July 17, 2015

SAS code & output for Practical Guide to Logistic Regression

Joseph M Hilbe, *Arizona State University*

Joseph M. Hilbe: *Practical Guide to Logistic Regression*
Full SAS Code & Output by Yang Liu

Table of Contents

Chapter 1 - SAS Code & Output 2
Section 1.4 2
Chapter 2 - SAS Code & Output 6
Section 2.1 6
Section 2.2 12
Section 2.3 14
Section 2.4 23
Section 2.5 33
Section 2.6 41
Chapter 3 - SAS Code & Output 47
Section 3.1 47
Section 3.2 51
Section 3.3 57
Section 3.4 60
Section 3.5 68
Chapter 4 - SAS Code & Output 73
Section 4.1 73
Section 4.2 86
Section 4.3 92
Section 4.4 95
Section 4.5 100
Section 4.6 108
Chapter 5 - SAS Code & Output 113
Section 5.2 113
Section 5.4 118
Section 5.5 122
Chapter 6 - SAS Code & Output 132
Section 6.2 132

Chapter 1 - SAS Code & Output

Section 1.4

```
*Import medpar as a temporary dataset;
proc import datafile="c:\data\medpar.dta" out=medpar dbms=dta replace;
run;

*Print the first six observations;
proc print data=medpar (obs=6);
run;
```

Obs	los	hmo	white	died	age80	type	type1	type2	type3	provnum
1	4	0	1	0	0	1	1	0	0	30001
2	9	1	1	0	0	1	1	0	0	30001
3	3	1	1	1	1	1	1	0	0	30001
4	9	0	1	0	0	1	1	0	0	30001
5	1	0	1	1	1	1	1	0	0	30001
6	4	0	1	1	0	1	1	0	0	30001

```
*Build the logistic model;
proc genmod data=medpar descending;
      model died=hmo white/ dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-960.3010	
Full Log Likelihood		-960.3010	
AIC (smaller is better)		1926.6020	
AICC (smaller is better)		1926.6181	
BIC (smaller is better)		1942.5316	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.9262	0.1974	-1.3131	-0.5393	22.02	<.0001
hmo	1	-0.0122	0.1489	-0.3041	0.2796	0.01	0.9345
white	1	0.3034	0.2052	-0.0988	0.7055	2.19	0.1392
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Another way to build the logistic model;
proc logistic data=medpar descending;
      model died=hmo white / clparm=both;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Response Variable	died
Number of Response Levels	2

Model Information	
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1495
Number of Observations Used	1495

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

Probability modeled is died='1'.

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1924.865	1926.602
SC	1930.175	1942.532
-2 Log L	1922.865	1920.602

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	2.2633	2	0.3225
Score	2.1999	2	0.3329
Wald	2.1859	2	0.3352

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-0.9261	0.1974	22.0149	<.0001
hmo	1	-0.0122	0.1489	0.0068	0.9345
white	1	0.3033	0.2052	2.1859	0.1393

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
hmo	0.988	0.738	1.323
white	1.354	0.906	2.025

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	20.5	**Somers' D**	0.025
Percent Discordant	17.9	**Gamma**	0.066
Percent Tied	61.6	**Tau-a**	0.011
Pairs	503766	**c**	0.513

Parameter Estimates and Profile-Likelihood Confidence Intervals			
Parameter	**Estimate**	**95% Confidence Limits**	
Intercept	-0.9261	-1.3251	-0.5488
hmo	-0.0122	-0.3075	0.2769
white	0.3033	-0.0900	0.7166

Parameter Estimates and Wald Confidence Intervals			
Parameter	**Estimate**	**95% Confidence Limits**	
Intercept	-0.9261	-1.3130	-0.5393
hmo	-0.0122	-0.3041	0.2796
white	0.3033	-0.0988	0.7055

Chapter 2 - SAS Code & Output

Section 2.1

```
*Create a new dataset with binary variables x and y;
data xdta;
      input x y @@;
      datalines;
 1 1 0 0 1 0 0 1 1
 0 1 1 1 0 0 1 0 1
 ;
 run;


*Build the logistic model;
proc genmod data=xdta descending;
      model y=x / dist=binomial link=logit;
      output out=residual resdev=deviance;
run;
```

Model Information	
Data Set	WORK.XDTA
Distribution	Binomial
Link Function	Logit
Dependent Variable	y

Number of Observations Read	9
Number of Observations Used	9
Number of Events	5
Number of Trials	9

Response Profile		
Ordered Value	**y**	**Total Frequency**
1	1	5
2	0	4

PROC GENMOD is modeling the probability that y='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-5.6144	
Full Log Likelihood		-5.6144	
AIC (smaller is better)		15.2288	
AICC (smaller is better)		17.2288	
BIC (smaller is better)		15.6232	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.0986	1.1547	-1.1646	3.3618	0.91	0.3414
x	1	-1.5041	1.4720	-4.3891	1.3809	1.04	0.3069
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Another way to build the logistic model;
proc logistic data=xdta descending;
      model y=x / clparm=both;
      output out=residual resdev=deviance;
run;
```

Model Information	
Data Set	WORK.XDTA
Response Variable	y
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	9
Number of Observations Used	9

Response Profile		
Ordered Value	y	Total Frequency
1	1	5
2	0	4

Probability modeled is y=1.

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	14.365	15.229
SC	14.563	15.623
-2 Log L	12.365	11.229

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1.1365	1	0.2864
Score	1.1025	1	0.2937
Wald	1.0441	1	0.3069

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.0986	1.1547	0.9052	0.3414
x	1	-1.5041	1.4720	1.0441	0.3069

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
x	0.222	0.012	3.979

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	45.0	Somers' D	0.350
Percent Discordant	10.0	Gamma	0.636
Percent Tied	45.0	Tau-a	0.194
Pairs	20	c	0.675

Parameter Estimates and Profile-Likelihood Confidence Intervals			
Parameter	Estimate	95% Confidence Limits	
Intercept	1.0986	-0.9565	4.1046
x	-1.5041	-4.9255	1.2199

Parameter Estimates and Wald Confidence Intervals			
Parameter	Estimate	95% Confidence Limits	
Intercept	1.0986	-1.1646	3.3618
x	-1.5041	-4.3891	1.3809

```
*Statistics of deviance residual;
proc means data=residual min q1 median q3 max maxdec=4;
      var deviance;
run;
```

Analysis Variable : deviance Deviance Residual				
Minimum	Lower Quartile	Median	Upper Quartile	Maximum
-1.6651	-1.0108	0.7585	0.7585	1.3537

```
*Generate a table of y by x;
proc freq data=xdta;
      tables y*x / norow nocol nocum nopercent;
run;
```

Table of y by x			
y	x		
Frequency	0	1	Total
0	1	3	4
1	3	2	5
Total	4	5	9

```
*Estimate statement with exp option provides the odds ratio;
proc genmod data=xdta descending;
      model y=x / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "x" x 1 / exp;
run;
```

Model Information	
Data Set	WORK.XDTA
Response Variable	y
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	9
Number of Observations Used	9

Response Profile		
Ordered Value	y	Total Frequency
1	1	5
2	0	4

Probability modeled is y=1.

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	14.365	15.229
SC	14.563	15.623
-2 Log L	12.365	11.229

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	1.1365	1	0.2864
Score	1.1025	1	0.2937
Wald	1.0441	1	0.3069

Analysis of Maximum Likelihood Estimates						
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**	**Exp(Est)**
Intercept	1	1.0986	1.1547	0.9052	0.3414	3.000
x	1	-1.5041	1.4720	1.0441	0.3069	0.222

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
x	0.222	0.012	3.979

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	45.0	**Somers' D**	0.350
Percent Discordant	10.0	**Gamma**	0.636
Percent Tied	45.0	**Tau-a**	0.194
Pairs	20	**c**	0.675

Section 2.2

```
*Build the logistic model;
proc genmod data=xdta descending;
      model y=x / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.XDTA
Distribution	Binomial
Link Function	Logit
Dependent Variable	y

Number of Observations Read	9
Number of Observations Used	9
Number of Events	5
Number of Trials	9

Response Profile		
Ordered Value	**y**	**Total Frequency**
1	1	5
2	0	4

PROC GENMOD is modeling the probability that y='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-5.6144	
Full Log Likelihood		-5.6144	
AIC (smaller is better)		15.2288	
AICC (smaller is better)		17.2288	
BIC (smaller is better)		15.6232	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.0986	1.1547	-1.1646	3.3618	0.91	0.3414
x	1	-1.5041	1.4720	-4.3891	1.3809	1.04	0.3069
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Create a dataset to make calculations;
data data1;
      set xdta;
      if x=1 then xb=1.0986-1.5041*1;
      else if x=0 then xb=1.0986-1.5041*0;
      mu=1/(1+exp(-xb));
      o=mu/(1-mu);
      or=o;
      if or < 1 then or=0.667/3;
      coeff=log(or);
      format mu 4.2 o or xb coeff 7.4;
run;

*Print the dataset;
proc print data=data1;
      var x o;
run;
```

Obs	x	o
1	1	0.6666
2	0	3.0000
3	1	0.6666
4	0	3.0000
5	1	0.6666
6	1	0.6666
7	1	0.6666
8	0	3.0000
9	0	3.0000

```
*Print the whole dataset;
proc print data=data1;
run;
```

Obs	x	y	xb	mu	o	or	coeff
1	1	1	-0.4055	0.40	0.6666	0.2223	-1.5036
2	0	0	1.0986	0.75	3.0000	3.0000	1.0986
3	1	0	-0.4055	0.40	0.6666	0.2223	-1.5036
4	0	1	1.0986	0.75	3.0000	3.0000	1.0986
5	1	0	-0.4055	0.40	0.6666	0.2223	-1.5036
6	1	1	-0.4055	0.40	0.6666	0.2223	-1.5036
7	1	0	-0.4055	0.40	0.6666	0.2223	-1.5036
8	0	1	1.0986	0.75	3.0000	3.0000	1.0986
9	0	1	1.0986	0.75	3.0000	3.0000	1.0986

Section 2.3

```
*Build the logistic model- covb option provides var-cov matrix;
proc genmod data=xdta descending;
      model y=x / dist=binomial  link=logit covb;
run;
```

Model Information	
Data Set	WORK.XDTA
Distribution	Binomial
Link Function	Logit
Dependent Variable	y

Number of Observations Read	9
Number of Observations Used	9
Number of Events	5
Number of Trials	9

Response Profile		
Ordered Value	y	Total Frequency
1	1	5
2	0	4

PROC GENMOD is modeling the probability that y='1'.

Parameter Information	
Parameter	**Effect**
Prm1	Intercept
Prm2	x

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-5.6144	
Full Log Likelihood		-5.6144	
AIC (smaller is better)		15.2288	
AICC (smaller is better)		17.2288	
BIC (smaller is better)		15.6232	

Algorithm converged.

Estimated Covariance Matrix		
	Prm1	**Prm2**
Prm1	1.33333	-1.33333
Prm2	-1.33333	2.16667

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.0986	1.1547	-1.1646	3.3618	0.91	0.3414
x	1	-1.5041	1.4720	-4.3891	1.3809	1.04	0.3069
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Use SAS interactive matrix language;
proc iml;
      vcov={1.33333 -1.33333,
           -1.33333  2.16667};
```

```
      se=sqrt(diag(vcov));
      print se;
quit;
```

se	
1.1546991	0
0	1.4719613

```
*Logistic regression with OIM standard error;
proc surveylogistic data=xdta;
      model y(event='1')=x;
run;
```

Model Information	
Data Set	WORK.XDTA
Response Variable	y
Number of Response Levels	2
Model	Binary Logit
Optimization Technique	Fisher's Scoring
Variance Adjustment	Degrees of Freedom (DF)

Variance Estimation	
Method	Taylor Series
Variance Adjustment	Degrees of Freedom (DF)

Number of Observations Read	9
Number of Observations Used	9

Response Profile		
Ordered Value	**y**	**Total Frequency**
1	0	4
2	1	5

Probability modeled is y=1.

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	14.365	15.229
SC	14.563	15.623
-2 Log L	12.365	11.229

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	1.1365	1	0.2864
Score	1.1025	1	0.2937
Wald	0.8121	1	0.3675

Analysis of Maximum Likelihood Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.0986	1.3093	0.7041	0.4014
x	1	-1.5041	1.6690	0.8121	0.3675

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
x	0.222	0.008	5.854

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	45.0	**Somers' D**	0.350
Percent Discordant	10.0	**Gamma**	0.636
Percent Tied	45.0	**Tau-a**	0.194
Pairs	20	**c**	0.675

```
*Covb option provides var-cov matrix;
proc genmod data=xdta descending;
      model y=x / dist=binomial link=logit covb;
run;
```

Model Information	
Data Set	WORK.XDTA
Distribution	Binomial
Link Function	Logit
Dependent Variable	y

Number of Observations Read	9
Number of Observations Used	9
Number of Events	5
Number of Trials	9

Response Profile		
Ordered Value	**y**	**Total Frequency**
1	1	5
2	0	4

PROC GENMOD is modeling the probability that y='1'.

Parameter Information	
Parameter	**Effect**
Prm1	Intercept
Prm2	x

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-5.6144	
Full Log Likelihood		-5.6144	
AIC (smaller is better)		15.2288	

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
AICC (smaller is better)		17.2288	
BIC (smaller is better)		15.6232	

Algorithm converged.

Estimated Covariance Matrix		
	Prm1	**Prm2**
Prm1	1.33333	-1.33333
Prm2	-1.33333	2.16667

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.0986	1.1547	-1.1646	3.3618	0.91	0.3414
x	1	-1.5041	1.4720	-4.3891	1.3809	1.04	0.3069
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Calculations of odds ratio and model statistics;
proc iml;
      vcov={1.33333 -1.33333,
           -1.33333  2.16667};
      coef={1.0986, -1.5041};
      or=exp(coef);
      se=sqrt(diag(vcov));
      ose=se*or;
      print or [format = 7.4] ose [format = 7.4];

      zscore=coef/se;
      delta=ose;
      z=zscore[,+];
      pvalue=2*(1-probnorm((abs(z))));
      print z pvalue;

      se1=se[,+];
      loci=coef-quantile('normal', 0.975)*se1;
      upci=coef+quantile('normal', 0.975)*se1;
      expl=exp(loci);
```

```
      expu=exp(upci);
      print or [format=7.4] delta [format=7.4] z [format=7.4]
            pvalue [format=7.4] expl [format=7.4] expu [format=7.4];
quit;
```

se	
1.1546991	0
0	1.4719613

or	ose
3	3.4640973
0.2222	0.3270698

zscore	
0.9514273	.
.	-1.021887

se	
1.1546991	0
0	1.4719613

loci	upci
-1.164556	3.3617809
-4.389074	1.3807191

```
*Clparm=both provides both PL and Wald confidence intervals;
proc logistic data=xdta descending;
      model y=x / clparm=both;
run;
```

Model Information	
Data Set	WORK.XDTA
Response Variable	y
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	9
Number of Observations Used	9

Response Profile		
Ordered Value	y	Total Frequency
1	1	5
2	0	4

Probability modeled is y=1.

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	14.365	15.229
SC	14.563	15.623
-2 Log L	12.365	11.229

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1.1365	1	0.2864
Score	1.1025	1	0.2937
Wald	1.0441	1	0.3069

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.0986	1.1547	0.9052	0.3414
x	1	-1.5041	1.4720	1.0441	0.3069

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
x	0.222	0.012	3.979

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	45.0	**Somers' D**	0.350
Percent Discordant	10.0	**Gamma**	0.636
Percent Tied	45.0	**Tau-a**	0.194
Pairs	20	**c**	0.675

Parameter Estimates and Profile-Likelihood Confidence Intervals			
Parameter	**Estimate**	**95% Confidence Limits**	
Intercept	1.0986	-0.9565	4.1046
x	-1.5041	-4.9255	1.2199

Parameter Estimates and Wald Confidence Intervals			
Parameter	**Estimate**	**95% Confidence Limits**	
Intercept	1.0986	-1.1646	3.3618
x	-1.5041	-4.3891	1.3809

Section 2.4

```
*Import medpar as a temporary dataset;
proc import datafile="c:\ado\medpar.dta" out=medpar dbms=dta replace;
run;

*Print the first six observations;
proc print data=medpar (obs=6);
run;
```

Obs	los	hmo	white	died	age80	type	type1	type2	type3	provnum
1	4	0	1	0	0	1	1	0	0	30001
2	9	1	1	0	0	1	1	0	0	30001
3	3	1	1	1	1	1	1	0	0	30001
4	9	0	1	0	0	1	1	0	0	30001
5	1	0	1	1	1	1	1	0	0	30001
6	4	0	1	1	0	1	1	0	0	30001

```
*Generate the frequency table of type and output the dataset;
proc freq data=medpar;
      tables type / out=freq;
run;
```

type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1134	75.85	1134	75.85
2	265	17.73	1399	93.58
3	96	6.42	1495	100.00

```
*Build the logistic model with class;
proc genmod data=medpar descending;
      class type (ref='1') / param = ref;
      model died=type / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "type2" type 1 0 / exp;
      estimate "type3" type 0 1 / exp;
      output out=residual resdev=deviance;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	**Value**	**Design Variables**	
type	**1**	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Parameter Information		
Parameter	**Effect**	**type**
Prm1	Intercept	
Prm2	type	2
Prm3	type	3

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-955.5718	
Full Log Likelihood		-955.5718	
AIC (smaller is better)		1917.1436	
AICC (smaller is better)		1917.1597	
BIC (smaller is better)		1933.0733	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7492	0.0636	-0.8739	-0.6246	138.74	<.0001
type	2	1	0.3122	0.1410	0.0359	0.5885	4.91	0.0268
type	3	1	0.6241	0.2142	0.2043	1.0439	8.49	0.0036
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	0.3210	0.2944	0.3487	-0.7492	0.0636	0.05	-0.8739	-0.6246	138.74	<.0001
Exp(Intercept)				0.4727	0.0301	0.05	0.4173	0.5355		
type2	0.5774	0.5090	0.6430	0.3122	0.1410	0.05	0.0359	0.5885	4.91	0.0268
Exp(type2)				1.3665	0.1926	0.05	1.0366	1.8013		
type3	0.6511	0.5509	0.7396	0.6241	0.2142	0.05	0.2043	1.0439	8.49	0.0036
Exp(type3)				1.8665	0.3998	0.05	1.2266	2.8402		

```
*Set up format for variable type;
proc format;
      value typefmt 1="Elective Admit"
                    2="Urgent Admit"
                    3="Emergency Admit";
run;

*Logistic regression with controlled reference;
proc genmod data=medpar descending;
      class type (ref='Elective Admit') / param = ref;
      model died=type / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "type2" type 1 0 / exp;
      estimate "type3" type 0 1 / exp;
      format type typefmt.;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	**Value**	**Design Variables**	
type	**Elective Admit**	0	0
	Emergency Admit	1	0
	Urgent Admit	0	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Parameter Information		
Parameter	**Effect**	**type**
Prm1	Intercept	
Prm2	type	Emergency Admit
Prm3	type	Urgent Admit

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-955.5718	
Full Log Likelihood		-955.5718	
AIC (smaller is better)		1917.1436	
AICC (smaller is better)		1917.1597	
BIC (smaller is better)		1933.0733	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7492	0.0636	-0.8739	-0.6246	138.74	<.0001
type	Emergency Admit	1	0.6241	0.2142	0.2043	1.0439	8.49	0.0036
type	Urgent Admit	1	0.3122	0.1410	0.0359	0.5885	4.91	0.0268
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
Label	**Mean Estimate**	**Mean Confidence Limits**		**L'Beta Estimate**	**Standard Error**	**Alpha**	**L'Beta Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
Intercept	0.3210	0.2944	0.3487	-0.7492	0.0636	0.05	-0.8739	-0.6246	138.74	<.0001
Exp(Intercept)				0.4727	0.0301	0.05	0.4173	0.5355		
type2	0.6511	0.5509	0.7396	0.6241	0.2142	0.05	0.2043	1.0439	8.49	0.0036
Exp(type2)				1.8665	0.3998	0.05	1.2266	2.8402		
type3	0.5774	0.5090	0.6430	0.3122	0.1410	0.05	0.0359	0.5885	4.91	0.0268
Exp(type3)				1.3665	0.1926	0.05	1.0366	1.8013		

```
*Logistic regression with controlled reference;
proc genmod data=medpar descending;
      class type (ref='Emergency Admit') / param = ref;
      model died=type / dist=binomial link=logit;
```

```
        estimate "Intercept" Intercept 1 / exp;
        estimate "type2" type 1 0 / exp;
        estimate "type3" type 0 1 / exp;
        format type typefmt.;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	**Value**	**Design Variables**	
type	**Elective Admit**	1	0
	Emergency Admit	0	0
	Urgent Admit	0	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Parameter Information		
Parameter	**Effect**	**type**
Prm1	Intercept	
Prm2	type	Elective Admit
Prm3	type	Urgent Admit

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-955.5718	
Full Log Likelihood		-955.5718	
AIC (smaller is better)		1917.1436	
AICC (smaller is better)		1917.1597	
BIC (smaller is better)		1933.0733	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.1252	0.2045	-0.5260	0.2757	0.37	0.5406
type	Elective Admit	1	-0.6241	0.2142	-1.0439	-0.2043	8.49	0.0036
type	Urgent Admit	1	-0.3119	0.2401	-0.7825	0.1588	1.69	0.1940
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
		Mean					**L'Beta**			
Label	**Mean Estimate**	**Confidence Limits**		**L'Beta Estimate**	**Standard Error**	**Alpha**	**Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
Intercept	0.4688	0.3714	0.5685	-0.1252	0.2045	0.05	-0.5260	0.2757	0.37	0.5406
Exp(Intercept)				0.8824	0.1805	0.05	0.5910	1.3174		
type2	0.3489	0.2604	0.4491	-0.6241	0.2142	0.05	-1.0439	-0.2043	8.49	0.0036

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Exp(type2)				0.5358	0.1148	0.05	0.3521	0.8152		
type3	0.4227	0.3138	0.5396	-0.3119	0.2401	0.05	-0.7825	0.1588	1.69	0.1940
Exp(type3)				0.7321	0.1758	0.05	0.4573	1.1721		

```
*Generate the frequency table for variable type;
proc freq data=medpar;
      tables type;
      format type typefmt.;
run;
```

type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Elective Admit	1134	75.85	1134	75.85
Urgent Admit	265	17.73	1399	93.58
Emergency Admit	96	6.42	1495	100.00

```
*Re-categorized variable type;
data medpar1;
      set medpar;
      if type in (2,3) then type=2;
run;

*Generate the frequency table for re-categorized variable type;
proc freq data=medpar1;
      tables type;
run;
```

type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Elective Admit	1134	75.85	1134	75.85
Urgent Admit	361	24.15	1495	100.00

