```
1          ************************************************************************;
2          *** EXST7034 Homework Example                                        ***;
3          *** Problem from Neter, Kutner, Nachtsheim, Wasserman 1996,14.9   ***;
4          ************************************************************************;
5
6          dm'log;clear;output;clear';
7          options nodate nocenter nonumber ps=512 ls=132 nolabel;
8          ODS HTML style=minimal rs=none
       body='C:\Geaghan\Current\EXST7034\Fall2005\SAS\Toxicity01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST7034\Fall2005\SAS\Toxicity01.html
9
10         DATA Toxicity indicator; INFILE CARDS MISSOVER;
11              TITLE1 'Logistic regression : Toxicity experiment';
12              LABEL X = 'Dose level (log scale)';
13              LABEL R = 'No of insects which died';
14              LABEL N = 'No of insects exposed';
15              LABEL P = 'Mortality proportion';
16            INPUT X R N;
17              P = R / N;
18              LOGIT = LOG(P/(1-P));
19              WT = N*P*(1-P);
20            output toxicity;
21            indicator = 1; output indicator;
22            R = N - R; indicator = 0; output indicator;
23                 *** NOTE: expected variance of P is 1/(n*p*(1-p)) ***;
24         CARDS;
NOTE: The data set WORK.TOXICITY has 6 observations and 7 variables.
NOTE: The data set WORK.INDICATOR has 12 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds
24       !        RUN;
31         ;
32         PROC PRINT DATA=toxicity; VAR X R N P;
33             TITLE2 'Sorted Raw Data Listing : original data'; RUN;
NOTE: There were 6 observations read from the data set WORK.TOXICITY.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.11 seconds
      cpu time            0.00 seconds
34         PROC PRINT DATA=indicator; VAR X R N P indicator;
35             TITLE2 'Sorted Raw Data Listing : with indicator variable'; RUN;
NOTE: There were 12 observations read from the data set WORK.INDICATOR.
NOTE: The PROCEDURE PRINT printed page 2.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.05 seconds
      cpu time            0.01 seconds
```

Logistic regression : Toxicity experiment
Sorted Raw Data Listing : original data

Obs	X	R	N	P
1	1	28	250	0.112
2	2	53	250	0.212
3	3	93	250	0.372
4	4	126	250	0.504
5	5	172	250	0.688
6	6	197	250	0.788

Logistic regression : Toxicity experiment
Sorted Raw Data Listing : with indicator

