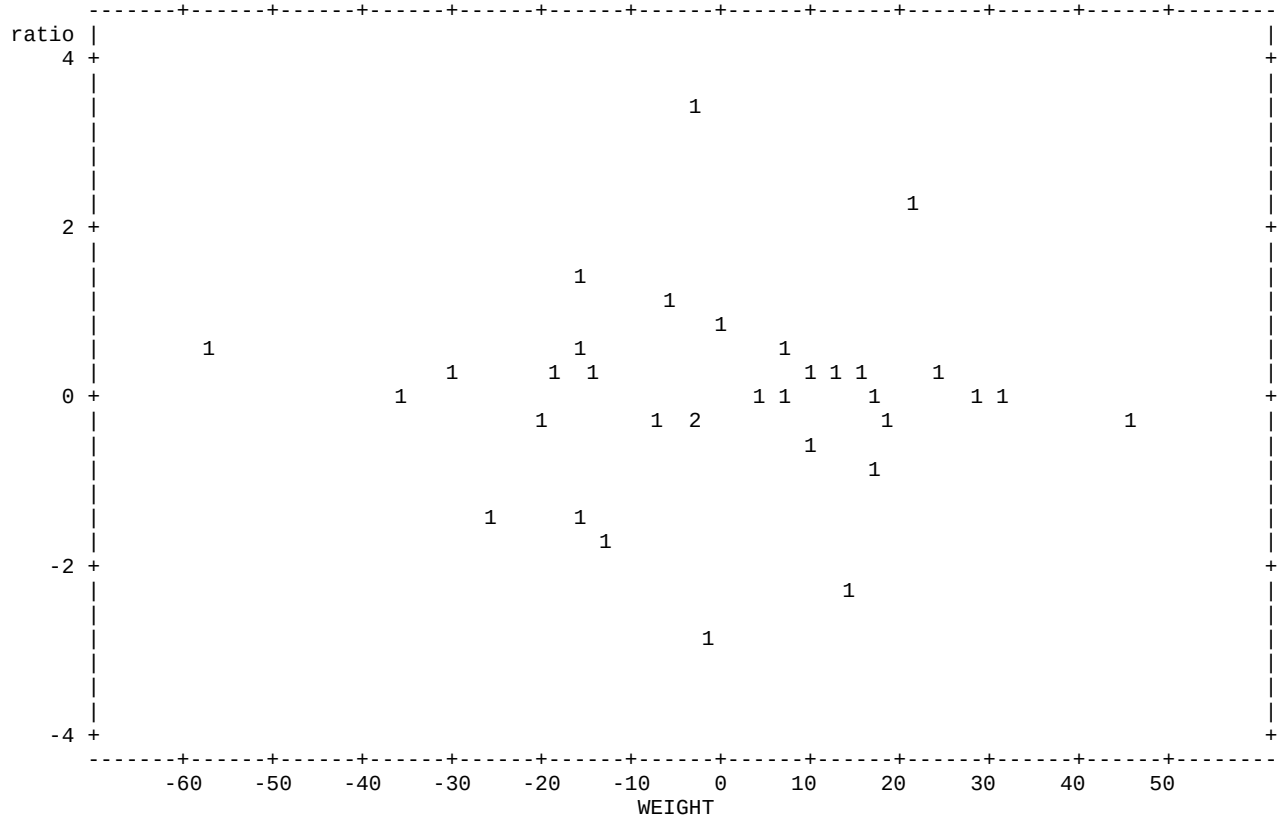




Chapter 11 : Alcohol Barrier in men and women
Fit of ratio on indicator variables with REG

The REG Procedure
Model: MODEL1
Partial Regression Residual Plot





Chapter 11 : Alcohol Barrier in men and women
Fit of ratio on indicator variables with REG

The REG Procedure
Model: MODEL1
Partial Regression Residual Plot