```
*Logistic regression with re-categorized type;
proc genmod data=medpar1 descending;
      class type (ref='1') / param = ref;
      model died=type / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "type2" type 1 0 / exp;
      estimate "type3" type 0 1 / exp;
run;
```

Model Information	
Data Set	WORK.MEDPAR1
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information		
Class	**Value**	**Design Variables**
type	**1**	0
	2	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Parameter Information		
Parameter	**Effect**	**type**
Prm1	Intercept	
Prm2	type	2

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-956.4123	
Full Log Likelihood		-956.4123	
AIC (smaller is better)		1916.8246	
AICC (smaller is better)		1916.8326	
BIC (smaller is better)		1927.4444	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7492	0.0636	-0.8739	-0.6246	138.74	<.0001
type	2	1	0.3966	0.1244	0.1528	0.6404	10.16	0.0014
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
		Mean					**L'Beta**			
Label	**Mean Estimate**	**Confidence Limits**		**L'Beta Estimate**	**Standard Error**	**Alpha**	**Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
Intercept	0.3210	0.2944	0.3487	-0.7492	0.0636	0.05	-0.8739	-0.6246	138.74	<.0001
Exp(Intercept)				0.4727	0.0301	0.05	0.4173	0.5355		
type2	0.5979	0.5381	0.6548	0.3966	0.1244	0.05	0.1528	0.6404	10.16	0.0014
Exp(type2)				1.4868	0.1849	0.05	1.1651	1.8973		
type3	0.5000	0.5000	0.5000	0.0000	0.0000	0.05	0.0000	0.0000	.	.
Exp(type3)				1.0000	0.0000	0.05	1.0000	1.0000		

Section 2.5

```
*Summary for variable los;
proc means data=medpar min q1 median mean q3 max maxdec=3;
      var los;
run;
```

Analysis Variable : los					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
1.000	4.000	8.000	9.854	13.000	116.000

```
*Build the generalized additive model;
proc gam data=medpar;
      model died (event='1')=spline(los) / dist=binomial;
run;
```

Summary of Input Data Set	
Number of Observations	1495
Number of Missing Observations	0
Distribution	Binomial
Link Function	Logit

Response Profile		
Ordered Value	died	Total Frequency
1	0	982
2	1	513

Note: PROC GAM is modeling the probability that died=1. One way to change this to model the probability that died=0 is to speci
response variable option EVENT='0'.

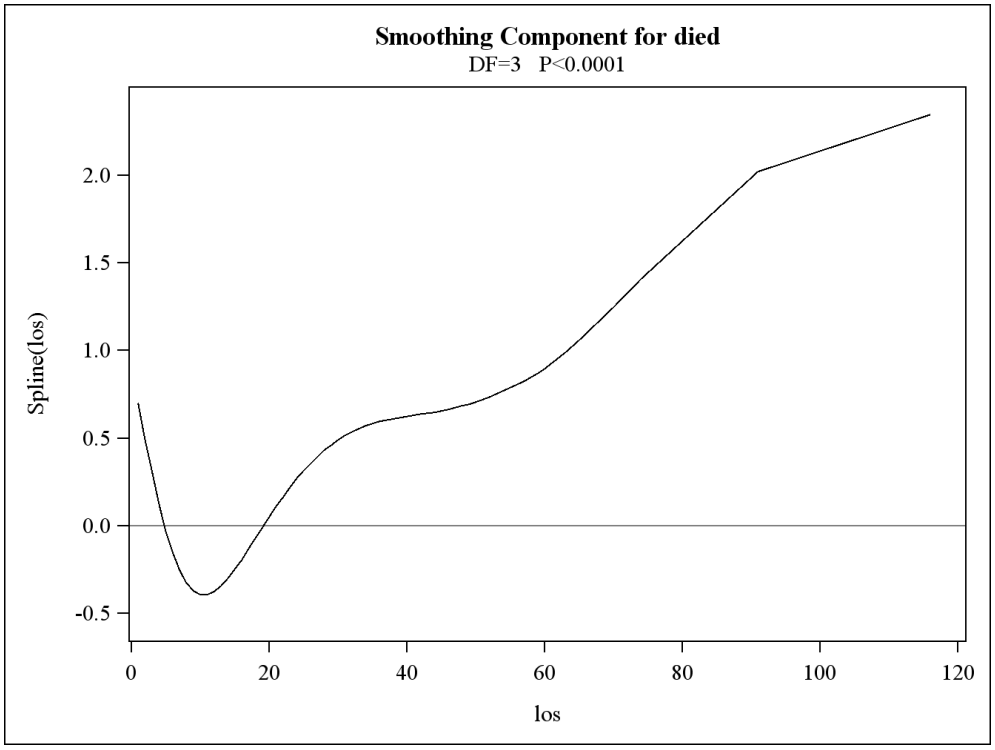
Iteration Summary and Fit Statistics	
Number of local scoring iterations	5
Local scoring convergence criterion	1.404456E-12
Final Number of Backfitting Iterations	1
Final Backfitting Criterion	1.24205E-11
The Deviance of the Final Estimate	1815.5324882

The local scoring algorithm converged.

Regression Model Analysis Parameter Estimates				
Parameter	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	-0.38942	0.08169	-4.77	<.0001
Linear(los)	-0.02582	0.00633	-4.08	<.0001

Smoothing Model Analysis Fit Summary for Smoothing Components				
Component	**Smoothing Parameter**	**DF**	**GCV**	**Num Unique Obs**
Spline(los)	0.999495	3.000000	6.130299	52

Smoothing Model Analysis Analysis of Deviance				
Source	**DF**	**Sum of Squares**	**Chi-Square**	**Pr > ChiSq**
Spline(los)	3.00000	89.017950	89.0180	<.0001



```
*Import badhealth as a temporary dataset;
proc import datafile="c:\data\badhealth.dta" out=badhealth dbms=dta
replace;
run;

*Print the first six observations;
proc print data=badhealth (obs=6);
run;
```

Obs	numvisit	badh	age
1	30	0	58
2	20	0	54
3	16	0	44
4	20	0	57
5	15	0	33
6	15	0	28

```
*Generate the frequency table for variable badh;
proc freq data=badhealth;
      tables badh / nocum nopercent;
run;
```

badh	Frequency
0	1015
1	112

```
*Summary for variable age;
proc means data=badhealth min q1 median mean q3 max maxdec=2;
      var age;
      output out=center mean=;
run;
```

Analysis Variable : age					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
20.00	28.00	35.00	37.23	46.00	60.00

```
*Create a macro variable;
proc sql;
      select age into: mean
      from center;
quit;
```

age
37.228926353

```
*Summary for variable numvisit;
proc means data=badhealth min q1 median mean q3 max maxdec=3;
      var numvisit;
run;
```

Analysis Variable : numvisit					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
0.000	0.000	1.000	2.353	3.000	40.000

```
*Center the age;
data badhealth1;
      set badhealth;
      cage=age-&mean;
run;

*Summary for centered age;
proc means data=badhealth1 min q1 median mean q3 max maxdec=3;
      var cage;
run;
```

Analysis Variable : cage					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
-17.229	-9.229	-2.229	0.000	8.771	22.771

```
*Provide the std;
proc means data=badhealth std;
      var age;
      output out=stderror std=;
run;
```

Analysis Variable : age
Std Dev
10.8289190

```
*Create a macro variable;
proc sql;
      select age into: std
      from stderror;
quit;
```

age
10.828918958

```
*Scale age with a different way;
proc standard data=badhealth mean=0 std=&std out=cenage;
      var age;
```

```
run;

*Build the logistic model;
proc genmod data=badhealth descending;
      model badh=age / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.BADHEALTH
Distribution	Binomial
Link Function	Logit
Dependent Variable	badh

Number of Observations Read	1127
Number of Observations Used	1127
Number of Events	112
Number of Trials	1127

Response Profile		
Ordered Value	**badh**	**Total Frequency**
1	1	112
2	0	1015

PROC GENMOD is modeling the probability that badh='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-343.8462	
Full Log Likelihood		-343.8462	
AIC (smaller is better)		691.6925	
AICC (smaller is better)		691.7031	
BIC (smaller is better)		701.7471	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.5887	0.4173	-5.4065	-3.7708	120.92	<.0001
age	1	0.0595	0.0095	0.0410	0.0781	39.62	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Build the logistic model with centered age;
proc genmod data=badhealth1 descending ;
      model badh=cage / dist=binomial  link=logit;
run;
```

Model Information	
Data Set	WORK.BADHEALTH1
Distribution	Binomial
Link Function	Logit
Dependent Variable	badh

Number of Observations Read	1127
Number of Observations Used	1127
Number of Events	112
Number of Trials	1127

Response Profile		
Ordered Value	badh	Total Frequency
1	1	112
2	0	1015

PROC GENMOD is modeling the probability that badh='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-343.8462	
Full Log Likelihood		-343.8462	
AIC (smaller is better)		691.6925	
AICC (smaller is better)		691.7031	
BIC (smaller is better)		701.7471	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.3717	0.1144	-2.5959	-2.1475	429.93	<.0001
cage	1	0.0595	0.0095	0.0410	0.0781	39.62	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Standardize age and output the sage dataset;
proc standard data=badhealth mean=0 std=1 out=sage;
      var age;
run;

*Build the logistic model with standardized age;
proc genmod data=sage descending ;
      model badh=age / dist=binomial  link=logit;
run;
```

Model Information	
Data Set	WORK.SAGE
Distribution	Binomial
Link Function	Logit
Dependent Variable	badh

Number of Observations Read	1127
Number of Observations Used	1127
Number of Events	112
Number of Trials	1127

Response Profile		
Ordered Value	**badh**	**Total Frequency**
1	1	112
2	0	1015

PROC GENMOD is modeling the probability that badh='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-343.8462	
Full Log Likelihood		-343.8462	
AIC (smaller is better)		691.6925	
AICC (smaller is better)		691.7031	
BIC (smaller is better)		701.7471	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.3717	0.1144	-2.5959	-2.1475	429.93	<.0001
age	1	0.6449	0.1024	0.4441	0.8456	39.62	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Section 2.6

```
*Build the logistic model and output model prediction;
proc genmod data=medpar descending;
      model died=white / dist=binomial link=logit;
      output out=etab pred=fitb;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-960.3044	
Full Log Likelihood		-960.3044	
AIC (smaller is better)		1924.6088	
AICC (smaller is better)		1924.6168	
BIC (smaller is better)		1935.2285	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.9273	0.1969	-1.3132	-0.5414	22.18	<.0001
white	1	0.3025	0.2049	-0.0991	0.7041	2.18	0.1398
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Generate the frequency table for fitb;
proc freq data=etab;
      tables fitb / nocum nopercent;
run;
```

Predicted Value	
fitb	Frequency
0.2834645677	127
0.3486842127	1368

```
*Build the logistic model and output model prediction;
proc genmod data=medpar descending;
      model died=white / dist=binomial link=logit;
      output out=etac pred=fitc;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	Died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-960.3044	
Full Log Likelihood		-960.3044	
AIC (smaller is better)		1924.6088	
AICC (smaller is better)		1924.6168	
BIC (smaller is better)		1935.2285	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.9273	0.1969	-1.3132	-0.5414	22.18	<.0001
white	1	0.3025	0.2049	-0.0991	0.7041	2.18	0.1398
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Summary for variable fitc;
proc means data=etac min q1 median mean q3 max maxdec=5;
      var fitc;
run;
```

Analysis Variable : fitc Predicted Value					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
0.28346	0.34868	0.34868	0.34314	0.34868	0.34868

```
*Create a dataset to make calculations;
data prob;
      xb20=-0.3617 - 0.0305*20;
      mu20=1/(1+exp(-xb20));
run;

*Print the variable mu20;
proc print data=prob;
      var mu20;
run;
```

Obs	mu20
1	0.27454

```
*Build the logistic model and output confidence intervals;
proc genmod data=medpar descending;
      model died=los / dist=binomial  link=logit;
      output out=cl pred=mu lower=loci upper=upci;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	Died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-952.2752	
Full Log Likelihood		-952.2752	
AIC (smaller is better)		1908.5505	
AICC (smaller is better)		1908.5585	
BIC (smaller is better)		1919.1703	

Algorithm converged.

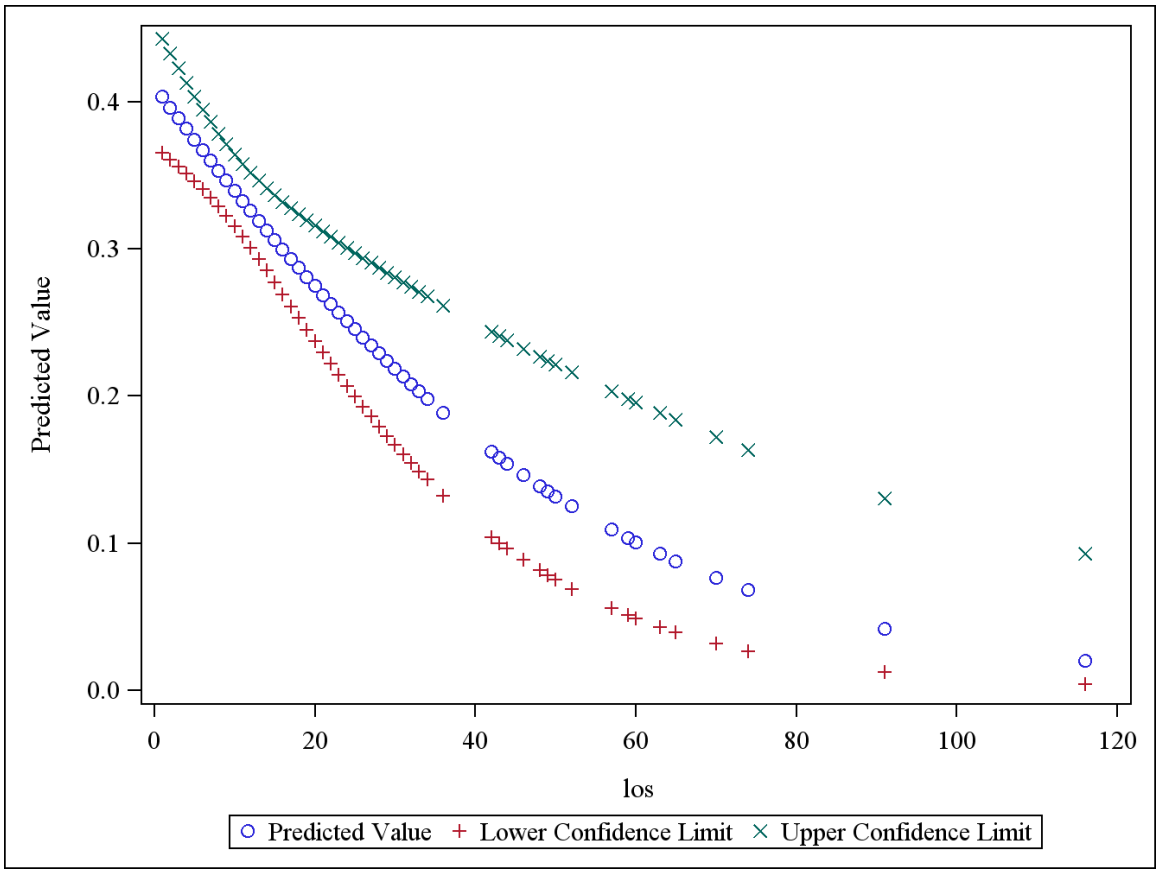
Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-0.3617	0.0884	-0.5350	-0.1884	16.73	<.0001
los	1	-0.0305	0.0077	-0.0456	-0.0154	15.71	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Summary for confidence intervals;
proc means data=cl min q1 median mean q3 max maxdec=5;
      var loci mu upci;
run;
```

Variable	Label	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
loci	Lower Confidence Limit	0.00401	0.29298	0.32870	0.31289	0.35091	0.36495
mu	Predicted Value	0.01988	0.31908	0.35307	0.34314	0.38139	0.40319
upci	Upper Confidence Limit	0.09265	0.34637	0.37823	0.37541	0.41284	0.44265

```
*Graph scatter plot;
proc sgplot data=cl;
      scatter x=los y=mu;
      scatter x=los y=loci;
      scatter x=los y=upci;
run;
```


Predicted Value
0.4
0.3
0.2
0.1
0.0
0
20
40
60
80
100
120
los
Predicted Value
Lower Confidence Limit
Upper Confidence Limit

Chapter 3 - SAS Code & Output

Section 3.1

```
*Import edrelig as a temporary dataset;
proc import datafile="c:\data\edrelig.dta" out=edrelig dbms=dta
replace;
run;

*Print the first six observations;
proc print data=edrelig (obs=6);
run;
```

Obs	male	age	kids	educlevel	religious
1	1	37	0	MA/PhD	0
2	0	27	0	AA	1
3	1	27	0	MA/PhD	0
4	0	32	1	AA	0
5	0	27	1	BA	0
6	1	57	1	MA/PhD	1

```
*Generate a table of educlevel;
proc freq data=edrelig;
      tables educlevel / nocum nopercent;
run;
```

educlevel	Frequency
AA	205
BA	204
MA/PhD	192

```
*Build logistic model and obtain odds ratio & covariance matrix;
proc genmod data=edrelig descending;
      class educlevel (ref='AA') / param = ref;
      model religious=age male kids educlevel/dist=binomial link=logit
      covb;
      estimate "Intercept" Intercept 1 / exp;
      estimate "Age" age 1 / exp;
      estimate "Male" male 1 / exp;
      estimate "Kid" kids 1 / exp;
      estimate "BA" educlevel 1 0 / exp;
      estimate "MA/PhD" educlevel 0 1 / exp;
run;
```

Model Information	
Data Set	WORK.EDRELIG
Distribution	Binomial
Link Function	Logit
Dependent Variable	religious

Number of Observations Read	601
Number of Observations Used	601
Number of Events	260
Number of Trials	601

Class Level Information			
Class	**Value**	**Design Variables**	
educlevel	**AA**	0	0
	BA	1	0
	MA/PhD	0	1

Response Profile		
Ordered Value	**religious**	**Total Frequency**
1	1	260
2	0	341

PROC GENMOD is modeling the probability that religious='1'.

Parameter Information		
Parameter	**Effect**	**educlevel**
Prm1	Intercept	
Prm2	age	
Prm3	male	
Prm4	kids	
Prm5	educlevel	BA
Prm6	educlevel	MA/PhD

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-396.4198	
Full Log Likelihood		-396.4198	
AIC (smaller is better)		804.8396	
AICC (smaller is better)		804.9810	
BIC (smaller is better)		831.2311	

Algorithm converged.

Estimated Covariance Matrix						
	Prm1	**Prm2**	**Prm3**	**Prm4**	**Prm5**	**Prm6**
Prm1	0.10887	-0.002631	-0.003488	-0.006413	-0.02009	-0.008667
Prm2	-0.002631	0.0001073	-0.000201	-0.000882	-0.000099	-0.000310
Prm3	-0.003488	-0.000201	0.03449	0.0000180	-0.004585	-0.01575
Prm4	-0.006413	-0.000882	0.0000180	0.04425	0.005386	0.003250
Prm5	-0.02009	-0.000099	-0.004585	0.005386	0.04336	0.02292
Prm6	-0.008667	-0.000310	-0.01575	0.003250	0.02292	0.05117

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.4352	0.3300	-2.0819	-0.7885	18.92	<.0001
age		1	0.0398	0.0104	0.0195	0.0601	14.78	0.0001
male		1	0.1900	0.1857	-0.1740	0.5540	1.05	0.3064
kids		1	0.1239	0.2104	-0.2884	0.5362	0.35	0.5558
educlevel	BA	1	-0.4723	0.2082	-0.8804	-0.0642	5.15	0.0233
educlevel	MA/PhD	1	-0.4954	0.2262	-0.9388	-0.0521	4.80	0.0285
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	0.1923	0.1109	0.3125	-1.4352	0.3300	0.05	-2.0819	-0.7885	18.92	<.0001
Exp(Intercept)				0.2381	0.0786	0.05	0.1247	0.4545		
Age	0.5100	0.5049	0.5150	0.0398	0.0104	0.05	0.0195	0.0601	14.78	0.0001
Exp(Age)				1.0406	0.0108	0.05	1.0197	1.0620		
Male	0.5473	0.4566	0.6351	0.1900	0.1857	0.05	-0.1740	0.5540	1.05	0.3064
Exp(Male)				1.2092	0.2246	0.05	0.8403	1.7402		
Kid	0.5309	0.4284	0.6309	0.1239	0.2104	0.05	-0.2884	0.5362	0.35	0.5558
Exp(Kid)				1.1319	0.2381	0.05	0.7495	1.7096		
BA	0.3841	0.2931	0.4840	-0.4723	0.2082	0.05	-0.8804	-0.0642	5.15	0.0233
Exp(BA)				0.6236	0.1298	0.05	0.4146	0.9378		
MA/PhD	0.3786	0.2811	0.4870	-0.4954	0.2262	0.05	-0.9388	-0.0521	4.80	0.0285
Exp(MA/PhD)				0.6093	0.1378	0.05	0.3911	0.9493		

```
*Calculations of odds ratio and model statistics;
proc iml;
      vcov={0.10887 -0.002631 -0.003488 -0.006413 -0.02009 -0.008667,
           -0.002631 0.0001073 -0.000201 -0.000882 -0.000099 -0.000310,
           -0.003488 -0.000201 0.03449 0.0000180 -0.004585 -0.01575,
           -0.006413 -0.000882 0.0000180 0.04425 0.005386 0.003250,
           -0.02009 -0.000099 -0.004585 0.005386 0.04336 0.02292,
           -0.008667 -0.000310 -0.01575 0.003250 0.02292 0.05117};
      coef={-1.4352, 0.0398, 0.1900, 0.1239, -0.4723, -0.4954};
      se=sqrt(diag(vcov));
      se1=se[,+];
      zscore=coef/se;
      z=zscore[,+];
      or=exp(coef);
      print or;
      delta=se*or;
      pvalue=2*(1-probnorm((abs(z))));
      loci=coef-quantile('normal', 0.975)*se1;
      upci=coef+quantile('normal', 0.975)*se1;
      expl=exp(loci);
      expu=exp(upci);
      print or [format=7.4] delta [format=7.4] z [format=7.4]
            pvalue [format=7.4] expl [format=7.4] expu [format=7.4];
      odds1=1/or[5];
      odds2=1/or[6];
      print odds1 odds2;
quit;
```

or
0.2380677
1.0406026
1.2092496
1.1319027
0.6235664
0.6093271

or	delta	z	pvalue	expl	expu
0.2381	0.0786	-4.3497	0.0000	0.1247	0.4545
1.0406	0.0108	3.8422	0.0001	1.0197	1.0619
1.2092	0.2246	1.0231	0.3063	0.8403	1.7402
1.1319	0.2381	0.5890	0.5559	0.7495	1.7095
0.6236	0.1298	-2.2682	0.0233	0.4146	0.9378
0.6093	0.1378	-2.1900	0.0285	0.3911	0.9493

odds1	odds2
1.6036784	1.6411546

Section 3.2

```
*Build the logistic model and obtain the deviance residual;
proc genmod data=edrelig descending;
      class educlevel (ref='AA') / param = ref;
      model religious=age male kids educlevel/dist=binomial link=logit;
      output out=residual resdev=deviance;
run;
```

Model Information	
Data Set	WORK. EDRELIG
Distribution	Binomial
Link Function	Logit
Dependent Variable	religious

Number of Observations Read	601
Number of Observations Used	601
Number of Events	260
Number of Trials	601

Class Level Information			
Class	Value	Design Variables	
educlevel	AA	0	0
	BA	1	0
	MA/PhD	0	1