Obs	X	R	N	P	indicator
1	1	28	250	0.112	1
2	1	222	250	0.112	0
3	2	53	250	0.212	1
4	2	197	250	0.212	0
5	3	93	250	0.372	1
6	3	157	250	0.372	0
7	4	126	250	0.504	1
8	4	124	250	0.504	0
9	5	172	250	0.688	1
10	5	78	250	0.688	0
11	6	197	250	0.788	1
12	6	53	250	0.788	0

```
37         PROC REG  DATA=toxicity;
38            TITLE2 'Using PROC REG';
39            TITLE3 'As a Simple linear regression with proportion';
40              MODEL P = X;
41              output out=next1 p=pred1 r=resid1;
42         RUN;
42       !       QUIT;
NOTE: The data set WORK.NEXT1 has 6 observations and 9 variables.
NOTE: The PROCEDURE REG printed page 3.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.11 seconds
      cpu time            0.05 seconds


Logistic regression : Toxicity experiment
Using PROC REG
As a Simple linear regression with proportion

The REG Procedure
Model: MODEL1
Dependent Variable: P

Number of Observations Read           6
Number of Observations Used           6

Analysis of Variance
                                  Sum of           Mean
Source                     DF        Squares         Square    F Value    Pr > F
Model                       1        0.34862        0.34862     677.88    <.0001
Error                       4        0.00206     0.00051429
Corrected Total             5        0.35068

Root MSE              0.02268    R-Square     0.9941
Dependent Mean        0.44600    Adj R-Sq     0.9927
Coeff Var             5.08472

Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     1       -0.04800        0.02111      -2.27      0.0854
X             1        0.14114        0.00542      26.04      <.0001


44         PROC REG  DATA=indicator; freq R;
45            TITLE3 'As a Simple linear regression with indicator variable';
46              MODEL indicator = X;
47              output out=next2 p=pred2 r=resid2;
48         RUN;
48       !       QUIT;
NOTE: The data set WORK.NEXT2 has 12 observations and 9 variables.
NOTE: The PROCEDURE REG printed page 4.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.13 seconds
      cpu time            0.05 seconds


Logistic regression : Toxicity experiment
Using PROC REG
As a Simple linear regression with indicator variable

The REG Procedure
Model: MODEL1
Dependent Variable: indicator

Number of Observations Read          12
Number of Observations Used          12
Sum of Frequencies Read            1500
Sum of Frequencies Used            1500
```

```
Frequency: R

Analysis of Variance
                                   Sum of           Mean
Source                     DF      Squares         Square    F Value    Pr > F
Model                       1     87.15571       87.15571     460.57    <.0001
Error                    1498    283.47029        0.18923
Corrected Total          1499    370.62600

Root MSE              0.43501    R-Square     0.2352
Dependent Mean        0.44600    Adj R-Sq     0.2346
Coeff Var            97.53557

Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     1       -0.04800        0.02561      -1.87      0.0611
X             1        0.14114        0.00658      21.46      <.0001


50         PROC REG  DATA=toxicity; WEIGHT WT;
51            TITLE3 'As a weighted regression on logits';
52              MODEL LOGIT = X;
53              OUTPUT OUT=NEXT3 PRED=PRED3 R=RESID3;
54         RUN;
54       !       QUIT;
NOTE: The data set WORK.NEXT3 has 6 observations and 9 variables.
NOTE: The PROCEDURE REG printed page 5.
NOTE: PROCEDURE REG used (Total process time):
      real time           0.14 seconds
      cpu time            0.08 seconds

Logistic regression : Toxicity experiment
Using PROC REG
As a weighted regression on logits

The REG Procedure
Model: MODEL1
Dependent Variable: LOGIT

Number of Observations Read           6
Number of Observations Used           6

Weight: WT

Analysis of Variance
                                   Sum of           Mean
Source                     DF      Squares         Square    F Value    Pr > F
Model                       1    293.43459      293.43459     809.32    <.0001
Error                       4      1.45028        0.36257
Corrected Total             5    294.88487

Root MSE              0.60214    R-Square     0.9951
Dependent Mean       -0.13651    Adj R-Sq     0.9939
Coeff Var          -441.09139

Parameter Estimates
                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|
Intercept     1       -2.64011        0.09501     -27.79      <.0001
X             1        0.67308        0.02366      28.45      <.0001
```

NOTE: exp(0.673076) = 1.960257809

NOTE: LD_{50} = Prob @ 50% : Log odds = log(0.5 / (1-0.5)) = log(1) = 0

Then: 0 = -2.64011 + 0.67308*Dose50 so Dose50 = 2.64011 / 0.67308 = 3.923

```
68         PROC NLIN  DATA=toxicity; _WEIGHT_=WT;
69            TITLE2 'Using Nonlinear regression with PROC NLIN';
70              PARMS B0=-2.64 B1=.67;
71              MODEL P = EXP(B0+B1*X)/(1+EXP(B0+B1*X));
72              OUTPUT OUT=NEXT6 PRED=pred6 r=resid;
73         RUN;
NOTE: DER.B0 not initialized or missing. It will be computed automatically.
NOTE: DER.B1 not initialized or missing. It will be computed automatically.
NOTE: PROC NLIN grid search time was  0: 0: 0.
NOTE: Convergence criterion met.
NOTE: The data set WORK.NEXT6 has 6 observations and 9 variables.
NOTE: The PROCEDURE NLIN printed page 8.
NOTE: PROCEDURE NLIN used (Total process time):
      real time           0.13 seconds
      cpu time            0.05 seconds
73       !       QUIT;


Logistic regression : Toxicity experiment
Using Nonlinear regression with PROC NLIN

The NLIN Procedure
Dependent Variable P
Method: Gauss-Newton

             Iterative Phase
                                 Weighted
 Iter          B0          B1          SS
    0     -2.6400      0.6700      0.0618
    1     -2.5882      0.6635      0.0525
    2     -2.5899      0.6639      0.0525
    3     -2.5899      0.6639      0.0525
NOTE: Convergence criterion met.

         Estimation Summary
Method                  Gauss-Newton
Iterations                         3
R                           1.798E-6
PPC(B0)                      1.51E-7
RPC(B0)                     9.237E-6
Object                      1.357E-8
Objective                   0.052534
Observations Read                  6
Observations Used                  6
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                      2     77.4282     38.7141    2947.72    <.0001
Error                      4      0.0525      0.0131
Uncorrected Total          6     77.4807

                             Approx
Parameter      Estimate    Std Error    Approximate 95% Confidence Limits
B0              -2.5899       0.1111     -2.8985     -2.2813
B1               0.6639       0.0275      0.5875      0.7404

  Approximate Correlation Matrix
               B0              B1
B0       1.0000000      -0.9552149
B1      -0.9552149       1.0000000
```

```
56         PROC LOGISTIC DATA=toxicity;
57            TITLE2 'Using proc logistic on R/N';
58              MODEL R/N = X;
59              OUTPUT OUT=NEXT4 PRED=PRED4 Resdev=RESID4 u=ucl4 l=lcl4;
60         RUN;
NOTE: Convergence criterion (GCONV=1E-8) satisfied.
NOTE: There were 6 observations read from the data set WORK.TOXICITY.
NOTE: The data set WORK.NEXT4 has 6 observations and 11 variables.
NOTE: The PROCEDURE LOGISTIC printed page 6.
NOTE: PROCEDURE LOGISTIC used (Total process time):
      real time           0.10 seconds
      cpu time            0.04 seconds
60       !       QUIT;


Logistic regression : Toxicity experiment
Using proc logistic on R/N

The LOGISTIC Procedure

Model Information
Data Set                          WORK.TOXICITY
Response Variable (Events)        R
Response Variable (Trials)        N
Model                             binary logit
Optimization Technique            Fisher's scoring

Number of Observations Read           6
Number of Observations Used           6
Sum of Frequencies Read            1500
Sum of Frequencies Used            1500

          Response Profile
 Ordered     Binary           Total
   Value     Outcome      Frequency
       1     Event              669
       2     Nonevent           831

Model Convergence Status
         Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics         Intercept
             Intercept            and
Criterion          Only     Covariates
AIC            2063.911       1684.291
SC             2069.225       1694.917
-2 Log L       2061.911       1680.291

Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio       381.6204        1         <.0001
Score                  352.7372        1         <.0001
Wald                   296.9818        1         <.0001

Analysis of Maximum Likelihood Estimates
                                 Standard          Wald
Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq
Intercept     1     -2.6435      0.1561      286.7841        <.0001
X             1      0.6740      0.0391      296.9818        <.0001

Odds Ratio Estimates
             Point          95% Wald
Effect    Estimate      Confidence Limits
X            1.962       1.817       2.118
```