Response Profile		
Ordered Value	religious	Total Frequency
1	1	260
2	0	341

PROC GENMOD is modeling the probability that religious='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-396.4198	
Full Log Likelihood		-396.4198	
AIC (smaller is better)		804.8396	
AICC (smaller is better)		804.9810	
BIC (smaller is better)		831.2311	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.4352	0.3300	-2.0819	-0.7885	18.92	<.0001
age		1	0.0398	0.0104	0.0195	0.0601	14.78	0.0001
male		1	0.1900	0.1857	-0.1740	0.5540	1.05	0.3064
kids		1	0.1239	0.2104	-0.2884	0.5362	0.35	0.5558
educlevel	BA	1	-0.4723	0.2082	-0.8804	-0.0642	5.15	0.0233
educlevel	MA/PhD	1	-0.4954	0.2262	-0.9388	-0.0521	4.80	0.0285
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Statistics of deviance residual;
proc means data=residual min q1 median mean q3 max maxdec=4;
      var deviance;
run;
```

Analysis Variable : deviance Deviance Residual					
Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
-1.6877	-1.0359	-0.8467	-0.0433	1.2388	1.6452

```
*Build the logistic model and obtain the Person residual;
proc genmod data=edrelig descending;
      class educlevel (ref='AA') / param = ref;
      model religious=age male kids educlevel/dist=binomial link=logit;
      output out=residuals reschi=pearson;
run;
```

Model Information	
Data Set	WORK.EDRELIG
Distribution	Binomial
Link Function	Logit
Dependent Variable	religious

Number of Observations Read	601
Number of Observations Used	601
Number of Events	260
Number of Trials	601

Class Level Information			
Class	**Value**	**Design Variables**	
educlevel	**AA**	0	0
	BA	1	0
	MA/PhD	0	1

Response Profile		
Ordered Value	**religious**	**Total Frequency**
1	1	260
2	0	341

PROC GENMOD is modeling the probability that religious='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-396.4198	
Full Log Likelihood		-396.4198	
AIC (smaller is better)		804.8396	
AICC (smaller is better)		804.9810	
BIC (smaller is better)		831.2311	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.4352	0.3300	-2.0819	-0.7885	18.92	<.0001
age		1	0.0398	0.0104	0.0195	0.0601	14.78	0.0001
male		1	0.1900	0.1857	-0.1740	0.5540	1.05	0.3064
kids		1	0.1239	0.2104	-0.2884	0.5362	0.35	0.5558
educlevel	BA	1	-0.4723	0.2082	-0.8804	-0.0642	5.15	0.0233
educlevel	MA/PhD	1	-0.4954	0.2262	-0.9388	-0.0521	4.80	0.0285
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Pearson Chi2 statistic;
proc sql;
      create table pr as
      select sum(pearson**2) as pchi2, sum(pearson**2)/595 as disp
      from residuals;
quit;

*Print the Chi2 statistic;
proc print data=pr;
run;
```

Obs	pchi2	disp
1	600.179	1.00870

```
*Build the logistic model and obtain statistic;
proc genmod data=edrelig descending;
      class educlevel (ref='AA') / param = ref;
      model religious=age male kids educlevel/dist=binomial link=logit;
      output out=obstats leverage=hat stdreschi=stdp stdresdev=stddev;
run;
```

Model Information	
Data Set	WORK.EDRELIG
Distribution	Binomial
Link Function	Logit
Dependent Variable	religious

Number of Observations Read	601
Number of Observations Used	601
Number of Events	260
Number of Trials	601

Class Level Information			
Class	Value	Design Variables	
educlevel	AA	0	0
	BA	1	0
	MA/PhD	0	1

Response Profile		
Ordered Value	religious	Total Frequency
1	1	260
2	0	341

PROC GENMOD is modeling the probability that religious='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-396.4198	
Full Log Likelihood		-396.4198	
AIC (smaller is better)		804.8396	
AICC (smaller is better)		804.9810	
BIC (smaller is better)		831.2311	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.4352	0.3300	-2.0819	-0.7885	18.92	<.0001
age		1	0.0398	0.0104	0.0195	0.0601	14.78	0.0001
male		1	0.1900	0.1857	-0.1740	0.5540	1.05	0.3064
kids		1	0.1239	0.2104	-0.2884	0.5362	0.35	0.5558
educlevel	BA	1	-0.4723	0.2082	-0.8804	-0.0642	5.15	0.0233
educlevel	MA/PhD	1	-0.4954	0.2262	-0.9388	-0.0521	4.80	0.0285
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Summary for statistic;
proc means data=obstats min q1 median mean q3 max maxdec=6;
      var hat stdp stddev;
run;
```

Variable	Label	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
hat	Leverage	0.006333	0.007721	0.009291	0.009983	0.011224	0.037907
stdp	Standardized Pearson Residual	-1.791322	-0.845671	-0.660307	-0.002086	1.077679	1.704935
stddev	Standardized Deviance Residual	-1.702215	-1.039587	-0.851493	-0.043556	1.242782	1.655601

Section 3.3

```
*Build the logistic model with class;
proc genmod data=medpar descending;
      class type (ref='1') / param = ref;
      model died=white hmo los type / dist=binomial link=logit covb;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Parameter Information		
Parameter	Effect	type
Prm1	Intercept	
Prm2	white	
Prm3	hmo	
Prm4	los	
Prm5	type	2
Prm6	type	3

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-940.5755	
Full Log Likelihood		-940.5755	
AIC (smaller is better)		1893.1509	

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
AICC (smaller is better)		1893.2074	
BIC (smaller is better)		1925.0102	

Algorithm converged.

Estimated Covariance Matrix						
	Prm1	**Prm2**	**Prm3**	**Prm4**	**Prm5**	**Prm6**
Prm1	0.04799	-0.04125	-0.003194	-0.000563	-0.005717	-0.002297
Prm2	-0.04125	0.04373	-0.001326	0.0000612	0.002537	0.0007290
Prm3	-0.003194	-0.001326	0.02287	0.0000272	0.001369	0.003236
Prm4	-0.000563	0.0000612	0.0000272	0.0000608	-0.000122	-0.000374
Prm5	-0.005717	0.002537	0.001369	-0.000122	0.02083	0.005140
Prm6	-0.002297	0.0007290	0.003236	-0.000374	0.005140	0.05263

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7202	0.2191	-1.1495	-0.2908	10.81	0.0010
white		1	0.3037	0.2091	-0.1062	0.7135	2.11	0.1465
hmo		1	0.0272	0.1512	-0.2692	0.3236	0.03	0.8573
los		1	-0.0372	0.0078	-0.0525	-0.0219	22.75	<.0001
type	2	1	0.4179	0.1443	0.1350	0.7007	8.38	0.0038
type	3	1	0.9338	0.2294	0.4842	1.3835	16.57	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Section 3.4

```
*Create a macro variable;
proc sql;
      select disp into: disp
      from pr;
quit;
```

disp
1.008705

```
*Calculations of scaling standard error;
proc iml;
      vcov={0.04799 -0.04125 -0.003194 -0.000563 -0.005717 -0.002297,
           -0.04125  0.04373 -0.001326  0.0000612 0.002537  0.0007290,
           -0.003194 -0.001326 0.02287   0.0000272 0.001369  0.003236,
           -0.000563 0.0000612 0.0000272 0.0000608 -0.000122 -0.000374,
           -0.005717 0.002537  0.001369 -0.000122  0.02083   0.005140,
           -0.002297 0.0007290 0.003236 -0.000374  0.005140  0.05263};
      coef={-0.7202,0.3037,0.0272,-0.0372,0.4179,0.9338};
      se=sqrt(diag(vcov));
      se1=se[,+];
      print coef se1;

      scse=se1*sqrt(&disp);
      print coef se1 scse;
quit;
```

coef	se1
-0.7202	0.2190662
0.3037	0.2091172
0.0272	0.1512283
-0.0372	0.0077974
0.4179	0.144326
0.9338	0.2294123

coef	se1	scse
-0.7202	0.2190662	0.2200176
0.3037	0.2091172	0.2100254
0.0272	0.1512283	0.1518851
-0.0372	0.0077974	0.0078313
0.4179	0.144326	0.1449528
0.9338	0.2294123	0.2304086

```
*Sort the dataset;
proc sort data=medpar;
      by descending type;
run;

*Use quasilikelihood function to generate scaling standard error;
proc glimmix data=medpar order=data;
      class type;
      model died (event='1')=white hmo los type/dist=binary link=logit
      solution;
      random _RESIDUAL_;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Response Variable	died
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	**Levels**	**Values**
type	3	3 2 1

Number of Observations Read	1495
Number of Observations Used	1495

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	0	982
2	1	513
The GLIMMIX procedure is modeling the probability that died='1'.		

Dimensions	
Covariance Parameters	1
Columns in X	7
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	1495

Optimization Information	
Optimization Technique	Newton-Raphson
Parameters in Optimization	6
Lower Boundaries	0
Upper Boundaries	0
Fixed Effects	Not Profiled

Iteration History					
Iteration	**Restarts**	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	**0**	4	942.24502063	.	30.04988
1	**0**	3	940.57613467	1.66888596	0.595459
2	**0**	3	940.57547422	0.00066046	0.000241
3	**0**	3	940.57547422	0.00000000	5.79E-11

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

Fit Statistics	
-2 Log Likelihood	1881.15
AIC (smaller is better)	1893.15
AICC (smaller is better)	1893.21
BIC (smaller is better)	1925.01
CAIC (smaller is better)	1931.01
HQIC (smaller is better)	1905.02
Pearson Chi-Square	1519.45
Pearson Chi-Square / DF	1.02

Parameter Estimates						
Effect	**type**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept		-0.7201	0.2213	1489	-3.25	0.0012
white		0.3037	0.2112	1489	1.44	0.1508
hmo		0.02720	0.1528	1489	0.18	0.8587
los		-0.03719	0.007878	1489	-4.72	<.0001
type	3	0.9338	0.2317	1489	4.03	<.0001
type	2	0.4179	0.1458	1489	2.87	0.0042
type	1	0	.	.	.	.
Residual		1.0205	.	.	.	.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
white	1	1489	2.07	0.1508
hmo	1	1489	0.03	0.8587
los	1	1489	22.29	<.0001
type	2	1489	10.69	<.0001

```
*Generate the robust standard errors;
proc surveylogistic data=medpar;
      class type (ref='1') / param=ref;
      model died (event='1')=white hmo los type;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Response Variable	died
Number of Response Levels	2
Model	Binary Logit
Optimization Technique	Fisher's Scoring
Variance Adjustment	Degrees of Freedom (DF)

Variance Estimation	
Method	Taylor Series
Variance Adjustment	Degrees of Freedom (DF)

Number of Observations Read	1495
Number of Observations Used	1495

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	0	982
2	1	513

Probability modeled is died='1'.

Class Level Information			
Class	**Value**	**Design Variables**	
type	**1**	0	0
	2	1	0
	3	0	1

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

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1924.865	1893.151
SC	1930.175	1925.010
-2 Log L	1922.865	1881.151

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	41.7143	5	<.0001
Score	40.1134	5	<.0001
Wald	32.2851	5	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
white	1	2.0747	0.1498
hmo	1	0.0323	0.8573
los	1	14.5630	0.0001
type	2	21.8681	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7202	0.2199	10.7273	0.0011
white		1	0.3037	0.2108	2.0747	0.1498
hmo		1	0.0272	0.1513	0.0323	0.8573
los		1	-0.0372	0.00975	14.5630	0.0001
type	**2**	1	0.4179	0.1455	8.2443	0.0041
type	**3**	1	0.9338	0.2298	16.5177	<.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
white	1.355	0.896	2.048
hmo	1.028	0.764	1.382
los	0.963	0.945	0.982
type 2 vs 1	1.519	1.142	2.020
type 3 vs 1	2.544	1.622	3.991

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	60.8	**Somers' D**	0.236
Percent Discordant	37.2	**Gamma**	0.240
Percent Tied	1.9	**Tau-a**	0.106
Pairs	503766	**c**	0.618

```
*Generate the bootstrapped standard errors;
%macro bootstrap (Nsamples);
proc surveyselect data=medpar out=boot
      seed=30459584 method=urs samprate=1 rep=&nsamples.;
run;

proc genmod data=boot descending;
      class type (ref='1') / param=ref;
      model died =white hmo los type / dist=binomial link=logit;
      freq numberhits;
      by replicate;
      ods output ParameterEstimates=est;
run;

data est1;
      set est;
      parameter1=parameter;
      if parameter="Scale" then delete;
      if level1=2 then parameter1="type2";
      else if level1=3 then parameter1="type3";
run;

proc means data=est1 mean;
      class parameter1;
      var StdErr;
run;
%mend;
%bootstrap(100);
```

Selection Method	Unrestricted Random Sampling

Input Data Set	MEDPAR
Random Number Seed	30459584
Sampling Rate	1
Sample Size	1495
Expected Number of Hits	1
Sampling Weight	1
Number of Replicates	100

Total Sample Size	149500
Output Data Set	BOOT

Model Information		
Data Set	WORK.BOOT	
Distribution	Binomial	
Link Function	Logit	
Dependent Variable	died	
Frequency Weight Variable	NumberHits	Number of Hits (Selections)

Number of Observations Read	936
Number of Observations Used	936
Sum of Frequencies Read	1495
Sum of Frequencies Used	1495
Number of Events	502
Number of Trials	1495

Class Level Information			
Class	**Value**	**Design Variables**	
type	**1**	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	502
2	0	993

PROC GENMOD is modeling the probability that died='1'.

The output for the 100 sample replication is omitted here.

Analysis Variable : StdErr Standard Error		
parameter1	N Obs	Mean
Intercept	100	0.2216722
hmo	100	0.1520079
los	100	0.0078152
type2	100	0.1455158
type3	100	0.2319434
white	100	0.2117522

Section 3.5

```
*Build the logistic model;
proc genmod data=medpar descending;
      model died =white los / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-951.5059	
Full Log Likelihood		-951.5059	
AIC (smaller is better)		1909.0118	
AICC (smaller is better)		1909.0279	
BIC (smaller is better)		1924.9415	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-0.5987	0.2133	-1.0167	-0.1807	7.88	0.0050
white	1	0.2527	0.2066	-0.1522	0.6575	1.50	0.2212
los	1	-0.0300	0.0077	-0.0451	-0.0149	15.15	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Build the logistic model with interaction;
proc genmod data=medpar descending;
      model died =white los white*los/ dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-948.5448	
Full Log Likelihood		-948.5448	
AIC (smaller is better)		1905.0897	
AICC (smaller is better)		1905.1165	
BIC (smaller is better)		1926.3292	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.0483	0.2802	-1.5976	-0.4991	13.99	0.0002
white	1	0.7709	0.2956	0.1916	1.3503	6.80	0.0091
los	1	0.0100	0.0162	-0.0217	0.0417	0.38	0.5360
white*los	1	-0.0478	0.0183	-0.0836	-0.0119	6.82	0.0090
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Generate odds ratios for los from 1 to 40;
data ior;
      do i=1 to 40;
      or=exp(0.7709+(-0.0478*i));
      output;
      end;
run;

*Print the calculated odds ratios;
proc print data=ior;
run;
```

Obs	i	or
1	1	2.06081
2	2	1.96462
3	3	1.87292
4	4	1.78550
5	5	1.70216
6	6	1.62271
7	7	1.54697
8	8	1.47477
9	9	1.40593
10	10	1.34031
11	11	1.27775
12	12	1.21811
13	13	1.16125
14	14	1.10705
15	15	1.05538
16	16	1.00612
17	17	0.95916
18	18	0.91439
19	19	0.87171
20	20	0.83102
21	21	0.79223
22	22	0.75525
23	23	0.72000
24	24	0.68640
25	25	0.65436

Obs	i	or
26	26	0.62382
27	27	0.59470
28	28	0.56694
29	29	0.54048
30	30	0.51525
31	31	0.49120
32	32	0.46827
33	33	0.44642
34	34	0.42558
35	35	0.40572
36	36	0.38678
37	37	0.36873
38	38	0.35152
39	39	0.33511
40	40	0.31947

Chapter 4 - SAS Code & Output

Section 4.1

```
*Build the logistic model and obtain the Person residual;
proc genmod data=medpar descending;
      class type (ref='1') / param=ref;
      model died=white hmo los type / dist=binomial link=logit;
      output out=residuals reschi=pearson;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	**Value**	**Design Variables**	
type	**1**	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-940.5755	
Full Log Likelihood		-940.5755	
AIC (smaller is better)		1893.1509	
AICC (smaller is better)		1893.2074	
BIC (smaller is better)		1925.0102	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7202	0.2191	-1.1495	-0.2908	10.81	0.0010
white		1	0.3037	0.2091	-0.1062	0.7135	2.11	0.1465
hmo		1	0.0272	0.1512	-0.2692	0.3236	0.03	0.8573
los		1	-0.0372	0.0078	-0.0525	-0.0219	22.75	<.0001
type	2	1	0.4179	0.1443	0.1350	0.7007	8.38	0.0038
type	3	1	0.9338	0.2294	0.4842	1.3835	16.57	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Pearson Chi2 statistic;
proc sql;
      create table pr as
      select sum(pearson**2) as pchi2, 1492 as df,
            1-probchi(sum(pearson**2), 1492) as pvalue
      from residuals;
quit;

*Print the Chi2 statistic;
proc print data=pr;
      title "Pearson Chi GOF";
run;
```

Pearson Chi GOF

Obs	pchi2	df	pvalue
1	1519.45	1492	0.30445

```
*Type3 option provides the likelihood ratio test;
proc genmod data=medpar descending;
      class type (ref='1') / param=ref;
      model died=white hmo los type / dist=binomial link=logit type3;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	**Value**	**Design Variables**	
type	**1**	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-940.5755	
Full Log Likelihood		-940.5755	
AIC (smaller is better)		1893.1509	
AICC (smaller is better)		1893.2074	
BIC (smaller is better)		1925.0102	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7202	0.2191	-1.1495	-0.2908	10.81	0.0010
white		1	0.3037	0.2091	-0.1062	0.7135	2.11	0.1465
hmo		1	0.0272	0.1512	-0.2692	0.3236	0.03	0.8573
los		1	-0.0372	0.0078	-0.0525	-0.0219	22.75	<.0001
type	2	1	0.4179	0.1443	0.1350	0.7007	8.38	0.0038
type	3	1	0.9338	0.2294	0.4842	1.3835	16.57	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
white	1	2.18	0.1400
hmo	1	0.03	0.8574
los	1	26.76	<.0001
type	2	21.77	<.0001

```
*Anscombe residuals can be obtained as a model output in the
SAS/Insight, not in SAS command language;

*Create new variables;
data mylgg;
      set medpar;
      if died=1 then dead=1;
```

```
      else if died=0 then alive=1;
      drop died;
      m=sum(alive, dead);
run;

*Transform the dataset;
proc sql;
      create table mylgg1 as
      select white as white, hmo as hmo, type as type, count(alive) as
      alive, count(dead) as dead, count(m) as m
      from mylgg
      group by white, hmo, type;
quit;

*Print the dataset with selected variables;
proc print data=mylgg1;
      var white hmo type m;
run;

*Print the whole dataset;
proc print data=mylgg1;
run;
```

Obs	white	hmo	type	m
1	0	0	1	72
2	0	0	2	33
3	0	0	3	10
4	0	1	1	8
5	0	1	2	4
6	1	0	1	857
7	1	0	2	201
8	1	0	3	83
9	1	1	1	197
10	1	1	2	27
11	1	1	3	3

Obs	white	hmo	type	alive	dead	m
1	0	0	1	55	17	72
2	0	0	2	22	11	33
3	0	0	3	6	4	10
4	0	1	1	7	1	8
5	0	1	2	1	3	4
6	1	0	1	580	277	857

Obs	white	hmo	type	alive	dead	m
7	1	0	2	119	82	201
8	1	0	3	43	40	83
9	1	1	1	128	69	197
10	1	1	2	19	8	27
11	1	1	3	2	1	3

```
*Obstats option provides all the residuals and statistics in Table 4.2;
proc genmod data=medpar descending;
      class type (ref='1') / param=ref;
      model died=white hmo los type / dist=binomial link=logit obstats;
      ods output obstats=stats;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Parameter Information		
Parameter	Effect	type
Prm1	Intercept	
Prm2	white	
Prm3	hmo	
Prm4	los	
Prm5	type	2
Prm6	type	3

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-940.5755	
Full Log Likelihood		-940.5755	
AIC (smaller is better)		1893.1509	
AICC (smaller is better)		1893.2074	
BIC (smaller is better)		1925.0102	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7202	0.2191	-1.1495	-0.2908	10.81	0.0010
white		1	0.3037	0.2091	-0.1062	0.7135	2.11	0.1465
hmo		1	0.0272	0.1512	-0.2692	0.3236	0.03	0.8573
los		1	-0.0372	0.0078	-0.0525	-0.0219	22.75	<.0001

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
type	2	1	0.4179	0.1443	0.1350	0.7007	8.38	0.0038
type	3	1	0.9338	0.2294	0.4842	1.3835	16.57	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

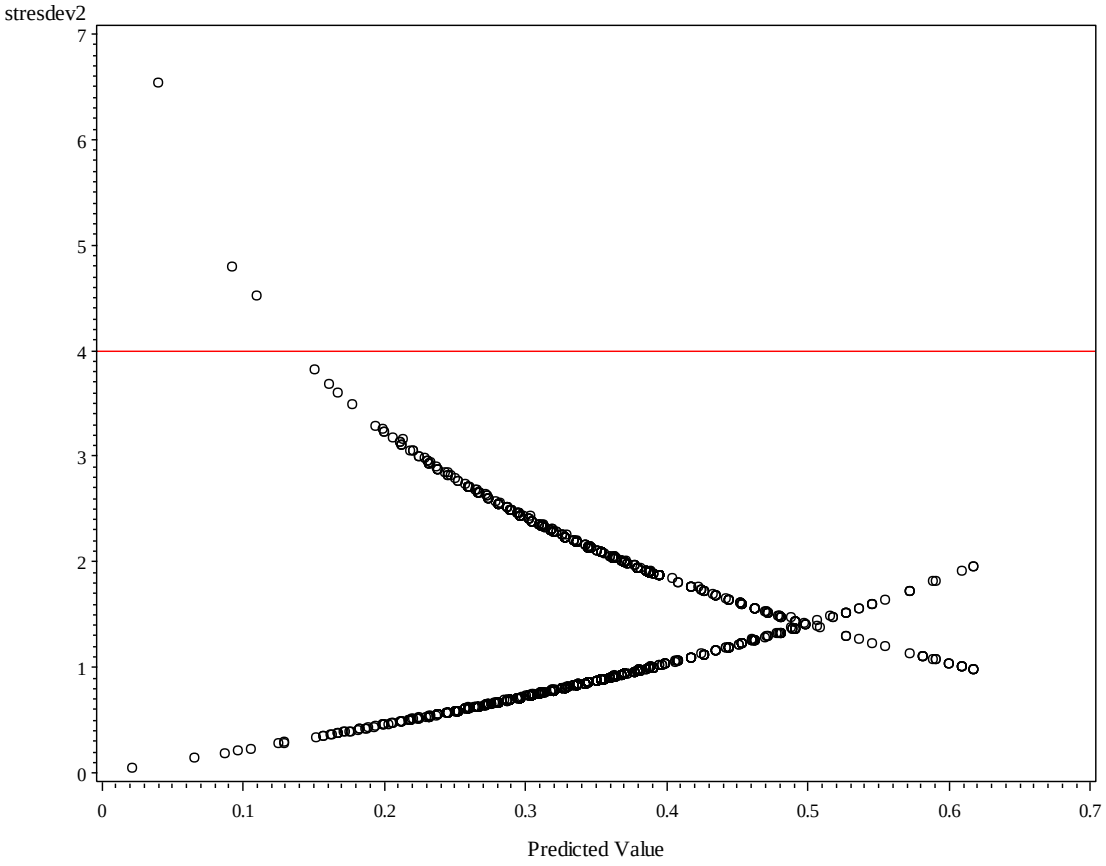
Note: The scale parameter was held fixed.