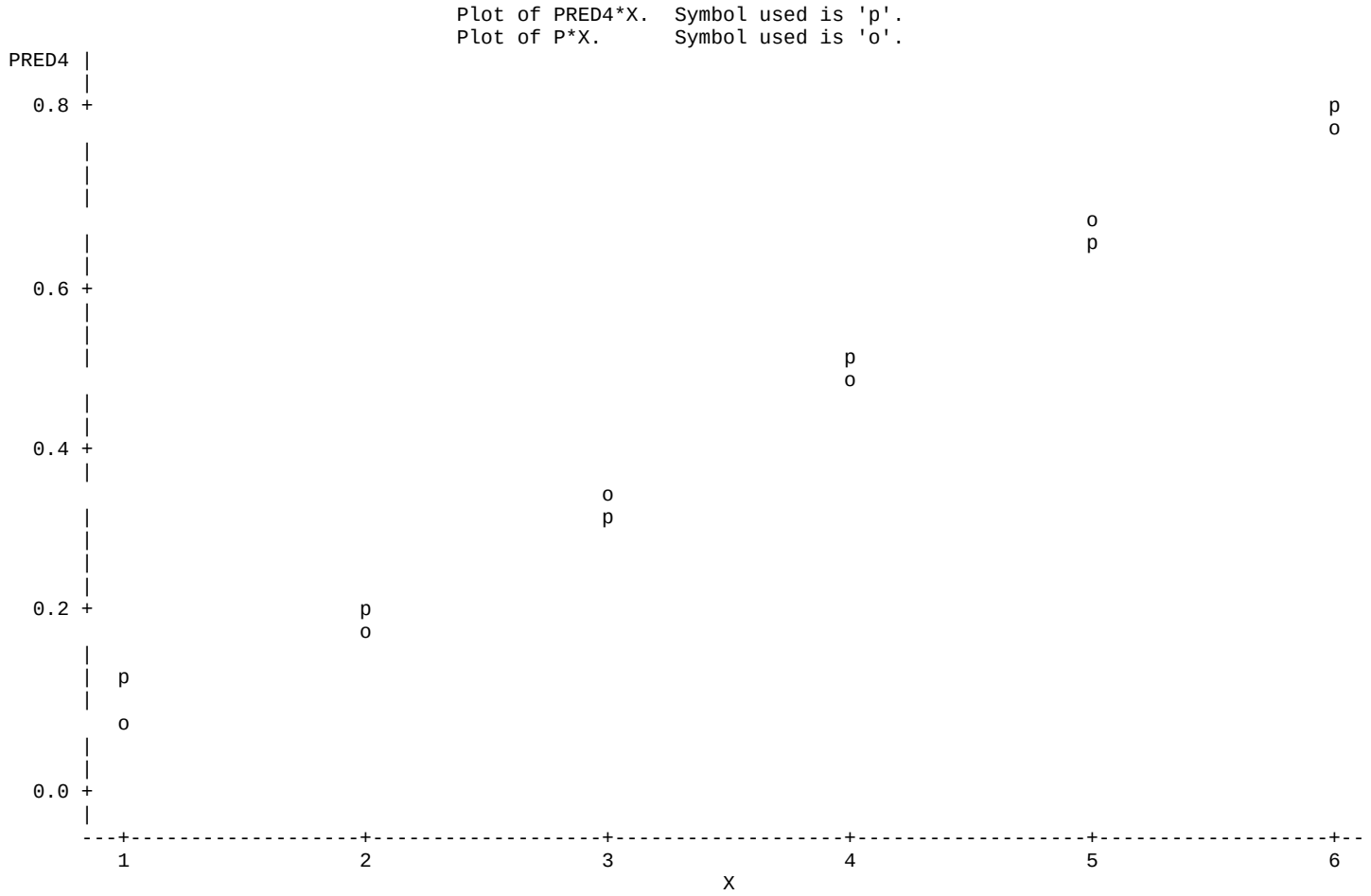
```
Association of Predicted Probabilities and Observed Responses
Percent Concordant      71.4    Somers' D    0.555
Percent Discordant      15.9    Gamma        0.636
Percent Tied            12.7    Tau-a        0.275
Pairs                 555939    c            0.778


62         PROC LOGISTIC DATA=indicator DESCENDING; FREQ R;
63            TITLE2 'Using proc logistic on indicator variable';
64              MODEL indicator = X;
65              OUTPUT OUT=NEXT5 PRED=PRED5 Resdev=RESID5 u=ucl5 l=lcl5;
66         RUN;
NOTE: PROC LOGISTIC is modeling the probability that indicator=1.
NOTE: Convergence criterion (GCONV=1E-8) satisfied.
NOTE: There were 12 observations read from the data set WORK.INDICATOR.
NOTE: The data set WORK.NEXT5 has 12 observations and 12 variables.
NOTE: The PROCEDURE LOGISTIC printed page 7.
NOTE: PROCEDURE LOGISTIC used (Total process time):
      real time           0.15 seconds
      cpu time            0.05 seconds
66       !       QUIT;
67


Logistic regression : Toxicity experiment
Using proc logistic on indicator variable

The LOGISTIC Procedure

Model Information
Data Set                      WORK.INDICATOR
Response Variable             indicator
Number of Response Levels     2
Frequency Variable            R
Model                         binary logit
Optimization Technique        Fisher's scoring

Number of Observations Read          12
Number of Observations Used          12
Sum of Frequencies Read            1500
Sum of Frequencies Used            1500

          Response Profile
 Ordered                        Total
   Value     indicator     Frequency
       1             1           669
       2             0           831

Probability modeled is indicator=1.

                    Model Convergence Status
         Convergence criterion (GCONV=1E-8) satisfied.

         ModelFit Statistics
                               Intercept
             Intercept               and
Criterion          Only     Covariates
AIC             2063.911       1684.291
SC              2069.225       1694.917
-2 Log L        2061.911       1680.291

Testing Global Null Hypothesis: BETA=0
Test                 Chi-Square       DF     Pr > ChiSq
Likelihood Ratio       381.6204        1         <.0001
Score                  352.7372        1         <.0001
Wald                   296.9818        1         <.0001
```

```
Analysis of Maximum Likelihood Estimates
                              Standard          Wald
Parameter    DF    Estimate       Error    Chi-Square    Pr > ChiSq
Intercept     1     -2.6435      0.1561      286.7841        <.0001
X             1      0.6740      0.0391      296.9818        <.0001

Odds Ratio Estimates
            Point          95% Wald
Effect    Estimate      Confidence Limits
X            1.962       1.817       2.118

Association of Predicted Probabilities and Observed Responses
Percent Concordant      71.4    Somers' D    0.555
Percent Discordant      15.9    Gamma        0.636
Percent Tied            12.7    Tau-a        0.275
Pairs                 555939    c            0.778


75         data next2; set next2; if indicator = 1; run;
NOTE: There were 12 observations read from the data set WORK.NEXT2.
NOTE: The data set WORK.NEXT2 has 6 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.04 seconds
      cpu time            0.04 seconds
76         data next3; set next3; odds3=exp(pred3); prob3=exp(pred3)/(1+exp(pred3)); run;
NOTE: There were 6 observations read from the data set WORK.NEXT3.
NOTE: The data set WORK.NEXT3 has 6 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.04 seconds
77         data next5; set next5; if indicator = 1; run;
NOTE: There were 12 observations read from the data set WORK.NEXT5.
NOTE: The data set WORK.NEXT5 has 6 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      cpu time            0.03 seconds
78
79         data all; merge next1 next2 next3 next4 next5 next6; by x; run;
NOTE: There were 6 observations read from the data set WORK.NEXT1.
NOTE: There were 6 observations read from the data set WORK.NEXT2.
NOTE: There were 6 observations read from the data set WORK.NEXT3.
NOTE: There were 6 observations read from the data set WORK.NEXT4.
NOTE: There were 6 observations read from the data set WORK.NEXT5.
NOTE: There were 6 observations read from the data set WORK.NEXT6.
NOTE: The data set WORK.ALL has 6 observations and 26 variables.
NOTE: DATA statement used (Total process time):
      real time           0.06 seconds
      cpu time            0.07 seconds
80
81    PROC PRINT DATA=all; var X R N P PRED1 PRED2 PRED3 ODDS3 PROB3 PRED4 PRED5 pred6 LCL4 UCL4;
82         RUN;
NOTE: There were 6 observations read from the data set WORK.ALL.
NOTE: The PROCEDURE PRINT printed page 9.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.07 seconds
      cpu time            0.01 seconds
83
84         OPTIONS PS=40;
84       !               PROC PLOT DATA=all; PLOT PRED4*X='p' P*X='o' / OVERLAY;
85              TITLE2 'Plot of the raw data'; RUN;
85       !                                          OPTIONS PS=256;
86
NOTE: There were 6 observations read from the data set WORK.ALL.
NOTE: The PROCEDURE PLOT printed page 10.
NOTE: PROCEDURE PLOT used (Total process time):
      real time           0.06 seconds
      cpu time            0.00 seconds
```