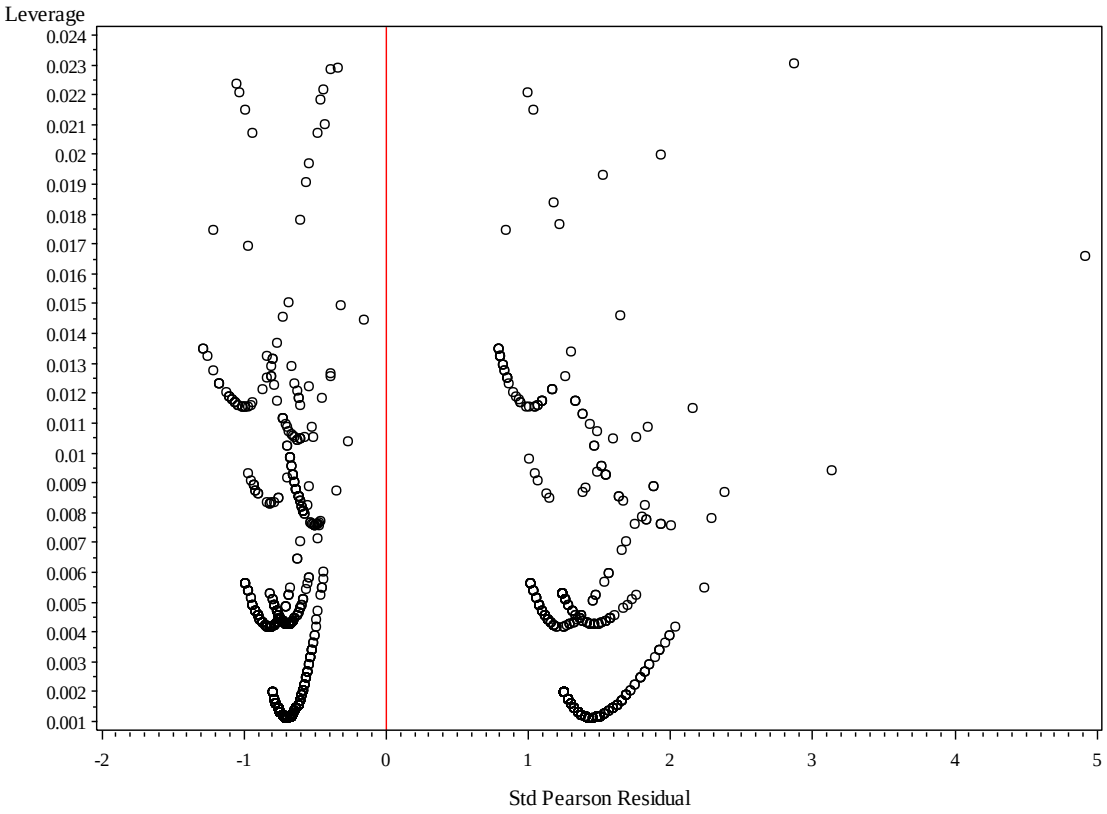
The output for observation statistics is omitted here.

```
*Square the standardized deviance residual;
data stats1;
      set stats;
      stresdev2=stresdev**2;
run;

*Plot the square of standardized deviance residuals and mu;
proc gplot data=stats1;
      symbol v=circle color=black;
      plot stresdev2*pred / vref=4 cvref=red;
run;


*Plot the leverage and std Pearson residual;
proc gplot data=stats1;
      symbol v=circle color=black;
      plot leverage*streschi / href=0 chref=red;
run;
```


stresdev2
7
6
5
4
3
2
1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Predicted Value


Leverage
0.024
0.023
0.022
0.021
0.02
0.019
0.018
0.017
0.016
0.015
0.014
0.013
0.012
0.011
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
-2
-1
0
1
2
3
4
5
Std Pearson Residual

```
*Sort the dataset;
proc sort data=medpar out=medpar1;
      by white hmo los type;
run;

*Calculate the sum of the dead;
proc means data=medpar1 sum;
      by white hmo los type;
      var died;
      output out=summary sum=dead;
run;
```

The output for proc means is omitted here.

```
*Create a new variable alive;
data summary1;
      set summary;
      alive=_freq_-dead;
      drop _type_ _freq_;
run;

*Print the dataset;
proc print data=summary1;
run;
```

The output for dataset summary1 is omitted here.

```
*Build the logistic model with numeric variables;
proc genmod data=medpar descending;
      model died=los type / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-941.7672	
Full Log Likelihood		-941.7672	
AIC (smaller is better)		1889.5345	
AICC (smaller is better)		1889.5506	
BIC (smaller is better)		1905.4641	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-0.8714	0.1414	-1.1485	-0.5944	38.00	<.0001
los	1	-0.0376	0.0078	-0.0529	-0.0224	23.33	<.0001
type	1	0.4386	0.0954	0.2515	0.6256	21.12	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

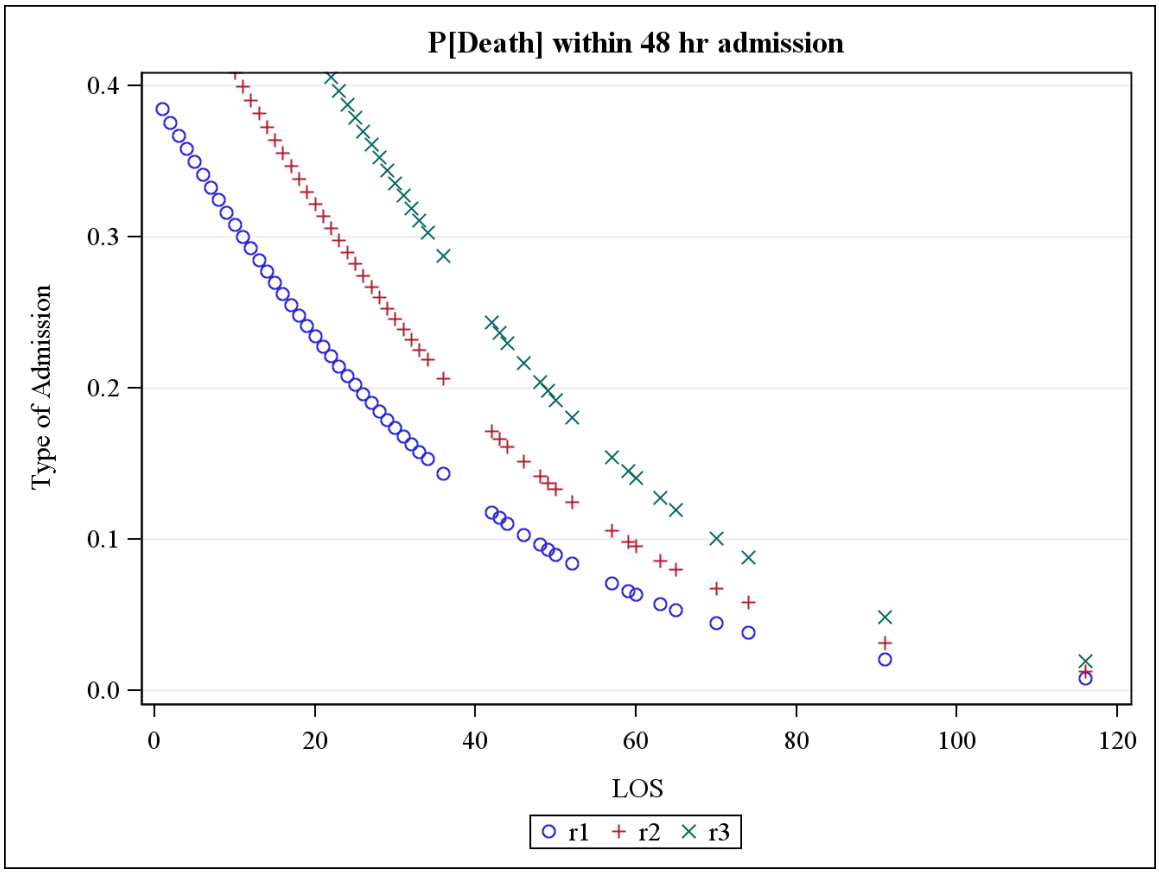
```
*Output the los;
proc freq data=medpar;
      tables los / out=los;
run;
```

los	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	126	8.43	126	8.43
2	71	4.75	197	13.18
3	75	5.02	272	18.19
4	104	6.96	376	25.15
5	123	8.23	499	33.38
6	97	6.49	596	39.87
7	116	7.76	712	47.63
8	92	6.15	804	53.78
9	74	4.95	878	58.73
10	89	5.95	967	64.68
11	70	4.68	1037	69.36
12	70	4.68	1107	74.05
13	43	2.88	1150	76.92
14	49	3.28	1199	80.20
15	41	2.74	1240	82.94
16	43	2.88	1283	85.82
17	29	1.94	1312	87.76
18	23	1.54	1335	89.30
19	24	1.61	1359	90.90
20	19	1.27	1378	92.17
21	18	1.20	1396	93.38
22	15	1.00	1411	94.38
23	10	0.67	1421	95.05
24	11	0.74	1432	95.79
25	4	0.27	1436	96.05
26	7	0.47	1443	96.52
27	7	0.47	1450	96.99
28	5	0.33	1455	97.32
29	3	0.20	1458	97.53
30	1	0.07	1459	97.59
31	2	0.13	1461	97.73
32	6	0.40	1467	98.13

los	Frequency	Percent	Cumulative Frequency	Cumulative Percent
33	2	0.13	1469	98.26
34	5	0.33	1474	98.60
36	1	0.07	1475	98.66
42	1	0.07	1476	98.73
43	1	0.07	1477	98.80
44	2	0.13	1479	98.93
46	3	0.20	1482	99.13
48	1	0.07	1483	99.20
49	1	0.07	1484	99.26
50	1	0.07	1485	99.33
52	1	0.07	1486	99.40
57	1	0.07	1487	99.46
59	1	0.07	1488	99.53
60	1	0.07	1489	99.60
63	1	0.07	1490	99.67
65	1	0.07	1491	99.73
70	1	0.07	1492	99.80
74	1	0.07	1493	99.87
91	1	0.07	1494	99.93
116	1	0.07	1495	100.00

```
*Prepare for the conditional effects plot;
data effect;
      set los;
      k1=-0.8714+(-0.0376)*los+0.4386*1;
      r1=1/(1+exp(-k1));
      k2=-0.8714+(-0.0376)*los+0.4386*2;
      r2=1/(1+exp(-k2));
      k3=-0.8714+(-0.0376)*los+0.4386*3;
      r3=1/(1+exp(-k3));
run;

*Graph the conditional effects plot;
proc sgplot data=effect;
      scatter x=los y=r1;
      scatter x=los y=r2;
      scatter x=los y=r3;
      xaxis label='LOS';
      yaxis label='Type of Admission' grid values=(0 to 0.4 by 0.1);
      title 'P[Death] within 48 hr admission';
run;
```



Section 4.2

```
*Build the logistic model and output model prediction;
proc genmod data=medpar descending;
      class type (ref='1') / param=ref;
      model died=white hmo los type / dist=binomial link=logit;
      output out=fit pred=mu;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	1495
Number of Observations Used	1495
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-940.5755	
Full Log Likelihood		-940.5755	
AIC (smaller is better)		1893.1509	
AICC (smaller is better)		1893.2074	
BIC (smaller is better)		1925.0102	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7202	0.2191	-1.1495	-0.2908	10.81	0.0010
white		1	0.3037	0.2091	-0.1062	0.7135	2.11	0.1465
hmo		1	0.0272	0.1512	-0.2692	0.3236	0.03	0.8573
los		1	-0.0372	0.0078	-0.0525	-0.0219	22.75	<.0001
type	2	1	0.4179	0.1443	0.1350	0.7007	8.38	0.0038

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
type	3	1	0.9338	0.2294	0.4842	1.3835	16.57	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Calculate the mean;
proc means data=fit mean;
      var mu;
run;
```

Analysis Variable : mu Predicted Value
Mean
0.3431439

```
*Build the logistic model and output classification table & ROC curve;
proc logistic data=medpar descending plots(only)=ROC;
      class type (ref='1') / param=ref;
      model died=white hmo los type / outroc=ROCdata ctable pprob=(0 to
1 by 0.0025);
      ods output classification=ctable;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Response Variable	died
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1495
Number of Observations Used	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

Probability modeled is died='1'.

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1924.865	1893.151
SC	1930.175	1925.010
-2 Log L	1922.865	1881.151

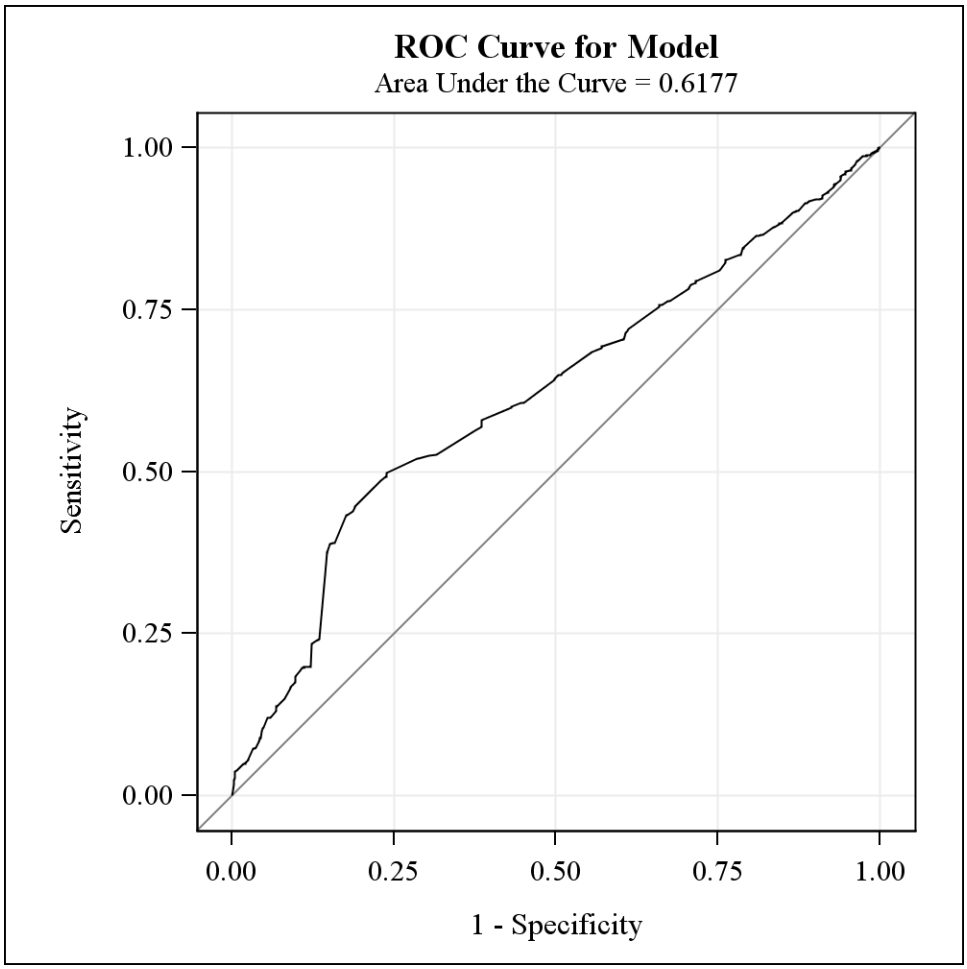
Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.7143	5	<.0001
Score	40.1134	5	<.0001
Wald	38.4987	5	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
white	1	2.1086	0.1465
hmo	1	0.0324	0.8573
los	1	22.7457	<.0001
type	2	21.8192	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7202	0.2191	10.8061	0.0010
white		1	0.3037	0.2091	2.1086	0.1465
hmo		1	0.0272	0.1512	0.0324	0.8573
los		1	-0.0372	0.00780	22.7457	<.0001
type	**2**	1	0.4179	0.1443	8.3840	0.0038
type	**3**	1	0.9338	0.2294	16.5689	<.0001

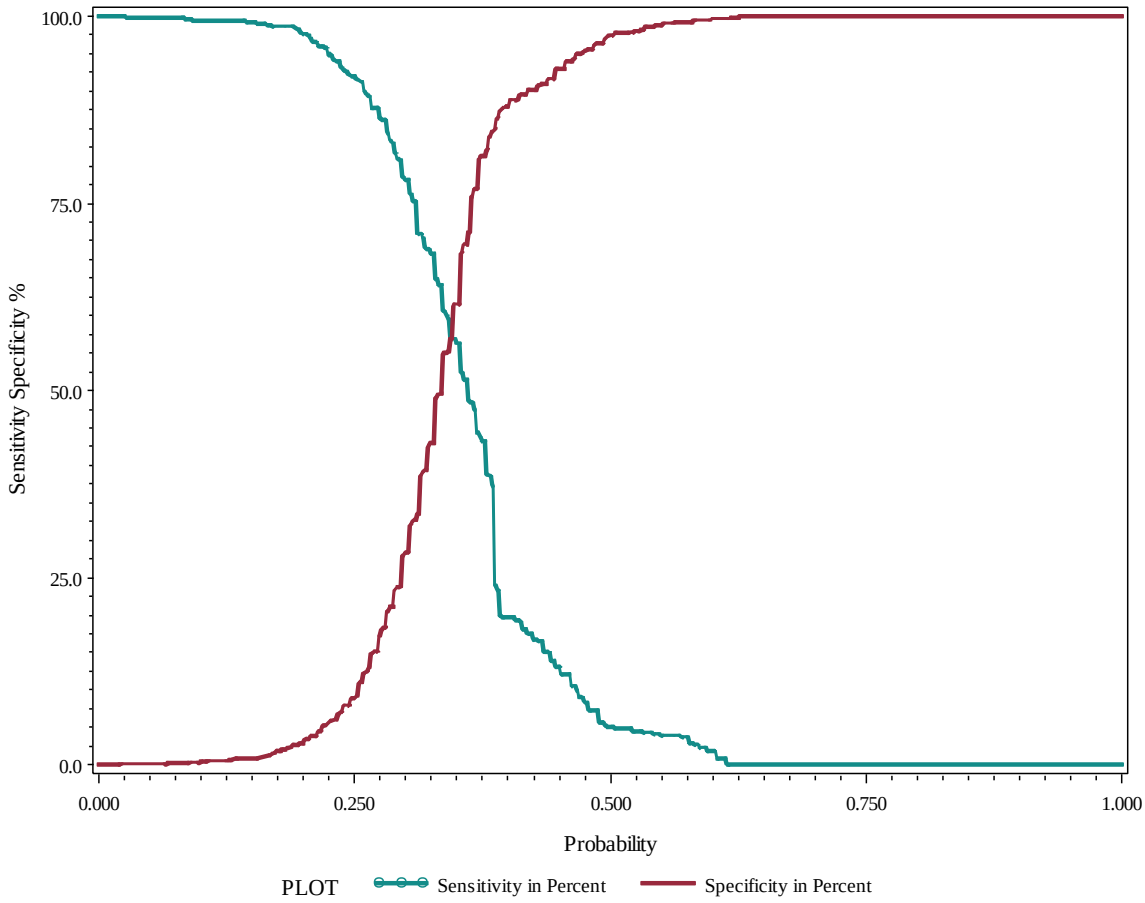
Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
white	1.355	0.899	2.041
hmo	1.028	0.764	1.382
los	0.963	0.949	0.978
type 2 vs 1	1.519	1.145	2.015
type 3 vs 1	2.544	1.623	3.989

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	60.9	**Somers' D**	0.235
Percent Discordant	37.4	**Gamma**	0.240
Percent Tied	1.7	**Tau-a**	0.106
Pairs	503766	**c**	0.618



The output for the classification table is omitted here.

```
*Sensitivity and specificity plot;
symbol1 interpol=join color=vibg height=0.1 width=2;
symbol2 interpol=join color=depk height=0.1 width=2;
axis1 label=("Probability") order=(0 to 1 by 0.25);
axis2 label=(angle=90 "Sensitivity Specificity %") order=(0 to 100 by
25);
proc gplot data=ctable;
      plot sensitivity*problevel specificity*problevel /
      overlay haxis=axis1 vaxis=axis2 legend;
run;
```




```
*Approximate cutoff point can be found when sensitivity and specificity
are closest/equal in the classification table;
```

Section 4.3

```
*Lackfit option provides the Hosmer-Lemeshow GOF test;
proc logistic data=medpar descending;
      class type (ref='1') / param=ref;
      model died=white hmo los type / lackfit;
run;
```

Model Information	
Data Set	WORK.MEDPAR
Response Variable	died
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	1495
Number of Observations Used	1495

Response Profile		
Ordered Value	died	Total Frequency
1	1	513
2	0	982

Probability modeled is died='1'.

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	1924.865	1893.151
SC	1930.175	1925.010
-2 Log L	1922.865	1881.151

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.7143	5	<.0001
Score	40.1134	5	<.0001
Wald	38.4987	5	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
white	1	2.1086	0.1465
hmo	1	0.0324	0.8573
los	1	22.7457	<.0001
type	2	21.8192	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.7202	0.2191	10.8061	0.0010
white		1	0.3037	0.2091	2.1086	0.1465
hmo		1	0.0272	0.1512	0.0324	0.8573
los		1	-0.0372	0.00780	22.7457	<.0001
type	**2**	1	0.4179	0.1443	8.3840	0.0038
type	**3**	1	0.9338	0.2294	16.5689	<.0001

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
white	1.355	0.899	2.041
hmo	1.028	0.764	1.382
los	0.963	0.949	0.978
type 2 vs 1	1.519	1.145	2.015
type 3 vs 1	2.544	1.623	3.989

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	60.8	**Somers' D**	0.236
Percent Discordant	37.2	**Gamma**	0.240
Percent Tied	1.9	**Tau-a**	0.106
Pairs	503766	**c**	0.618

Partition for the Hosmer and Lemeshow Test					
		died = 1		died = 0	
Group	Total	Observed	Expected	Observed	Expected
1	151	43	31.92	108	119.08
2	148	42	40.31	106	107.69
3	156	40	46.66	116	109.34
4	143	37	45.31	106	97.69
5	148	40	49.29	108	98.71
6	169	41	59.06	128	109.94
7	165	41	60.17	124	104.83
8	158	105	60.70	53	97.30
9	155	70	66.04	85	88.96
10	102	54	53.53	48	48.47

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
82.2059	8	<.0001

Section 4.4

```
*Import hivlgold as a temporary dataset;
proc import datafile="c:\data\hivlgold.dta" out=HIV dbms=dta replace;
run;

*Build the weighted logistic model;
proc genmod data=HIV descending;
      class cd4 (ref='0') cd8 (ref='0') / param = ref;
      weight cases;
      model infec= cd4 cd8 / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.HIV
Distribution	Binomial
Link Function	Logit
Dependent Variable	infec
Scale Weight Variable	cases

Number of Observations Read	12
Number of Observations Used	12
Sum of Weights	47
Number of Events	6
Number of Trials	12

Class Level Information			
Class	**Value**	**Design Variables**	
cd4	**0**	0	0
	1	1	0
	2	0	1
cd8	**0**	0	0
	1	1	0
	2	0	1

Response Profile			
Ordered Value	**infec**	**Total Frequency**	**Total Weight**
1	1	6	14
2	0	6	33

PROC GENMOD is modeling the probability that infec='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-18.5160	
Full Log Likelihood		-18.5160	
AIC (smaller is better)		47.0319	
AICC (smaller is better)		57.0319	
BIC (smaller is better)		49.4564	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2877	0.7638	-1.2093	1.7846	0.14	0.7064
cd4	1	1	-1.2040	1.1328	-3.4243	1.0164	1.13	0.2879
cd4	2	1	-28.1155	1.4434	-30.9445	-25.2866	379.43	<.0001
cd8	1	1	0.2231	1.0368	-1.8090	2.2553	0.05	0.8296
cd8	2	0	27.1347	0.0000	27.1347	27.1347	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Build the Firth's penalized logistic model;
proc logistic data=HIV descending;
      class cd4 (ref='0') cd8 (ref='0') / param = ref;
      weight cases;
      model infec= cd4 cd8 / firth clodds=pl;
run;
```

Model Information	
Data Set	WORK.HIV
Response Variable	infec
Number of Response Levels	2
Weight Variable	cases
Model	binary logit
Optimization Technique	Fisher's scoring
Likelihood Penalty	Firth's bias correction

Number of Observations Read	12
Number of Observations Used	12
Sum of Weights Read	47
Sum of Weights Used	47

Response Profile			
Ordered Value	infec	Total Frequency	Total Weight
1	1	6	14.000000
2	0	6	33.000000

Probability modeled is infec=1.