Logistic regression : Toxicity experiment
Using Nonlinear regression with PROC NLIN

Obs	X	R	N	P	pred1	pred2	PRED3	odds3	prob3	pred6	PRED4	PRED5	lcl4	ucl4
1	1	28	250	0.112	0.09314	0.09314	-1.96703	0.13987	0.12271	0.12720	0.12244	0.12244	0.09917	0.15025
2	2	53	250	0.212	0.23429	0.23429	-1.29395	0.27418	0.21518	0.22063	0.21491	0.21491	0.18696	0.24577
3	3	93	250	0.372	0.37543	0.37543	-0.62088	0.53747	0.34958	0.35478	0.34941	0.34941	0.32090	0.37903
4	4	126	250	0.504	0.51657	0.51657	0.05220	1.05358	0.51305	0.51646	0.51307	0.51307	0.48336	0.54270
5	5	172	250	0.688	0.65771	0.65771	0.72527	2.06530	0.67377	0.67476	0.67399	0.67399	0.63934	0.70683
6	6	197	250	0.788	0.79886	0.79886	1.39835	4.04852	0.80192	0.80119	0.80222	0.80222	0.76644	0.83371

Logistic regression : Toxicity experiment
Plot of the raw data



```
87         proc probit data=toxicity;
88            TITLE2 'The Probit Analysis alternative (PROC PROBIT)';
89                 MODEL R / N = X / lackfit inversecl itprint;
90         run;
NOTE: Algorithm converged.
NOTE: Since the chi-square is small (p > 0.1000), fiducial limits will be calculated using a t
      value of  1.96.
NOTE: The PROCEDURE PROBIT printed pages 11-12.
NOTE: PROCEDURE PROBIT used (Total process time):
      real time           0.09 seconds
      cpu time            0.04 seconds


Logistic regression : Toxicity experiment
The Probit Analysis alternative (PROC PROBIT)

Probit Procedure

          Iteration History forParameter Estimates
Iter    Ridge    Loglikelihood       Intercept               X
   0        0       -1039.7208               0               0
   1        0       -843.30756    -1.373632294    0.3537926765
   2        0       -840.02049    -1.589086086    0.4057285694
   3        0       -840.01726    -1.596151662    0.4073569795
   4        0       -840.01726    -1.596159362    0.4073587083
   5        0       -840.01726    -1.596159362    0.4073587083

Model Information
Data Set                  WORK.TOXICITY
Events Variable                       R
Trials Variable                       N
Number of Observations                6
Number of Events                    669
Number of Trials                   1500
Name of Distribution             Normal
Log Likelihood             -840.0172634

Number of Observations Read           6
Number of Observations Used           6
Number of Events                    669
Number of Trials                   1500

Parameter Information
Parameter    Effect
Intercept    Intercept
X            X

Last Evaluation of the Negative of the Gradient

   Intercept             X
1.4329231E-9  1.0690542E-9

Last Evaluation of the Negative of the Hessian

              Intercept             X
Intercept  801.20363321   2914.911468
X           2914.911468  12634.622901
Algorithm converged.

                    Goodness-of-Fit Tests
Statistic                          Value       DF    Pr > ChiSq
Pearson Chi-Square                1.1957        4        0.8788
L.R.    Chi-Square                1.1926        4        0.8793

Response-Covariate Profile
Response Levels               2
Number of Covariate Values    6
```

Since the chi-square is small (p > 0.1000), fiducial limits will be calculated using a t value of 1.96.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
X	1	336.8094	<.0001

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.5962	0.0881	-1.7689	-1.4234	327.92	<.0001
X	1	0.4074	0.0222	0.3639	0.4509	336.81	<.0001

Probit Model in Terms of Tolerance Distribution

MU	SIGMA
3.91831408	2.45483889

Estimated Covariance Matrix for Tolerance Parameters

	MU	SIGMA
MU	0.007754	0.002042
SIGMA	0.002042	0.017892

Logistic regression : Toxicity experiment
The Probit Analysis alternative (PROC PROBIT)

Probit Procedure

Probit Analysis on X

Probability	X	95% Fiducial Limits	
0.01	-1.7925	-2.4662	-1.2441
0.02	-1.1233	-1.7203	-0.6362
0.03	-0.6987	-1.2475	-0.2500
0.04	-0.3793	-0.8921	0.0408
0.05	-0.1195	-0.6034	0.2776
0.06	0.1016	-0.3578	0.4793
0.07	0.2955	-0.1426	0.6564
0.08	0.4691	0.0499	0.8152
0.09	0.6270	0.2248	0.9597
0.10	0.7723	0.3856	1.0929
0.15	1.3740	1.0495	1.6463
0.20	1.8523	1.5739	2.0894
0.25	2.2626	2.0202	2.4732
0.30	2.6310	2.4167	2.8220
0.35	2.9724	2.7793	3.1502
0.40	3.2964	3.1175	3.4673
0.45	3.6098	3.4385	3.7805
0.50	3.9183	3.7479	4.0952
0.55	4.2268	4.0512	4.4160
0.60	4.5402	4.3539	4.7474
0.65	4.8642	4.6621	5.0946
0.70	5.2056	4.9831	5.4643
0.75	5.5741	5.3263	5.8665
0.80	5.9844	5.7057	6.3171
0.85	6.4626	6.1454	6.8449
0.90	7.0643	6.6961	7.5115
0.91	7.2097	6.8289	7.6729
0.92	7.3675	6.9729	7.8482
0.93	7.5411	7.1312	8.0411
0.94	7.7350	7.3078	8.2567
0.95	7.9562	7.5092	8.5028
0.96	8.2160	7.7455	8.7920
0.97	8.5354	8.0359	9.1478
0.98	8.9599	8.4216	9.6211
0.99	9.6291	9.0288	10.3676