Class Level Information			
Class	Value	Design Variables	
cd4	0	0	0
	1	1	0
	2	0	1
cd8	0	0	0
	1	1	0
	2	0	1

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

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	55.428	47.003
SC	55.913	49.428
-2 Log L	53.428	37.003

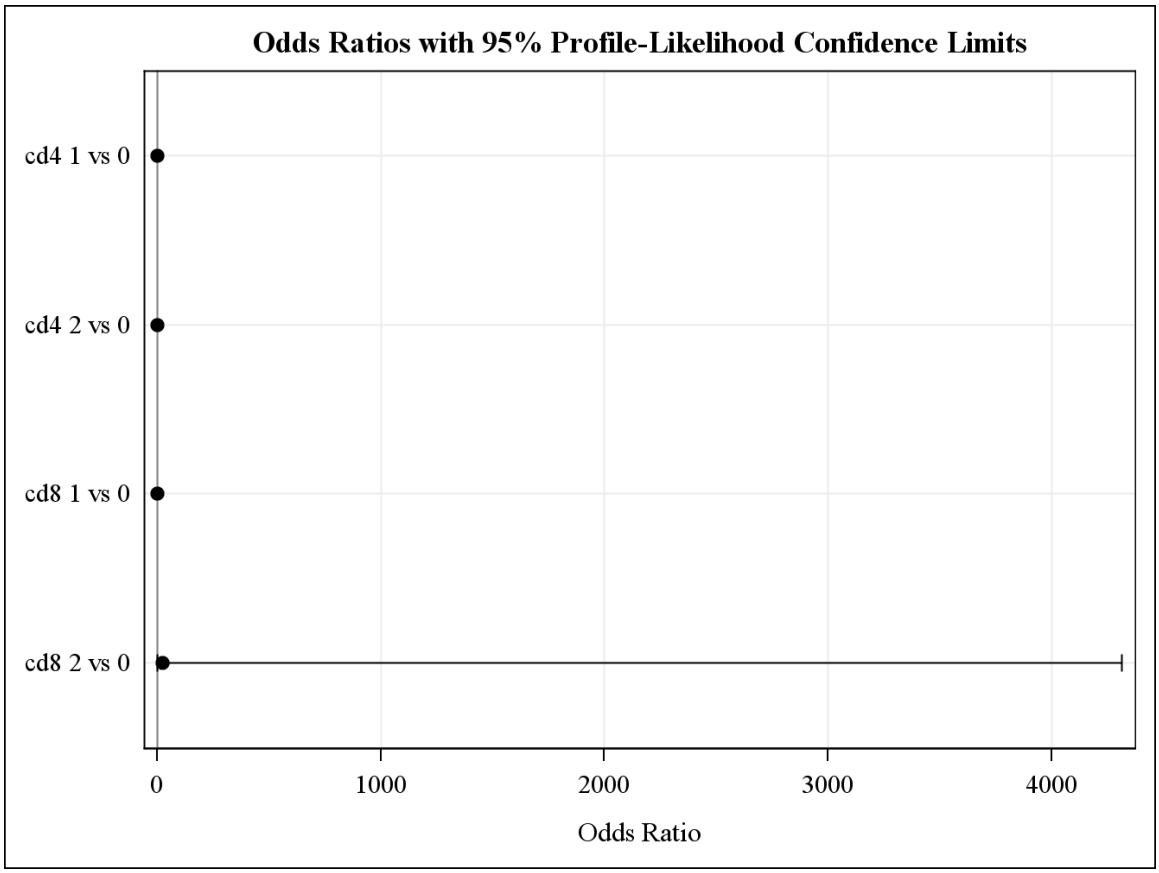
Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	16.4253	4	0.0025
Score	15.6517	4	0.0035
Wald	6.5245	4	0.1633

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
cd4	2	5.4072	0.0670
cd8	2	4.0456	0.1323

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	0.2431	0.7557	0.1035	0.7476
cd4	**1**	1	-1.0207	1.0904	0.8762	0.3492
cd4	**2**	1	-4.0138	1.7546	5.2329	0.0222
cd8	**1**	1	0.1320	0.9853	0.0179	0.8934
cd8	**2**	1	3.2265	1.7200	3.5190	0.0607

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	72.2	**Somers' D**	0.556
Percent Discordant	16.7	**Gamma**	0.625
Percent Tied	11.1	**Tau-a**	0.303
Pairs	36	**c**	0.778

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals				
Effect	**Unit**	**Estimate**	**95% Confidence Limits**	
cd4 1 vs 0	1.0000	0.360	0.040	2.644
cd4 2 vs 0	1.0000	0.018	<0.001	0.323
cd8 1 vs 0	1.0000	1.141	0.178	8.515
cd8 2 vs 0	1.0000	25.191	1.591	>999.999



Section 4.5

```
*Import azheart as a temporary dataset;
proc import datafile="c:\data\azheart.dta" out=azheart dbms=dta
replace;
run;

*Print the first six observations;
proc print data=azheart (obs=6);
run;
```

Obs	died	procedure	age	gender	los	type
1	1	1	65	1	10	0
2	0	1	69	1	7	1
3	0	0	76	0	7	1
4	0	1	65	1	8	0
5	0	0	69	1	1	0
6	0	1	67	1	7	1

```
*Generate a table of died by procedure and type;
proc freq data=azheart;
      tables died*procedure died*type / norow nocol nocum nopercent;
```

```
run;
```

Table of died by procedure			
died	procedure		
Frequency	0	1	Total
0	19	9	28
1	1	5	6
Total	20	14	34

Table of died by type			
died	type		
Frequency	0	1	Total
0	17	11	28
1	4	2	6
Total	21	13	34

```
*Build the logistic model and obtain odds ratio & statistics;
proc genmod data=azheart descending;
      model died=procedure type / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "Procedure" procedure 1 / exp;
      estimate "Type" type 1 / exp;
run;
```

Model Information	
Data Set	WORK.AZHEART
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	34
Number of Observations Used	34
Number of Events	6
Number of Trials	34

Response Profile		
Ordered Value	died	Total Frequency
1	1	6
2	0	28

PROC GENMOD is modeling the probability that died='1'.

Parameter Information	
Parameter	Effect
Prm1	Intercept
Prm2	procedure
Prm3	type

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-12.9785	
Full Log Likelihood		-12.9785	
AIC (smaller is better)		31.9570	
AICC (smaller is better)		32.7570	
BIC (smaller is better)		36.5361	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.2479	1.2411	-5.6804	-0.8153	6.85	0.0089
procedure	1	2.5380	1.2483	0.0915	4.9846	4.13	0.0420
type	1	0.5415	1.1228	-1.6592	2.7422	0.23	0.6296
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	0.0374	0.0034	0.3068	-3.2479	1.2411	0.05	-5.6804	-0.8153	6.85	0.0089
Exp(Intercept)				0.0389	0.0482	0.05	0.0034	0.4425		
Procedure	0.9268	0.5229	0.9932	2.5380	1.2483	0.05	0.0915	4.9846	4.13	0.0420
Exp(Procedure)				12.6548	15.7966	0.05	1.0958	146.1437		
Type	0.6322	0.1599	0.9395	0.5415	1.1228	0.05	-1.6592	2.7422	0.23	0.6296
Exp(Type)				1.7186	1.9296	0.05	0.1903	15.5208		

```
*Build the quasibinomial logistic model;
proc glimmix data=azheart;
      model died (event='1')=procedure type/dist=binary link=logit
      solution covb;
      random _RESIDUAL_;
run;
```

Model Information	
Data Set	WORK.AZHEART
Response Variable	died
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Number of Observations Read	34
Number of Observations Used	34

Response Profile		
Ordered Value	**died**	**Total Frequency**
1	0	28
2	1	6
The GLIMMIX procedure is modeling the probability that died='1'.		

Dimensions	
Covariance Parameters	1
Columns in X	3
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	34

Optimization Information	
Optimization Technique	Newton-Raphson
Parameters in Optimization	3
Lower Boundaries	0
Upper Boundaries	0
Fixed Effects	Not Profiled

Iteration History					
Iteration	**Restarts**	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	**0**	4	13.373927713	.	0.697824
1	**0**	3	13.000939828	0.37298788	0.219255
2	**0**	3	12.978667219	0.02227261	0.017184
3	**0**	3	12.978510793	0.00015643	0.000136
4	**0**	3	12.978510783	0.00000001	8.801E-9

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

Fit Statistics	
-2 Log Likelihood	25.96
AIC (smaller is better)	31.96
AICC (smaller is better)	32.76
BIC (smaller is better)	36.54
CAIC (smaller is better)	39.54
HQIC (smaller is better)	33.52
Pearson Chi-Square	30.41
Pearson Chi-Square / DF	0.98

Parameter Estimates					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	-3.2479	1.2294	31	-2.64	0.0128
procedure	2.5380	1.2364	31	2.05	0.0486
type	0.5415	1.1122	31	0.49	0.6298
Residual	0.9811	.	.	.	.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
procedure	1	31	4.21	0.0486
type	1	31	0.24	0.6298

Covariance Matrix for Fixed Effects				
Effect	**Row**	**Col1**	**Col2**	**Col3**
Intercept	**1**	1.5113	-1.3315	-0.7666
procedure	**2**	-1.3315	1.5288	0.4765
type	**3**	-0.7666	0.4765	1.2369

```
*Calculations of odds ratio and model statistics;
proc iml;
      vcov={1.5113 -1.3315 -0.7666,
           -1.3315  1.5288  0.4765,
           -0.7666  0.4765  1.2369};
      coef={-3.2479,2.5380,0.5415};
      se=sqrt(diag(vcov));
      se1=se[,+];
```

```
      zscore=coef/se;
      z=zscore[,+];
      or=exp(coef);
      delta=se*or;
      pvalue=2*(1-probnorm((abs(z))));
      loci=coef-quantile('normal', 0.975)*se1;
      upci=coef+quantile('normal', 0.975)*se1;
      expl=exp(loci);
      expu=exp(upci);
      print or [format=7.4] delta [format=7.4] z [format=7.4]
            pvalue [format=7.4] expl [format=7.4] expu [format=7.4];
quit;
```

or	delta	z	pvalue	expl	expu
0.0389	0.0478	-2.6420	0.0082	0.0035	0.4324
12.6543	15.6464	2.0527	0.0401	1.1214	142.792
1.7186	1.9113	0.4869	0.6263	0.1943	15.2000

```
*Build the exact logistic model;
proc genmod data=azheart descending;
      model died=procedure type / dist=binomial link=logit;
      exact procedure type / estimate=both;
run;
```

Model Information	
Data Set	WORK.AZHEART
Distribution	Binomial
Link Function	Logit
Dependent Variable	died

Number of Observations Read	34
Number of Observations Used	34
Number of Events	6
Number of Trials	34

Response Profile		
Ordered Value	died	Total Frequency
1	1	6
2	0	28

PROC GENMOD is modeling the probability that died='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-12.9785	
Full Log Likelihood		-12.9785	
AIC (smaller is better)		31.9570	
AICC (smaller is better)		32.7570	
BIC (smaller is better)		36.5361	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	-3.2479	1.2411	-5.6804	-0.8153	6.85	0.0089
procedure	1	2.5380	1.2483	0.0915	4.9846	4.13	0.0420
type	1	0.5415	1.1228	-1.6592	2.7422	0.23	0.6296
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Conditional Exact Tests				
			p-Value	
Effect	**Test**	**Statistic**	**Exact**	**Mid**
procedure	**Score**	5.0867	0.0542	0.0383
	Probability	0.0318	0.0542	0.0383
type	**Score**	0.2205	1.0000	0.8146
	Probability	0.3709	1.0000	0.8146

Exact Parameter Estimates					
Parameter	Estimate	Standard Error	95% Confidence Limits		Two-sided p-Value
procedure	2.3357	1.1779	-0.1188	6.3798	0.0679
type	0.5048	1.0826	-2.2967	3.3459	1.0000

Exact Odds Ratios				
Parameter	Estimate	95% Confidence Limits		Two-sided p-Value
procedure	10.336	0.888	589.811	0.0679
type	1.657	0.101	28.387	1.0000

Section 4.6

```
*Create a temporary dataset;
data mydata;
      input y x count @@;
      datalines;
1 1 8 1 0 6 0 1 5 0 0 4
;
run;

*Build the logistic model with weight and obtain odds ratio;
proc genmod data=mydata descending;
      weight count;
      model y=x / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "x" x 1 / exp;
run;
```

Model Information	
Data Set	WORK.MYDATA
Distribution	Binomial
Link Function	Logit
Dependent Variable	y
Scale Weight Variable	count

Number of Observations Read	4
Number of Observations Used	4
Sum of Weights	23

Number of Events	2
Number of Trials	4

Response Profile			
Ordered Value	**y**	**Total Frequency**	**Total Weight**
1	1	2	14
2	0	2	9

PROC GENMOD is modeling the probability that y='1'.

Parameter Information	
Parameter	**Effect**
Prm1	Intercept
Prm2	x

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-15.3917	
Full Log Likelihood		-15.3917	
AIC (smaller is better)		34.7835	
AICC (smaller is better)		46.7835	
BIC (smaller is better)		33.5561	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	0.4055	0.6455	-0.8597	1.6706	0.39	0.5299
x	1	0.0645	0.8612	-1.6234	1.7525	0.01	0.9403
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	0.6000	0.2974	0.8417	0.4055	0.6455	0.05	-0.8597	1.6706	0.39	0.5299
Exp(Intercept)				1.5000	0.9682	0.05	0.4233	5.3154		
x	0.5161	0.1647	0.8523	0.0645	0.8612	0.05	-1.6234	1.7525	0.01	0.9403
Exp(x)				1.0667	0.9186	0.05	0.1972	5.7688		

```
*Create a temporary dataset;
data mydata2;
      input grade gender type count @@;
      datalines;
0 0 1 3 0 0 2 4 0 0 3 2 0 1 1 2
0 1 2 4 0 1 3 3 1 0 1 2 1 0 2 1
1 0 3 6 1 1 1 3 1 1 2 2 1 1 3 4
;
run;

*Build the logistic model with weight and obtain odds ratio;
proc genmod data=mydata2 descending;
      class type (ref='1') / param=ref;
      weight count;
      model grade=gender type / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "Gender" gender 1 / exp;
      estimate "Type2" type 1 0 / exp;
      estimate "Type3" type 0 1 / exp;
run;
```

Model Information	
Data Set	WORK.MYDATA2
Distribution	Binomial
Link Function	Logit
Dependent Variable	grade
Scale Weight Variable	count

Number of Observations Read	12
Number of Observations Used	12
Sum of Weights	36
Number of Events	6
Number of Trials	12

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

Response Profile			
Ordered Value	grade	Total Frequency	Total Weight
1	1	6	18
2	0	6	18

PROC GENMOD is modeling the probability that grade='1'.

Parameter Information		
Parameter	Effect	type
Prm1	Intercept	
Prm2	gender	
Prm3	type	2
Prm4	type	3

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-22.9149	
Full Log Likelihood		-22.9149	
AIC (smaller is better)		53.8299	

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
AICC (smaller is better)		59.5442	
BIC (smaller is better)		55.7695	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.0494	0.7252	-1.4707	1.3719	0.00	0.9457
gender		1	0.0988	0.7089	-1.2906	1.4882	0.02	0.8891
type	2	1	-0.9859	0.9276	-2.8039	0.8322	1.13	0.2879
type	3	1	0.6968	0.8374	-0.9445	2.3382	0.69	0.4053
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
Label	**Mean Estimate**	**Mean Confidence Limits**		**L'Beta Estimate**	**Standard Error**	**Alpha**	**L'Beta Confidence Limits**		**Chi-Square**	**Pr > ChiSq**
Intercept	0.4876	0.1868	0.7977	-0.0494	0.7252	0.05	-1.4707	1.3719	0.00	0.9457
Exp(Intercept)				0.9518	0.6902	0.05	0.2298	3.9428		
Gender	0.5247	0.2158	0.8158	0.0988	0.7089	0.05	-1.2906	1.4882	0.02	0.8891
Exp(Gender)				1.1039	0.7825	0.05	0.2751	4.4292		
Type2	0.2717	0.0571	0.6968	-0.9859	0.9276	0.05	-2.8039	0.8322	1.13	0.2879
Exp(Type2)				0.3731	0.3461	0.05	0.0606	2.2983		
Type3	0.6675	0.2800	0.9120	0.6968	0.8374	0.05	-0.9445	2.3382	0.69	0.4053
Exp(Type3)				2.0074	1.6811	0.05	0.3889	10.3622		

Chapter 5 - SAS Code & Output

Section 5.2

```
*Create a temporary dataset;
data obser;
      input y x1 x2 x3 @@;
      datalines;
1 1 0 1 0 1 0 1 0 1 0 1 1 1 1 1 1 0 0 1
1 0 0 1 0 0 1 1 1 1 0 0 1 1 0 0 0 0 1 0
;
run;

*Build the logistic model;
proc genmod data=obser descending;
      model y=x1 x2 x3 / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.OBSER
Distribution	Binomial
Link Function	Logit
Dependent Variable	y

Number of Observations Read	10
Number of Observations Used	10
Number of Events	6
Number of Trials	10

Response Profile		
Ordered Value	**y**	**Total Frequency**
1	1	6
2	0	4

PROC GENMOD is modeling the probability that y='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Log Likelihood		-6.0251	
Full Log Likelihood		-6.0251	
AIC (smaller is better)		20.0501	
AICC (smaller is better)		28.0501	
BIC (smaller is better)		21.2604	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept	1	1.2050	1.8348	-2.3911	4.8011	0.43	0.5113
x1	1	0.1714	1.4909	-2.7507	3.0934	0.01	0.9085
x2	1	-1.5972	1.6011	-4.7352	1.5408	1.00	0.3185
x3	1	-0.5499	1.5817	-3.6499	2.5502	0.12	0.7281
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Create a temporary dataset;
data grp;
      input y cases x1 x2 x3 @@;
      datalines;
1 3 1 0 1 1 1 1 1 1
2 2 0 0 1 0 1 0 1 1
2 2 1 0 0 0 1 0 1 0
;
run;

*Build the logistic model;
proc genmod data=grp descending;
      model y/cases=x1 x2 x3 / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.GRP
Distribution	Binomial
Link Function	Logit

Model Information	
Response Variable (Events)	y
Response Variable (Trials)	cases

Number of Observations Read	6
Number of Observations Used	6
Number of Events	6
Number of Trials	10

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	6
2	Nonevent	4

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	2	8.2310	4.1155
Scaled Deviance	2	8.2310	4.1155
Pearson Chi-Square	2	6.6300	3.3150
Scaled Pearson X2	2	6.6300	3.3150
Log Likelihood		-6.0251	
Full Log Likelihood		-4.9264	
AIC (smaller is better)		17.8529	
AICC (smaller is better)		57.8529	
BIC (smaller is better)		17.0199	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.2050	1.8348	-2.3911	4.8011	0.43	0.5113
x1	1	0.1714	1.4909	-2.7507	3.0934	0.01	0.9085
x2	1	-1.5972	1.6011	-4.7352	1.5408	1.00	0.3185
x3	1	-0.5499	1.5817	-3.6499	2.5502	0.12	0.7281
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

```
*Build the logistic model;
proc genmod data=grp descending;
      class x1 (ref='0') x2 (ref='0') x3 (ref='0') / param=ref;
      model y/cases=x1 x2 x3 / dist=binomial link=logit;
run;
```

Model Information	
Data Set	WORK.GRP
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	y
Response Variable (Trials)	cases

Number of Observations Read	6
Number of Observations Used	6
Number of Events	6
Number of Trials	10

Class Level Information		
Class	Value	Design Variables
x1	0	0
	1	1
x2	0	0
	1	1

Class Level Information		
Class	**Value**	**Design Variables**
x3	**0**	0
	1	1

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	6
2	Nonevent	4

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	2	8.2310	4.1155
Scaled Deviance	2	8.2310	4.1155
Pearson Chi-Square	2	6.6300	3.3150
Scaled Pearson X2	2	6.6300	3.3150
Log Likelihood		-6.0251	
Full Log Likelihood		-4.9264	
AIC (smaller is better)		17.8529	
AICC (smaller is better)		57.8529	
BIC (smaller is better)		17.0199	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.2050	1.8348	-2.3911	4.8011	0.43	0.5113
x1	1	1	0.1714	1.4909	-2.7507	3.0934	0.01	0.9085
x2	1	1	-1.5972	1.6011	-4.7352	1.5408	1.00	0.3185
x3	1	1	-0.5499	1.5817	-3.6499	2.5502	0.12	0.7281
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Section 5.4

```
*Sort the dataset;
proc sort data=medpar out=medpar1;
      by white hmo los type;
run;

*Calculate the sum of the dead;
proc means data=medpar1 sum;
      by white hmo los type;
      var died;
      output out=summary sum=dead;
run;
```

The output for proc means is omitted here.

```
*Create a new variable alive;
data summary1;
      set summary;
      alive=_freq_-dead;
      cases=_freq_;
      drop _type_ _freq_;
run;

*Obstats option provides all the residuals and useful statistics;
proc genmod data=summary1 descending;
      class type (ref='1')/ param=ref;
      model dead/cases=white hmo los type / dist=binomial link=logit
      obstats;
      ods output obstats=allstats;
run;
```

Model Information	
Data Set	WORK.SUMMARY1
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	dead
Response Variable (Trials)	cases

Number of Observations Read	210
Number of Observations Used	210
Number of Events	513
Number of Trials	1495

Class Level Information			
Class	Value	Design Variables	
type	1	0	0
	2	1	0
	3	0	1

Response Profile		
Ordered Value	Binary Outcome	Total Frequency
1	Event	513
2	Nonevent	982

Parameter Information		
Parameter	Effect	type
Prm1	Intercept	
Prm2	white	
Prm3	hmo	
Prm4	los	
Prm5	type	2
Prm6	type	3

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	204	399.4412	1.9580
Scaled Deviance	204	399.4412	1.9580
Pearson Chi-Square	204	386.4857	1.8945
Scaled Pearson X2	204	386.4857	1.8945
Log Likelihood		-940.5755	
Full Log Likelihood		-314.6998	
AIC (smaller is better)		641.3995	
AICC (smaller is better)		641.8133	
BIC (smaller is better)		661.4822	

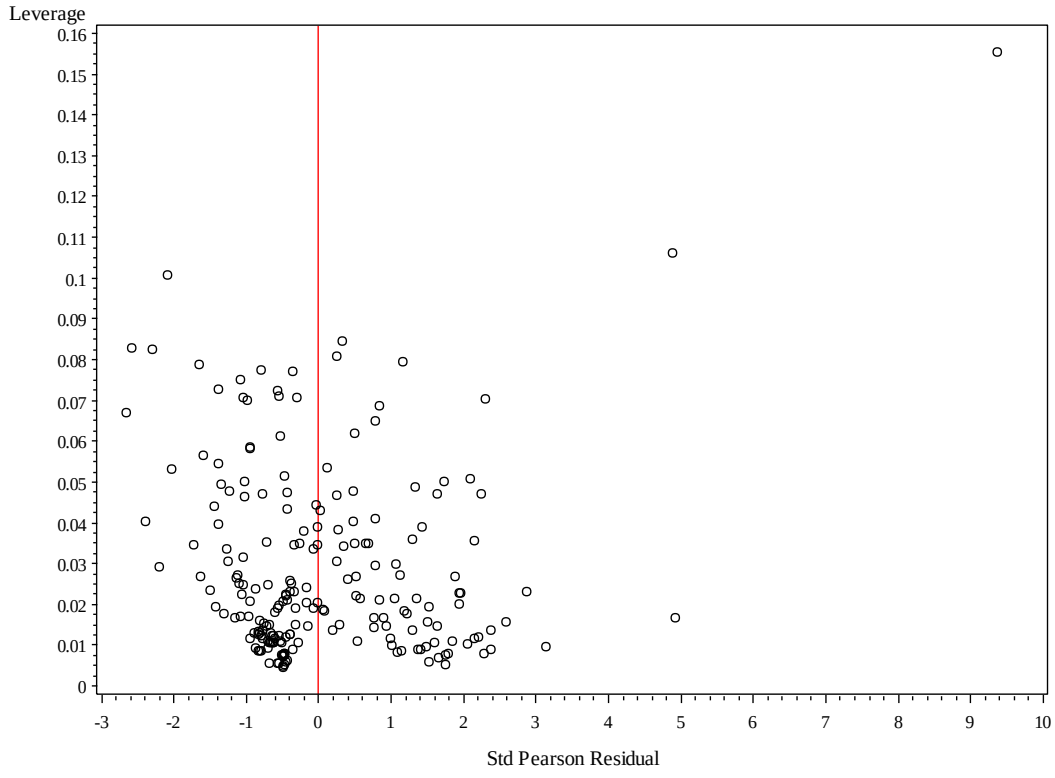
Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.7201	0.2191	-1.1495	-0.2908	10.81	0.0010
white		1	0.3037	0.2091	-0.1062	0.7135	2.11	0.1465
hmo		1	0.0272	0.1512	-0.2692	0.3236	0.03	0.8573
los		1	-0.0372	0.0078	-0.0525	-0.0219	22.75	<.0001
type	2	1	0.4179	0.1443	0.1350	0.7007	8.38	0.0038
type	3	1	0.9338	0.2294	0.4842	1.3835	16.57	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

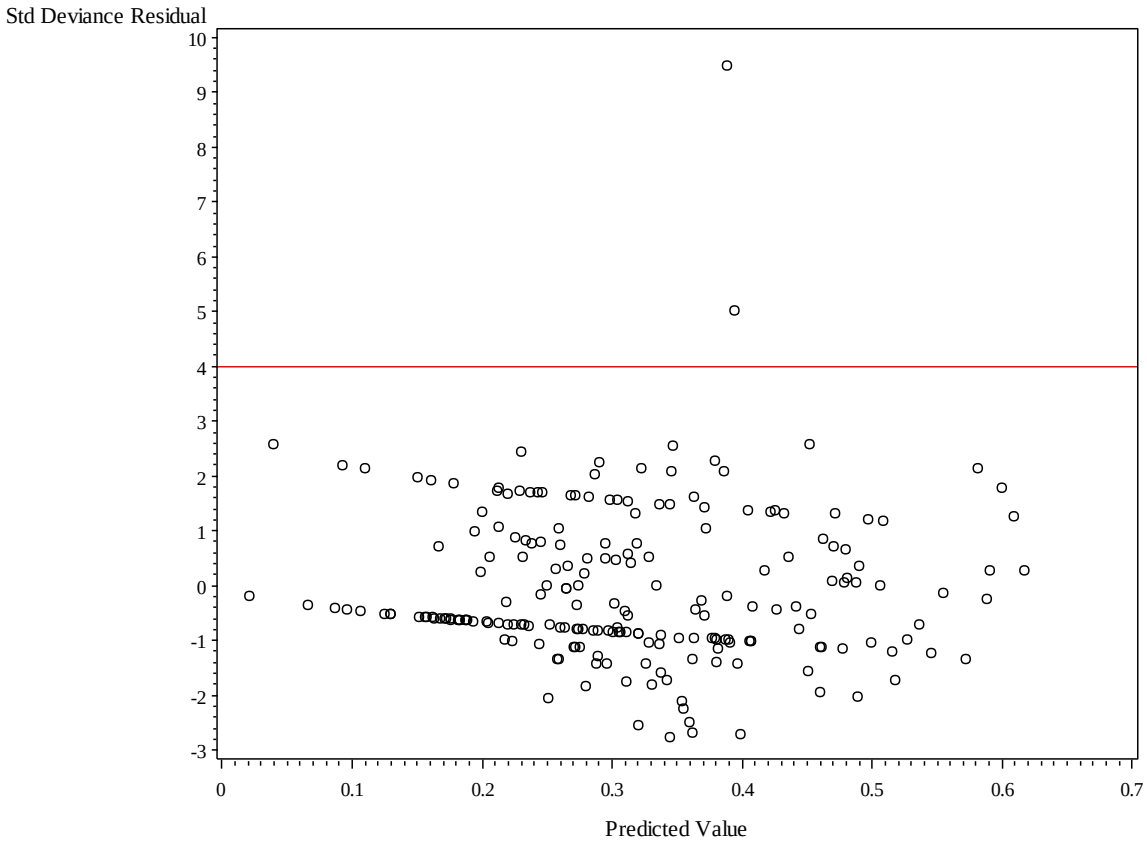
Note: The scale parameter was held fixed.

The output for observation statistics is omitted here.

```
*Plot the leverage and std Pearson residual;
proc gplot data=allstats;
      symbol v=circle color=black;
      plot leverage*streschi / href=0 chref=red;
run;
```



```
*Plot the standardized deviance residuals and mu;
proc gplot data=allstats;
      symbol v=circle color=black;
      plot stresdev*pred / vref=4 cvref=red;
run;
```



Section 5.5

```
*Import titanicgrp as a temporary dataset;
proc import datafile="c:\data\titanicgrp.dta" out=azheart dbms=dta
replace;
run;

*Print the dataset;
proc print data=titanicgrp;
run;
```

Obs	survive	cases	age	sex	class
1	1	1	0	0	1
2	13	13	0	0	2
3	14	31	0	0	3
4	5	5	0	1	1
5	11	11	0	1	2
6	13	48	0	1	3
7	140	144	1	0	1
8	80	93	1	0	2
9	76	165	1	0	3
10	57	175	1	1	1
11	14	168	1	1	2
12	75	462	1	1	3

```
*Build the logistic model and obtain odds ratio & covariance matrix;
proc genmod data=titanicgrp descending;
      class class (ref='3')/ param=ref;
      model survive/cases=age sex class/dist=binomial link=logit covb;
      estimate "Intercept" Intercept 1 / exp;
      estimate "ageadults" age 1 / exp;
      estimate "sexman" sex 1 / exp;
      estimate "class" class 1 0 / exp;
      estimate "class" class 0 1 / exp;
run;
```

Model Information	
Data Set	WORK.TITANICGRP
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	survive
Response Variable (Trials)	cases

Number of Observations Read	12
Number of Observations Used	12
Number of Events	499
Number of Trials	1316

Class Level Information			
Class	**Value**	**Design Variables**	
class	**1**	1	0
	2	0	1
	3	0	0

Response Profile		
Ordered Value	**Binary Outcome**	**Total Frequency**
1	Event	499
2	Nonevent	817

Parameter Information		
Parameter	**Effect**	**class**
Prm1	Intercept	
Prm2	age	
Prm3	sex	
Prm4	class	1
Prm5	class	2

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	7	110.8438	15.8348
Scaled Deviance	7	110.8438	15.8348
Pearson Chi-Square	7	100.8828	14.4118
Scaled Pearson X2	7	100.8828	14.4118
Log Likelihood		-638.1004	

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Full Log Likelihood		-73.8837	
AIC (smaller is better)		157.7673	
AICC (smaller is better)		167.7673	
BIC (smaller is better)		160.1918	

Algorithm converged.

Estimated Covariance Matrix					
	Prm1	**Prm2**	**Prm3**	**Prm4**	**Prm5**
Prm1	0.06143	-0.05164	-0.01337	-0.000882	-0.005907
Prm2	-0.05164	0.05888	0.002418	-0.007878	-0.003941
Prm3	-0.01337	0.002418	0.02110	-0.004657	-0.001562
Prm4	-0.000882	-0.007878	-0.004657	0.02914	0.01093
Prm5	-0.005907	-0.003941	-0.001562	0.01093	0.03071

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.2955	0.2478	0.8097	1.7813	27.32	<.0001
age		1	-1.0556	0.2427	-1.5312	-0.5800	18.92	<.0001
sex		1	-2.3695	0.1453	-2.6542	-2.0848	266.11	<.0001
class	1	1	1.7664	0.1707	1.4318	2.1010	107.06	<.0001
class	2	1	0.7558	0.1753	0.4123	1.0993	18.60	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
		Mean					L'Beta			
Label	Mean Estimate	Confidence Limits		L'Beta Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	0.7851	0.6921	0.8559	1.2955	0.2478	0.05	0.8097	1.7813	27.32	<.0001
Exp(Intercept)				3.6529	0.9053	0.05	2.2473	5.9374		
ageadults	0.2581	0.1778	0.3589	-1.0556	0.2427	0.05	-1.5312	-0.5800	18.92	<.0001
Exp(ageadults)				0.3480	0.0844	0.05	0.2163	0.5599		
sexman	0.0855	0.0657	0.1106	-2.3695	0.1453	0.05	-2.6542	-2.0848	266.11	<.0001
Exp(sexman)				0.0935	0.0136	0.05	0.0704	0.1243		
class	0.8540	0.8072	0.8910	1.7664	0.1707	0.05	1.4318	2.1010	107.06	<.0001
Exp(class)				5.8496	0.9986	0.05	4.1861	8.1741		
class	0.6804	0.6016	0.7501	0.7558	0.1753	0.05	0.4123	1.0993	18.60	<.0001
Exp(class)				2.1293	0.3732	0.05	1.5103	3.0021		

```
*Build the logistic mode with robust adjustment;
proc glimmix data=titanicgrp order=data empirical=hc0;
      class class;
      model survive/cases=age sex class / dist=binomial link=logit
      solution covb;
      random _RESIDUAL_;
run;
```

Model Information	
Data Set	WORK.TITANICGRP
Response Variable (Events)	survive
Response Variable (Trials)	cases
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual
Fixed Effects SE Adjustment	Sandwich - Classical

Class Level Information		
Class	**Levels**	**Values**
class	3	1 2 3

Number of Observations Read	12
Number of Observations Used	12
Number of Events	499
Number of Trials	1316

Dimensions	
Covariance Parameters	1
Columns in X	6
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	12

Optimization Information	
Optimization Technique	Newton-Raphson
Parameters in Optimization	5
Lower Boundaries	0
Upper Boundaries	0
Fixed Effects	Not Profiled

Iteration History					
Iteration	**Restarts**	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	**0**	4	78.356446026	.	12.7278
1	**0**	3	73.913673628	4.44277240	0.640774
2	**0**	3	73.883654007	0.03001962	0.005531
3	**0**	3	73.88365169	0.00000232	4.505E-7

Convergence criterion (ABSGCONV=0.00001) satisfied.

Fit Statistics	
-2 Log Likelihood	147.77
AIC (smaller is better)	157.77
AICC (smaller is better)	167.77
BIC (smaller is better)	160.19
CAIC (smaller is better)	165.19
HQIC (smaller is better)	156.87
Pearson Chi-Square	100.88
Pearson Chi-Square / DF	14.41

Parameter Estimates						
Effect	**class**	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept		1.2955	0.7614	7	1.70	0.1326
age		-1.0556	0.7134	7	-1.48	0.1825
sex		-2.3695	0.4719	7	-5.02	0.0015
class	1	1.7664	0.5221	7	3.38	0.0117
class	2	0.7558	0.5936	7	1.27	0.2436
class	3	0	.	.	.	.
Residual		14.4118	.	.	.	.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
age	1	7	2.19	0.1825
sex	1	7	25.21	0.0015
class	2	7	5.72	0.0336

Empirical Covariance Matrix for Fixed Effects								
Effect	**class**	**Row**	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**	**Col6**
Intercept		**1**	0.5798	-0.4722	-0.2001	0.04322	-0.05299	
age		**2**	-0.4722	0.5089	0.1235	-0.1162	-0.06026	
sex		**3**	-0.2001	0.1235	0.2227	-0.1112	-0.01121	
class	1	**4**	0.04322	-0.1162	-0.1112	0.2726	0.1219	

Empirical Covariance Matrix for Fixed Effects								
Effect	**class**	**Row**	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**	**Col6**
class	2	**5**	-0.05299	-0.06026	-0.01121	0.1219	0.3524	
class	3	**6**						

```
*Calculations of odds ratio and model statistics;
proc iml;
      vcov={0.5798 -0.4722 -0.2001 0.04322 -0.05299,
            -0.4722 0.5089 0.1235 -0.1162 -0.06026,
            -0.2001 0.1235 0.2227 -0.1112 -0.01121,
             0.04322 -0.1162 -0.1112 0.2726 0.1219,
            -0.05299 -0.06026 -0.01121 0.1219 0.3524};
      coef={1.2955, -1.0556, -2.3695, 1.7664, 0.7558};
      or=exp(coef);
      rse=sqrt(diag(vcov));
      rse1=rse[,+];
      ORrse=rse*or;
      pvalue=2*(1-probnorm((abs(1/rse1))));
      print or ORrse pvalue;
quit;
```

or	ORrse	pvalue
3.6528219	2.7814266	0.189085
0.3479836	0.2482418	0.1609772
0.0935275	0.0441366	0.0340868
5.8497563	3.0542227	0.0554541
2.1293143	1.264031	0.0920768

```
*Build the Beta binomial model;
proc fmm data=titanicgrp;
      class class;
      model survive/cases=age sex class / dist=betabinomial;
run;
```

Model Information	
Data Set	WORK.TITANICGRP
Response Variable (Events)	survive
Response Variable (Trials)	cases
Type of Model	Homogeneous Regression Mixture
Distribution	Beta-Binomial
Components	1
Link Function	Logit
Estimation Method	Maximum Likelihood

Class Level Information		
Class	**Levels**	**Values**
class	3	1 2 3

Number of Observations Read	12
Number of Observations Used	12
Number of Events	499
Number of Trials	1316

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	6
Mean Function Parameters	5
Scale Parameters	1
Lower Boundaries	1
Upper Boundaries	0
Number of Threads	4

Iteration History				
Iteration	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	5	59.639708268	.	15.98042
1	3	56.55987088	3.07983739	12.00596
2	2	55.379659115	1.18021176	11.03613
3	3	54.711536509	0.66812261	9.571647
4	3	54.561339723	0.15019679	7.160595
5	2	54.444150323	0.11718940	5.634089
6	4	53.83789298	0.60625734	1.824941
7	3	53.799578603	0.03831438	0.170995
8	3	53.796768005	0.00281060	0.134166
9	4	53.749941889	0.04682612	1.880966
10	2	53.689565786	0.06037610	0.992955
11	35	40.071078839	13.61848695	6.175032

Iteration History				
Iteration	Evaluations	Objective Function	Change	Max Gradient
12	5	39.930456938	0.14062190	6.434034
13	4	37.853599118	2.07685782	2.733451
14	5	37.364066476	0.48953264	1.559344
15	3	37.157870588	0.20619589	1.107901
16	3	37.03546346	0.12240713	0.98377
17	2	36.93574982	0.09971364	0.489551
18	3	36.911989561	0.02376026	0.398571
19	3	36.906597248	0.00539231	0.272466
20	2	36.902206064	0.00439118	0.100485
21	3	36.901218847	0.00098722	0.018134
22	3	36.901200284	0.00001856	0.015308
23	2	36.901182797	0.00001749	0.005728
24	3	36.901181327	0.00000147	0.003207
25	2	36.901180592	0.00000073	0.000713

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

Fit Statistics	
-2 Log Likelihood	73.8024
AIC (smaller is better)	85.8024
AICC (smaller is better)	102.6
BIC (smaller is better)	88.7118
Pearson Statistic	12.0534

Parameter Estimates for 'Beta-Binomial' Model					
Effect	class	Estimate	Standard Error	z Value	Pr > |z|
Intercept		1.5051	0.8304	1.81	0.0699
age		-2.2130	0.7890	-2.80	0.0050
sex		-2.1859	0.7611	-2.87	0.0041
class	1	2.7694	0.8764	3.16	0.0016
class	2	2.0292	0.7879	2.58	0.0100

Parameter Estimates for 'Beta-Binomial' Model					
Effect	**class**	**Estimate**	**Standard Error**	**z Value**	**Pr > |z|**
class	3	0	.	.	.
Scale Parameter		5.9955	3.6409		

Chapter 6 - SAS Code & Output

Section 6.2

```
*Import rwm1984 as a temporary dataset;
proc import datafile="c:\data\rwm1984.dta" out=rwm1984 dbms=dta
replace;
run;

*Print the first 6 observations;
proc print data=rwm1984 (obs=6);
run;
```

Obs	docvis	hospvis	edlevel	age	outwork	female	married	kids	hhninc	educ	self
1	1	0	3	54	0	0	1	0	3.049999952	15	0
2	0	0	1	44	1	1	1	0	3.049999952	9	0
3	0	0	1	58	1	1	0	0	1.434000015	11	0
4	7	2	1	64	0	0	0	0	1.5	10.5	0
5	6	0	3	30	1	0	0	0	2.400000095	13	0
6	9	0	3	26	1	0	0	0	1.049999952	13	0

Obs	edlevel1	edlevel2	edlevel3	edlevel4
1	0	0	1	0
2	1	0	0	0
3	1	0	0	0
4	1	0	0	0
5	0	0	1	0
6	0	0	1	0

```
*Generate the frequency table of outwork ;
proc freq data=rwm1984;
      tables outwork / norow nocol nocum nopercent;
run;
```

outwork	Frequency
0	2454
1	1420

```
*Summary for continuous variables;
proc means data=rwm1984 min q1 median mean q3 max maxdec=3;
      var docvis age;
      output out=center mean=;
run;
```

Variable	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum
docvis	0.000	0.000	1.000	3.163	4.000	121.000
age	25.000	35.000	44.000	43.996	54.000	64.000

```
*Create the macro variables;
proc sql;
      select age into: meanage from center;
      select docvis into: meandoc from center;
quit;
```

age
43.995869902

docvis
3.1628807434

```
*Center the continuous variables;
data R84;
      set rwm1984;
      cage=age-&meanage;
      cdoc=docvis-&meandoc;
run;

*Build the logistic model and obtain odds ratio & statistics;
proc genmod data=R84 descending;
      model outwork=cdoc female kids cage / dist=binomial link=logit;
      estimate "Intercept" Intercept 1 / exp;
      estimate "Cdoc" cdoc 1 / exp;
      estimate "Female" female 1 / exp;
      estimate "Kids" kids 1 / exp;
      estimate "Cage" cage 1 / exp;
run;
```

Model Information	
Data Set	WORK.R84
Distribution	Binomial
Link Function	Logit
Dependent Variable	outwork

Number of Observations Read	3874
Number of Observations Used	3874
Number of Events	1420
Number of Trials	3874

Response Profile		
Ordered Value	outwork	Total Frequency
1	1	1420
2	0	2454

PROC GENMOD is modeling the probability that outwork='1'.

Parameter Information	
Parameter	Effect
Prm1	Intercept
Prm2	cdoc
Prm3	female
Prm4	kids
Prm5	cage

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Log Likelihood		-1959.1076	
Full Log Likelihood		-1959.1076	
AIC (smaller is better)		3928.2151	
AICC (smaller is better)		3928.2306	
BIC (smaller is better)		3959.5253	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.0103	0.0811	-2.1692	-1.8513	614.58	<.0001
cdoc	1	0.0244	0.0063	0.0122	0.0367	15.22	<.0001
female	1	2.2568	0.0828	2.0946	2.4190	743.57	<.0001

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
kids	1	0.3580	0.0900	0.1817	0.5343	15.83	<.0001
cage	1	0.0544	0.0042	0.0462	0.0625	170.95	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

Contrast Estimate Results										
Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	0.1181	0.1025	0.1357	-2.0103	0.0811	0.05	-2.1692	-1.8513	614.58	<.0001
Exp(Intercept)				0.1340	0.0109	0.05	0.1143	0.1570		
Cdoc	0.5061	0.5030	0.5092	0.0244	0.0063	0.05	0.0122	0.0367	15.22	<.0001
Exp(Cdoc)				1.0247	0.0064	0.05	1.0122	1.0374		
Female	0.9052	0.8904	0.9183	2.2568	0.0828	0.05	2.0946	2.4190	743.57	<.0001
Exp(Female)				9.5525	0.7906	0.05	8.1221	11.2348		
Kids	0.5886	0.5453	0.6305	0.3580	0.0900	0.05	0.1817	0.5343	15.83	<.0001
Exp(Kids)				1.4304	0.1287	0.05	1.1992	1.7063		
Cage	0.5136	0.5116	0.5156	0.0544	0.0042	0.05	0.0462	0.0625	170.95	<.0001
Exp(Cage)				1.0559	0.0044	0.05	1.0473	1.0645		

```
*Build the quasibinomial logistic model;
proc glimmix data=R84;
      model outwork (event='1')=cdoc female kids cage / dist=binary
      link=logit solution covb;
      random _RESIDUAL_;
run;
```

Model Information	
Data Set	WORK.R84
Response Variable	outwork
Response Distribution	Binary
Link Function	Logit
Variance Function	Default

Model Information	
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Number of Observations Read	3874
Number of Observations Used	3874

Response Profile		
Ordered Value	**outwork**	**Total Frequency**
1	0	2454
2	1	1420
The GLIMMIX procedure is modeling the probability that outwork='1'.		

Dimensions	
Covariance Parameters	1
Columns in X	5
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	3874

Optimization Information	
Optimization Technique	Newton-Raphson
Parameters in Optimization	5
Lower Boundaries	0
Upper Boundaries	0
Fixed Effects	Not Profiled

Iteration History					
Iteration	**Restarts**	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	**0**	4	1966.345961	.	60.23161
1	**0**	3	1959.1436053	7.20235570	2.872108
2	**0**	3	1959.107567	0.03603825	0.017694
3	**0**	3	1959.1075656	0.00000149	7.493E-7

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

Fit Statistics	
-2 Log Likelihood	3918.22
AIC (smaller is better)	3928.22
AICC (smaller is better)	3928.23
BIC (smaller is better)	3959.53
CAIC (smaller is better)	3964.53
HQIC (smaller is better)	3939.33
Pearson Chi-Square	4205.00
Pearson Chi-Square / DF	1.09

Parameter Estimates					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	-2.0103	0.08454	3869	-23.78	<.0001
cdoc	0.02443	0.006529	3869	3.74	0.0002
female	2.2568	0.08628	3869	26.16	<.0001
kids	0.3580	0.09379	3869	3.82	0.0001
cage	0.05438	0.004336	3869	12.54	<.0001
Residual	1.0868	.	.	.	.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
cdoc	1	3869	14.00	0.0002
female	1	3869	684.15	<.0001
kids	1	3869	14.57	0.0001
cage	1	3869	157.29	<.0001

Covariance Matrix for Fixed Effects						
Effect	**Row**	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**
Intercept	**1**	0.007147	-6.01E-6	-0.00505	-0.00461	-0.00015
cdoc	**2**	-6.01E-6	0.000043	-0.00002	0.000022	-2.26E-6
female	**3**	-0.00505	-0.00002	0.007444	0.000595	0.000062
kids	**4**	-0.00461	0.000022	0.000595	0.008796	0.000192
cage	**5**	-0.00015	-2.26E-6	0.000062	0.000192	0.000019

```
*Calculations of odds ratio and model statistics;
proc iml;
      vcov={0.007147 -0.00000601 -0.00505 -0.00461 -0.00015,
            -0.00000601 0.000043 -0.00002 0.000022 -0.00000226,
            -0.00505 -0.00002 0.007444 0.000595 0.000062,
            -0.00461 0.000022 0.000595 0.008796 0.000192,
            -0.00015 -0.00000226 0.000062 0.000192 0.000019};
      coef={-2.0103,0.02443,2.2568,0.3580,0.05438};
      se=sqrt(diag(vcov));
      se1=se[,+];
      zscore=coef/se;
      z=zscore[,+];
      or=exp(coef);
      delta=se*or;
      pvalue=2*(1-probnorm((abs(z))));
      loci=coef-quantile('normal', 0.975)*se1;
      upci=coef+quantile('normal', 0.975)*se1;
      expl=exp(loci);
      expu=exp(upci);
      print or [format=7.4] delta [format=7.4] z [format=7.4]
            pvalue [format=7.4] expl [format=7.4] expu [format=7.4];
quit;
```

or	**delta**	**z**	**pvalue**	**expl**	**expu**
0.1339	0.0113	-23.779	0.0000	0.1135	0.1581
1.0247	0.0067	3.7255	0.0002	1.0116	1.0380
9.5525	0.8242	26.1571	0.0000	8.0663	11.3124

or	delta	z	pvalue	expl	expu
1.4305	0.1342	3.8172	0.0001	1.1903	1.7191
1.0559	0.0046	12.4756	0.0000	1.0469	1.0649

```
*Bayesian logistic regression;
proc genmod data=R84 descending;
      model outwork=cdoc female kids cage / dist=binomial link=logit;
      bayes seed=10231995 nbi=5000 nmc=100000
      coeffprior=uniform diagnostics=all
      statistics=(summary interval) plots=all;
run;
```

Model Information	
Data Set	WORK.R84
Burn-In Size	5000
MC Sample Size	100000
Thinning	1
Sampling Algorithm	ARMS
Distribution	Binomial
Link Function	Logit
Dependent Variable	outwork

Number of Observations Read	3874
Number of Observations Used	3874
Number of Events	1420
Number of Trials	3874

Response Profile		
Ordered Value	**outwork**	**Total Frequency**
1	1	1420
2	0	2454

PROC GENMOD is modeling the probability that outwork='1'.

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**	
Intercept	1	-2.0103	0.0811	-2.1692	-1.8513
cdoc	1	0.0244	0.0063	0.0122	0.0367
female	1	2.2568	0.0828	2.0946	2.4190
kids	1	0.3580	0.0900	0.1817	0.5343
cage	1	0.0544	0.0042	0.0462	0.0625
Scale	0	1.0000	0.0000	1.0000	1.0000

Note: The scale parameter was held fixed.

Uniform Prior for Regression Coefficients	
Parameter	**Prior**
Intercept	Constant
cdoc	Constant
female	Constant
kids	Constant
cage	Constant

Algorithm converged.

Initial Values of the Chain						
Chain	**Seed**	**Intercept**	**cdoc**	**female**	**kids**	**cage**
1	10231995	-2.01028	0.024432	2.256804	0.357976	0.054379
2		-1.8957	0.005644	2.008517	0.088084	0.041902
3		-2.12485	0.043221	2.505091	0.627868	0.066856

Fit Statistics	
DIC (smaller is better)	3928.284
pD (effective number of parameters)	5.032

Posterior Summaries						
				Percentiles		
Parameter	N	Mean	Standard Deviation	25%	50%	75%
Intercept	100000	-2.0140	0.0815	-2.0686	-2.0134	-1.9586
cdoc	100000	0.0247	0.00632	0.0204	0.0246	0.0289
female	100000	2.2605	0.0832	2.2043	2.2602	2.3166
kids	100000	0.3596	0.0907	0.2981	0.3590	0.4207
cage	100000	0.0545	0.00418	0.0516	0.0545	0.0573

Posterior Intervals					
		Equal-Tail Interval		HPD Interval	
Parameter	Alpha				
Intercept	0.050	-2.1755	-1.8557	-2.1710	-1.8520
cdoc	0.050	0.0124	0.0373	0.0124	0.0373
female	0.050	2.0989	2.4242	2.0971	2.4220
kids	0.050	0.1831	0.5382	0.1838	0.5386
cage	0.050	0.0463	0.0628	0.0464	0.0628

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
Intercept	0.2937	0.0021	0.0045	-0.0009
cdoc	0.0065	-0.0024	-0.0004	-0.0029
female	0.1193	-0.0024	-0.0034	-0.0013
kids	0.2308	0.0038	0.0004	0.0002
cage	0.2597	0.0059	0.0014	0.0029

Gelman-Rubin Diagnostics		
Parameter	Estimate	97.5% Bound
Intercept	1.0000	1.0000
cdoc	1.0000	1.0000
female	1.0000	1.0001
kids	1.0000	1.0000
cage	1.0000	1.0000

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Intercept	-1.4697	0.1417
cdoc	-1.1356	0.2561
female	0.1919	0.8478
kids	1.0302	0.3029
cage	2.3240	0.0201

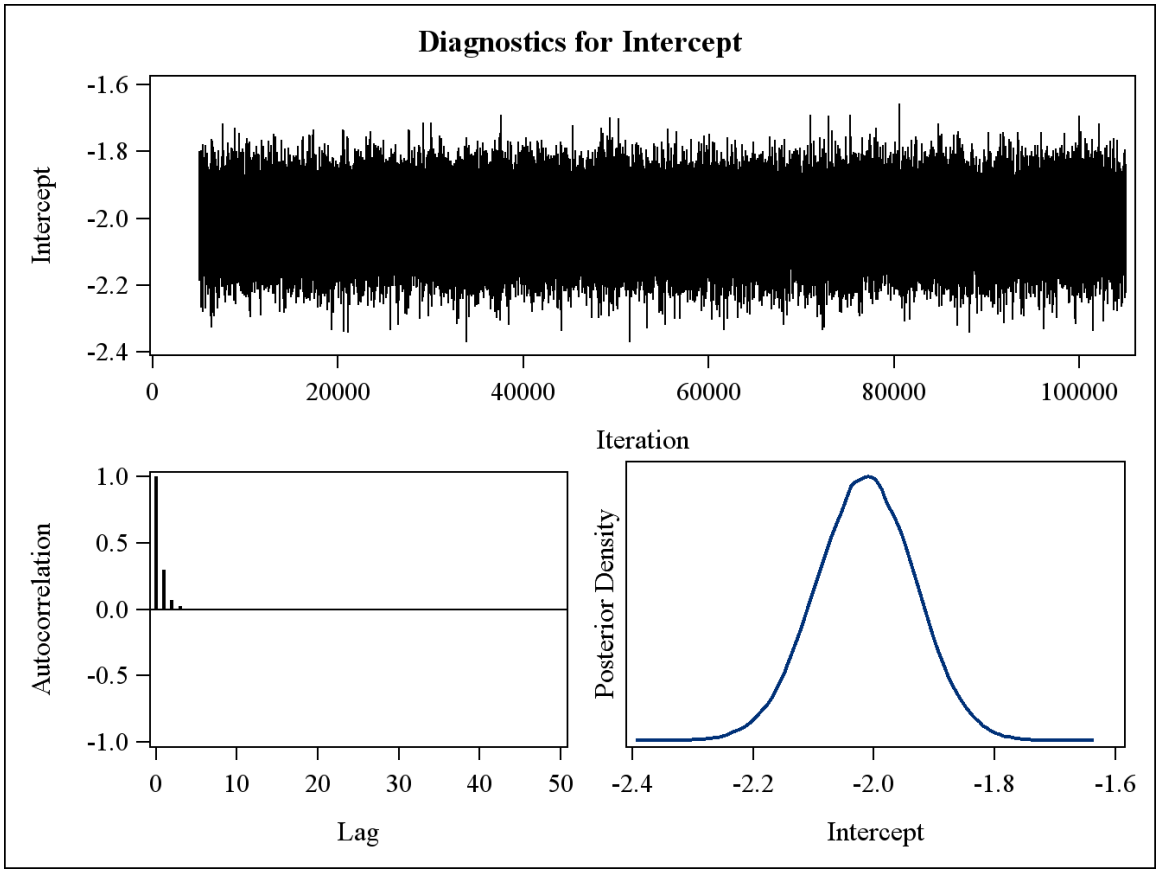
Raftery-Lewis Diagnostics				
Quantile=0.025 Accuracy=+/-0.005 Probability=0.95 Epsilon=0.001				
	Number of Samples			Dependence Factor
Parameter	**Burn-in**	**Total**	**Minimum**	
Intercept	3	4409	3746	1.1770
cdoc	2	3771	3746	1.0067
female	2	3865	3746	1.0318
kids	3	4095	3746	1.0932
cage	3	4140	3746	1.1052

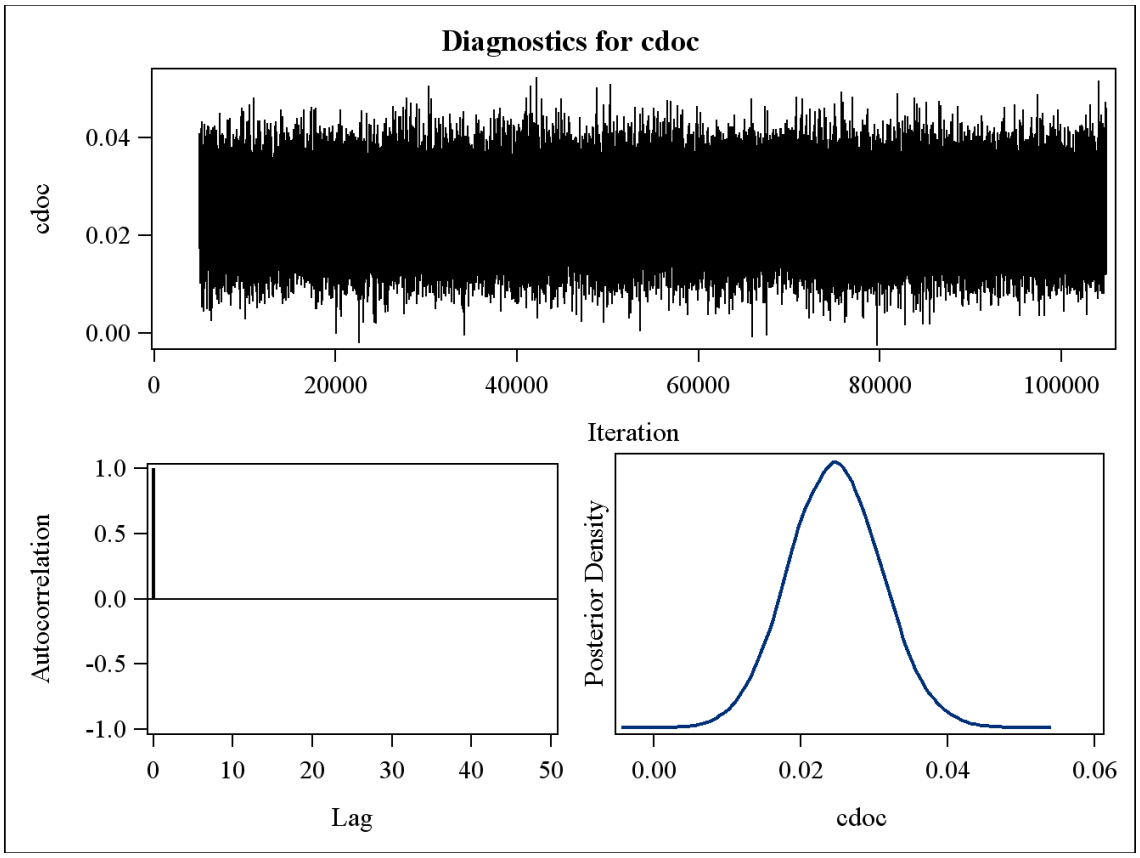
Heidelberger-Welch Diagnostics								
	Stationarity Test				Half-width Test			
Parameter	**Cramer-von-Mises Stat**	**p**	**Test Outcome**	**Iterations Discarded**	**Half-width**	**Mean**	**Relative Half-width**	**Test Outcome**
Intercept	0.3495	0.0986	Passed	0	0.000632	-2.0140	-0.00031	Passed
cdoc	0.1584	0.3649	Passed	0	0.000041	0.0247	0.00168	Passed
female	0.1102	0.5370	Passed	0	0.000579	2.2605	0.000256	Passed
kids	0.2728	0.1616	Passed	0	0.000745	0.3596	0.00207	Passed
cage	0.2726	0.1618	Passed	10000	0.000033	0.0545	0.000612	Passed

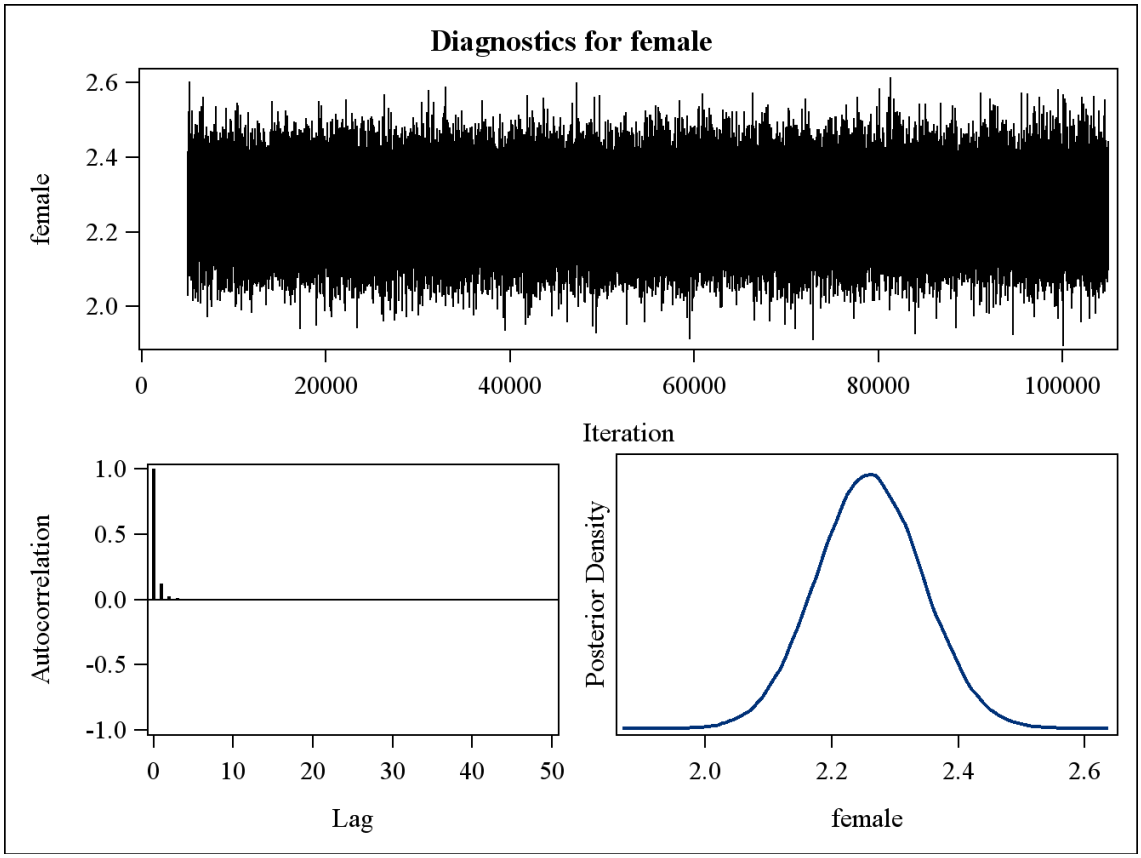
Effective Sample Sizes			
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
Intercept	56425.7	1.7722	0.5643
cdoc	98714.9	1.0130	0.9871
female	77017.1	1.2984	0.7702

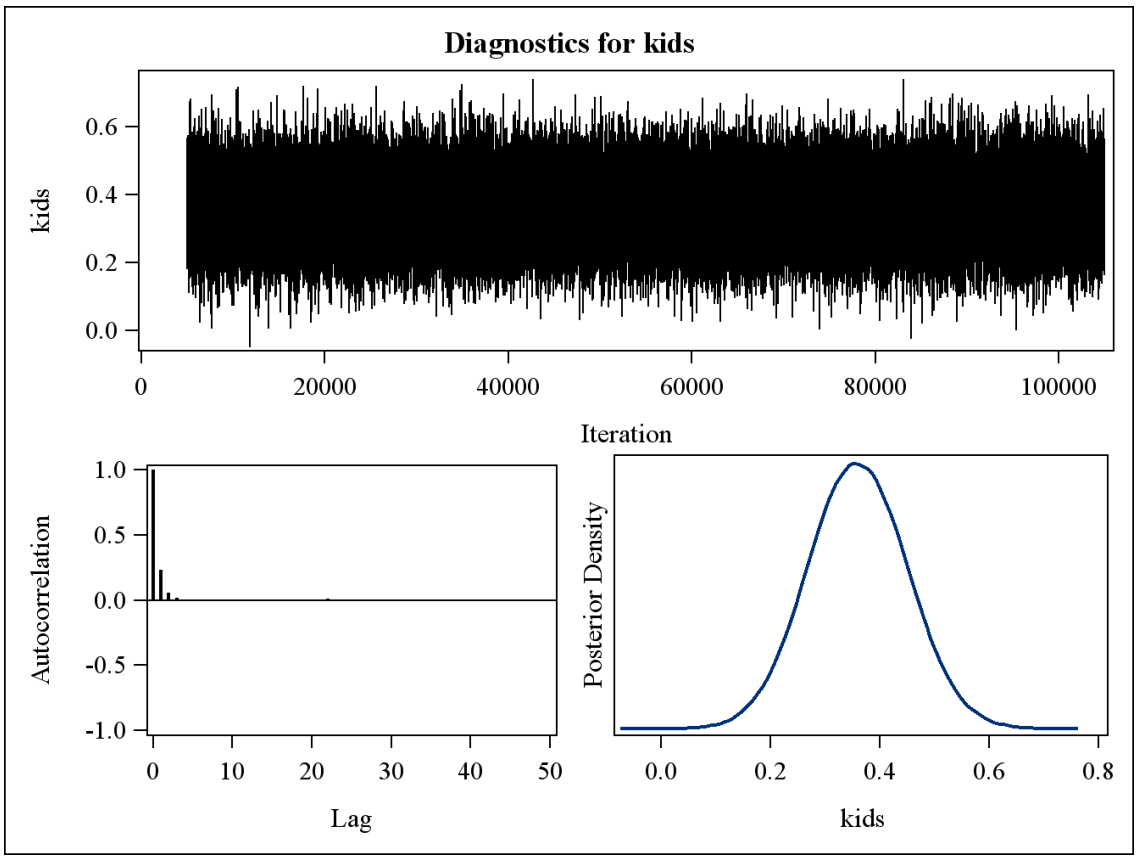
Effective Sample Sizes			
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
kids	61896.2	1.6156	0.6190
cage	58182.7	1.7187	0.5818

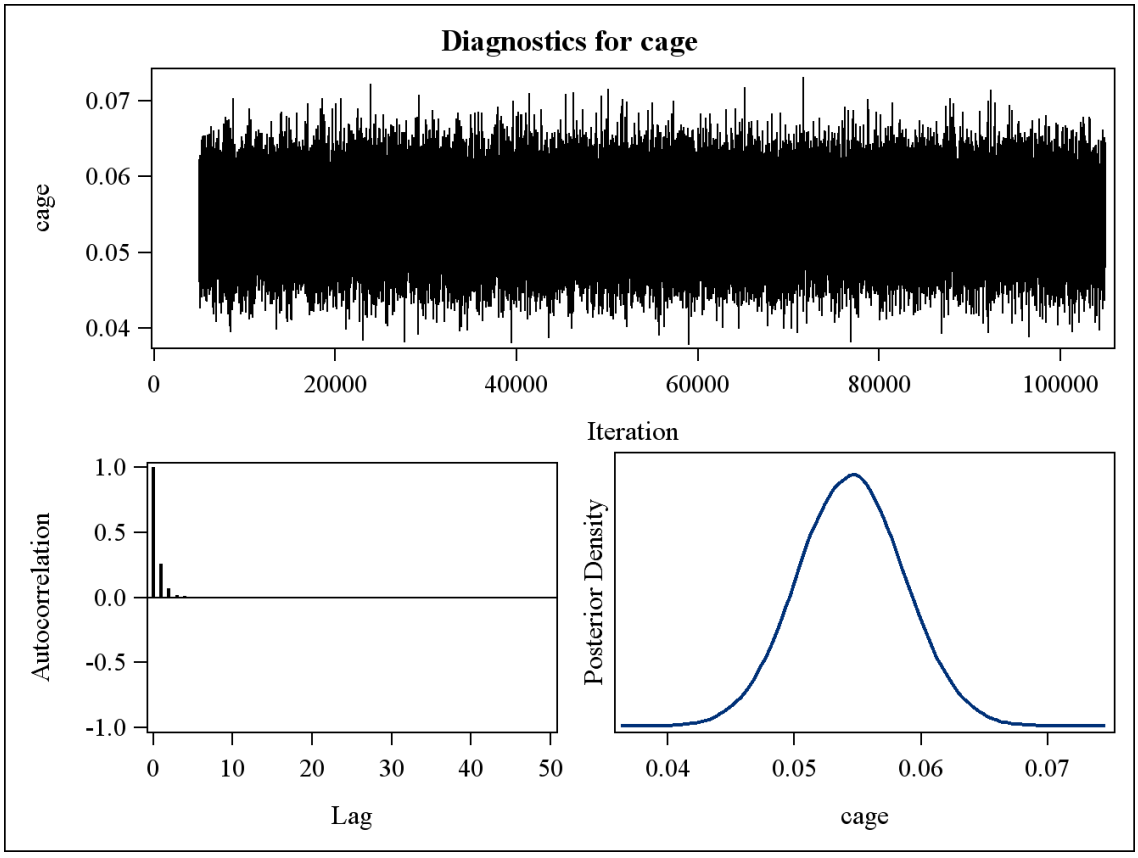
Monte Carlo Standard Errors			
Parameter	**MCSE**	**Standard Deviation**	**MCSE/SD**
Intercept	0.000343	0.0815	0.00421
cdoc	0.000020	0.00632	0.00318
female	0.000300	0.0832	0.00360
kids	0.000364	0.0907	0.00402
cage	0.000017	0.00418	0.00415




Diagnostics for cdoc
cdoc
0.04
0.02
0.00
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.00
0.02
0.04
0.06
cdoc


Diagnostics for female
female
2.6
2.4
2.2
2.0
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
2.0
2.2
2.4
2.6
female


Diagnostics for kids
kids
0.0
0.2
0.4
0.6
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.0
0.2
0.4
0.6
0.8
kids


Diagnostics for cage
cage
0.04
0.05
0.06
0.07
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.04
0.05
0.06
0.07
cage

```
*Create the normal prior;
data prior;
      input _type_ $ Intercept cdoc cage;
      datalines;
Var 1e5 1e5 1e5
Mean 0 0 0
;
run;

*Bayesian logistic regression with normal prior;
proc genmod  data=R84 descending;
      model outwork=cdoc female kids cage / dist=binomial link=logit;
      bayes seed=10231995 nbi=5000 nmc=100000
      coeffprior=normal(input=prior) diagnostics=all
      statistics=(summary interval) plots=all;
run;
```

Model Information	
Data Set	WORK.R84
Burn-In Size	5000
MC Sample Size	100000
Thinning	1
Sampling Algorithm	ARMS
Distribution	Binomial
Link Function	Logit
Dependent Variable	outwork

Number of Observations Read	3874
Number of Observations Used	3874
Number of Events	1420
Number of Trials	3874

Response Profile		
Ordered Value	**outwork**	**Total Frequency**
1	1	1420
2	0	2454

PROC GENMOD is modeling the probability that outwork='1'.

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**	
Intercept	1	-2.0103	0.0811	-2.1692	-1.8513
cdoc	1	0.0244	0.0063	0.0122	0.0367
female	1	2.2568	0.0828	2.0946	2.4190
kids	1	0.3580	0.0900	0.1817	0.5343
cage	1	0.0544	0.0042	0.0462	0.0625
Scale	0	1.0000	0.0000	1.0000	1.0000

Note: The scale parameter was held fixed.

Independent Normal Prior for Regression Coefficients		
Parameter	**Mean**	**Precision**
Intercept	0	0.00001
cdoc	0	0.00001
female	0	1E-6
kids	0	1E-6
cage	0	0.00001

Algorithm converged.

Initial Values of the Chain						
Chain	**Seed**	**Intercept**	**cdoc**	**female**	**kids**	**cage**
1	10231995	-2.01028	0.024432	2.256804	0.357976	0.054379
2		-1.8957	0.005644	2.008516	0.088084	0.041902
3		-2.12485	0.043221	2.505091	0.627868	0.066856

Fit Statistics	
DIC (smaller is better)	3928.284
pD (effective number of parameters)	5.032

Posterior Summaries						
				Percentiles		
Parameter	N	Mean	Standard Deviation	25%	50%	75%
Intercept	100000	-2.0140	0.0815	-2.0686	-2.0134	-1.9586
cdoc	100000	0.0247	0.00632	0.0204	0.0246	0.0289
female	100000	2.2605	0.0832	2.2043	2.2602	2.3166
kids	100000	0.3596	0.0907	0.2981	0.3590	0.4207
cage	100000	0.0545	0.00418	0.0516	0.0545	0.0573

Posterior Intervals					
		Equal-Tail Interval		HPD Interval	
Parameter	Alpha				
Intercept	0.050	-2.1755	-1.8557	-2.1710	-1.8520
cdoc	0.050	0.0124	0.0373	0.0124	0.0373
female	0.050	2.0989	2.4242	2.0971	2.4220
kids	0.050	0.1831	0.5382	0.1838	0.5386
cage	0.050	0.0463	0.0628	0.0464	0.0628

Posterior Autocorrelations				
Parameter	Lag 1	Lag 5	Lag 10	Lag 50
Intercept	0.2937	0.0021	0.0045	-0.0009
cdoc	0.0065	-0.0024	-0.0004	-0.0029
female	0.1193	-0.0024	-0.0034	-0.0013
kids	0.2308	0.0038	0.0004	0.0002
cage	0.2597	0.0059	0.0014	0.0029

Gelman-Rubin Diagnostics		
Parameter	Estimate	97.5% Bound
Intercept	1.0000	1.0000
cdoc	1.0000	1.0000
female	1.0000	1.0001
kids	1.0000	1.0000
cage	1.0000	1.0000

Geweke Diagnostics		
Parameter	**z**	**Pr > |z|**
Intercept	-1.4697	0.1417
cdoc	-1.1356	0.2561
female	0.1919	0.8478
kids	1.0302	0.3029
cage	2.3240	0.0201

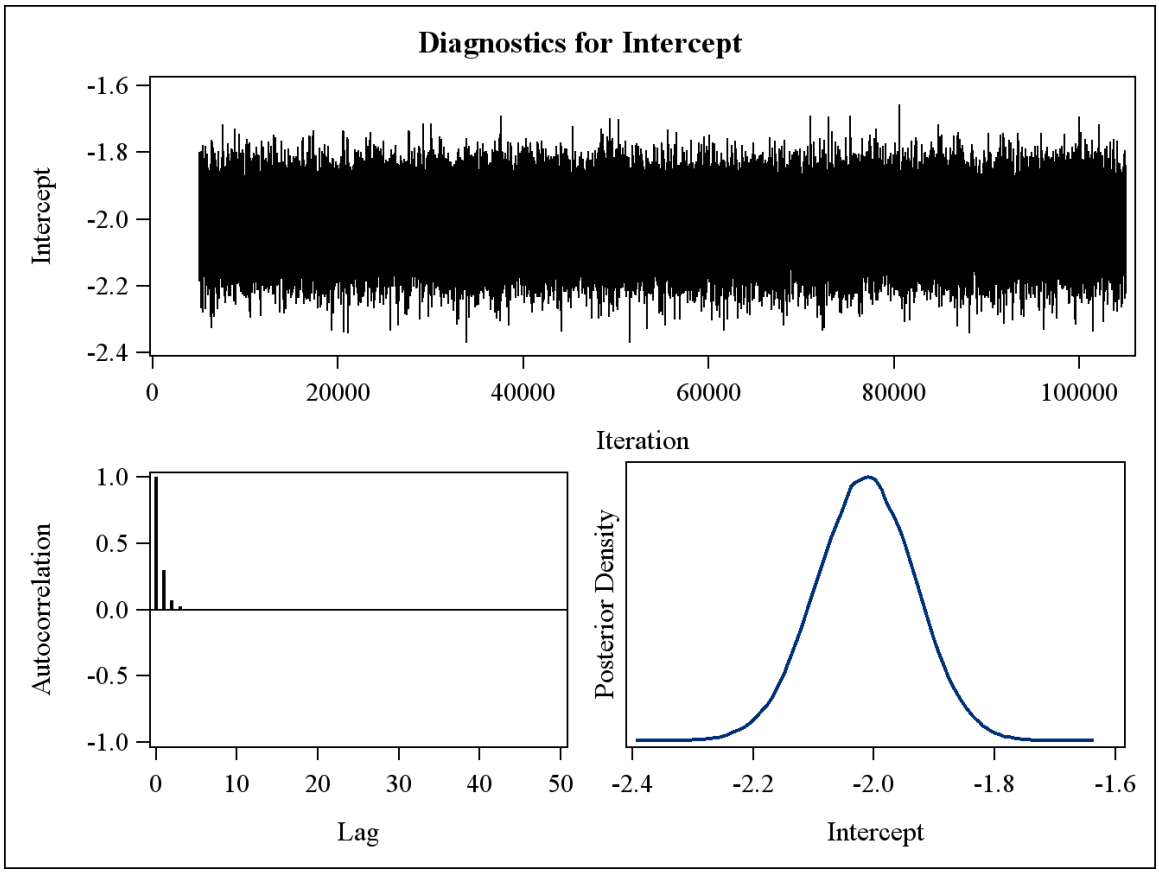
Raftery-Lewis Diagnostics				
Quantile=0.025 Accuracy=+/-0.005 Probability=0.95 Epsilon=0.001				
	Number of Samples			Dependence
Parameter	**Burn-in**	**Total**	**Minimum**	**Factor**
Intercept	3	4409	3746	1.1770
cdoc	2	3771	3746	1.0067
female	2	3865	3746	1.0318
kids	3	4095	3746	1.0932
cage	3	4140	3746	1.1052

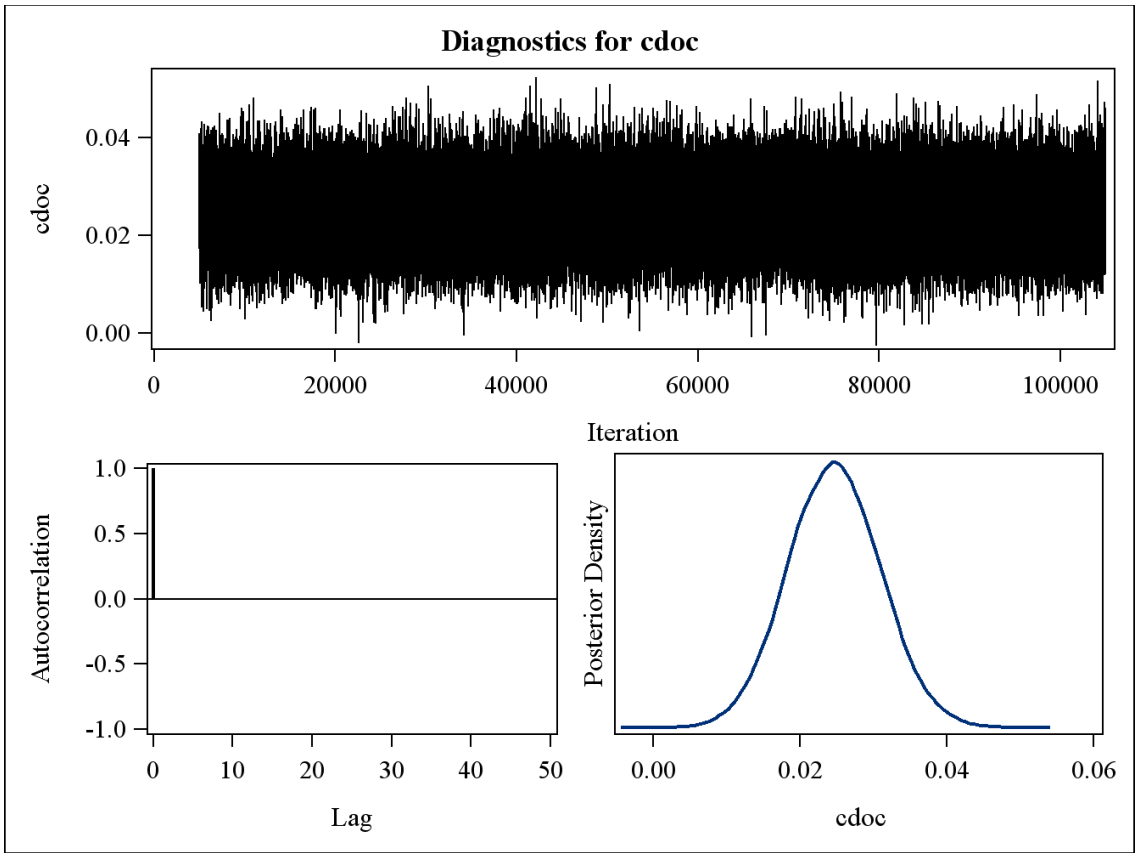
Heidelberger-Welch Diagnostics								
	Stationarity Test				Half-width Test			
Parameter	**Cramer-von-Mises Stat**	**p**	**Test Outcome**	**Iterations Discarded**	**Half-width**	**Mean**	**Relative Half-width**	**Test Outcome**
Intercept	0.3495	0.0986	Passed	0	0.000632	-2.0140	-0.00031	Passed
cdoc	0.1584	0.3649	Passed	0	0.000041	0.0247	0.00168	Passed
female	0.1102	0.5370	Passed	0	0.000579	2.2605	0.000256	Passed
kids	0.2728	0.1616	Passed	0	0.000745	0.3596	0.00207	Passed
cage	0.2726	0.1618	Passed	10000	0.000033	0.0545	0.000612	Passed

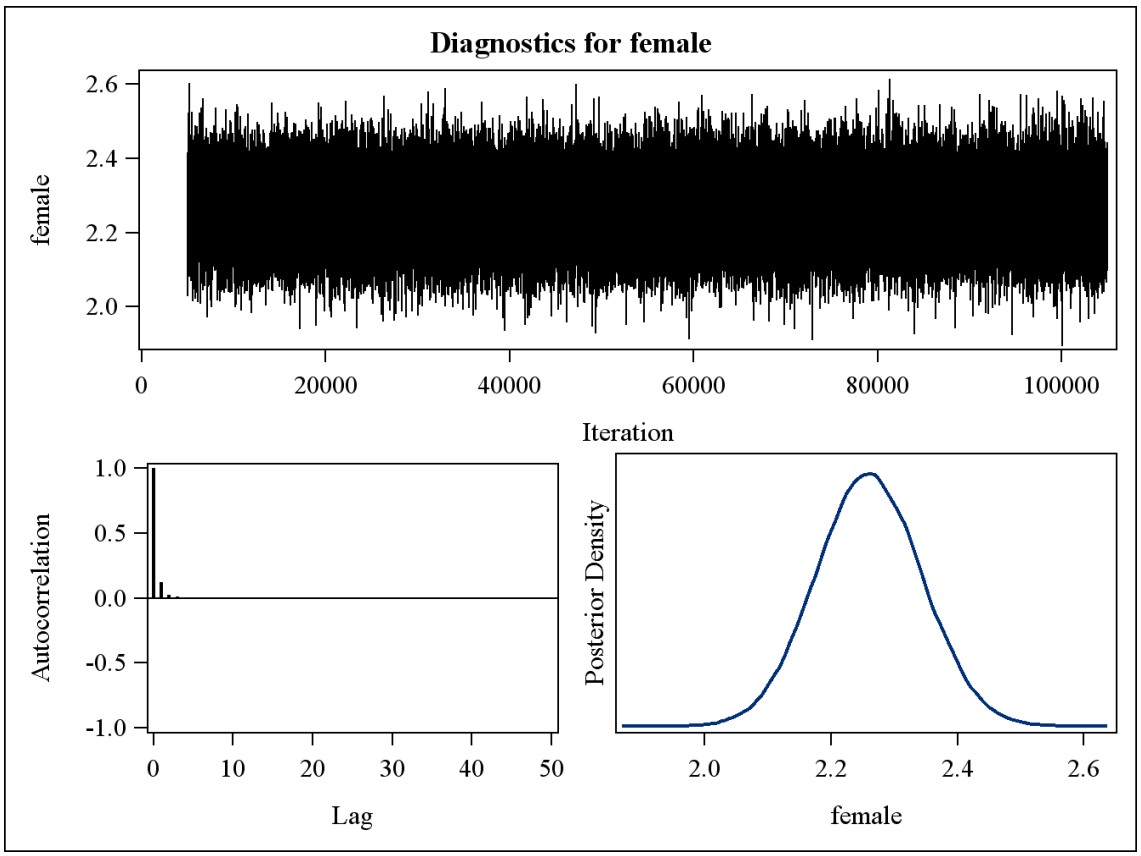
Effective Sample Sizes			
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
Intercept	56425.7	1.7722	0.5643
cdoc	98714.9	1.0130	0.9871

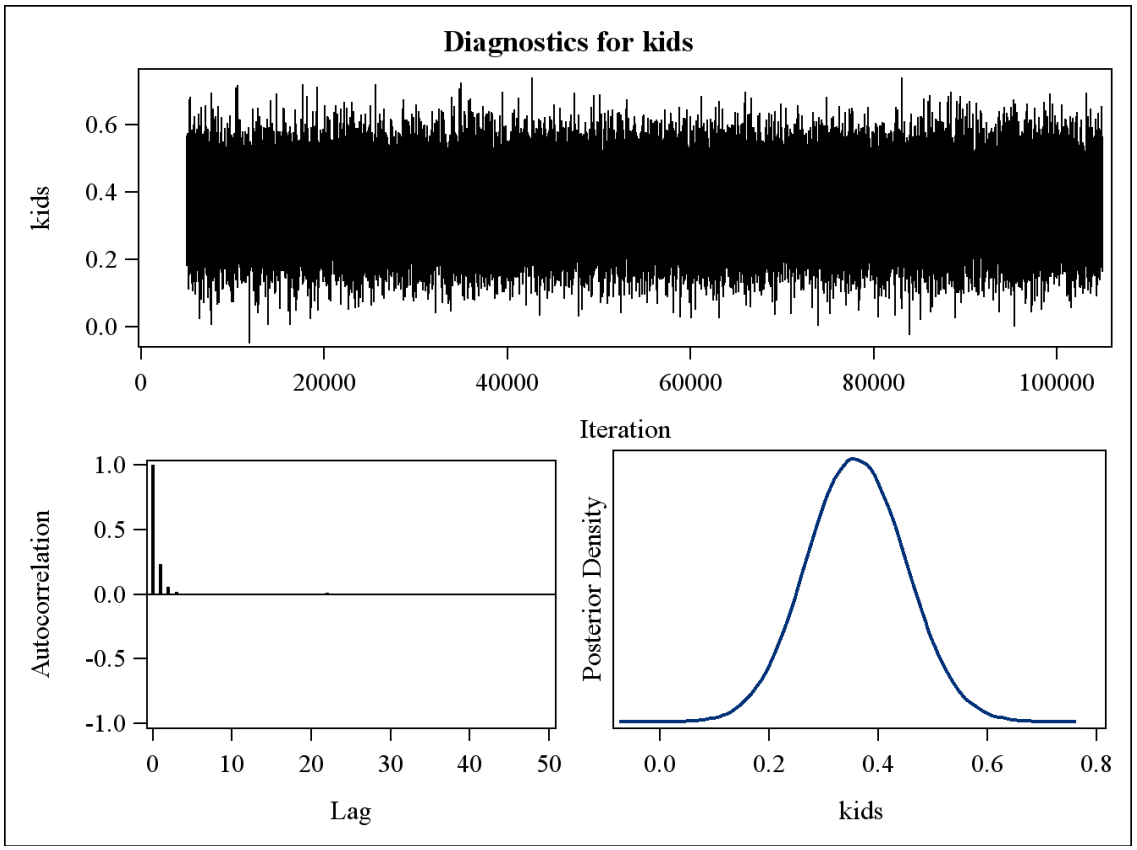
Effective Sample Sizes			
Parameter	**ESS**	**Autocorrelation Time**	**Efficiency**
female	77017.1	1.2984	0.7702
kids	61896.2	1.6156	0.6190
cage	58182.7	1.7187	0.5818

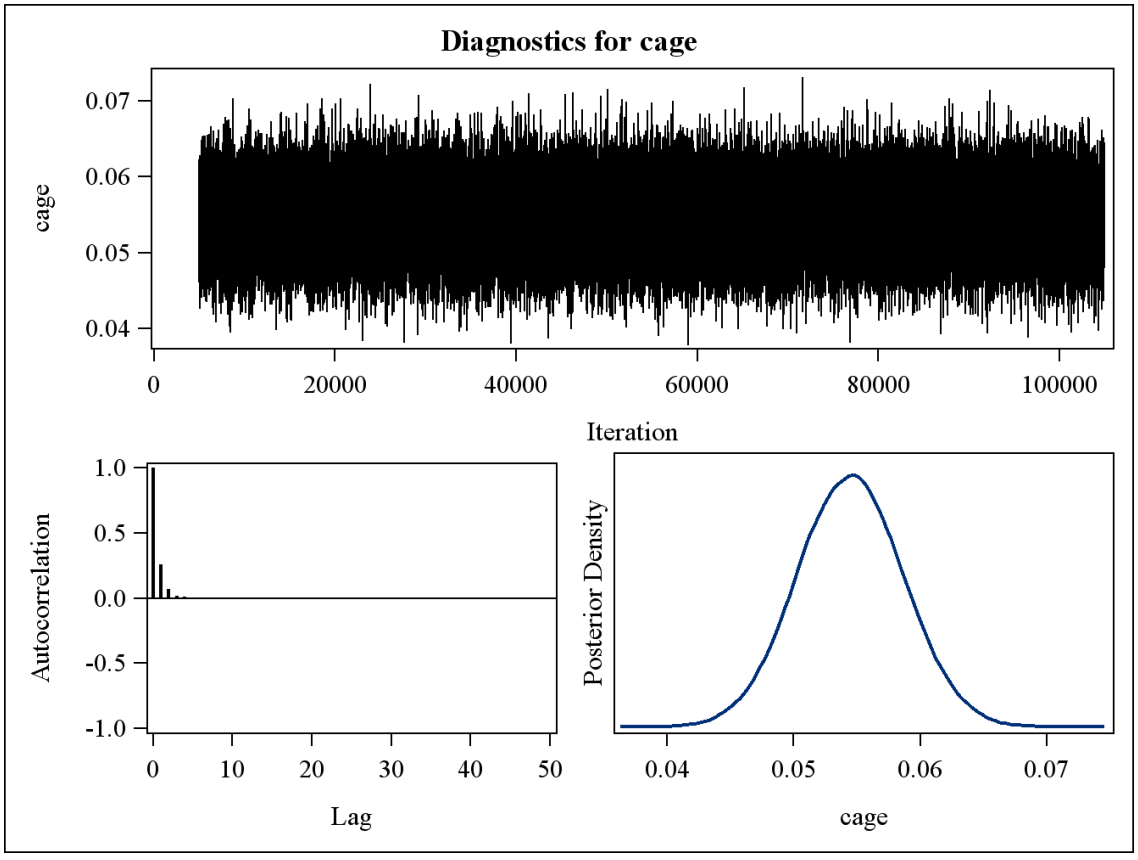
Monte Carlo Standard Errors			
Parameter	**MCSE**	**Standard Deviation**	**MCSE/SD**
Intercept	0.000343	0.0815	0.00421
cdoc	0.000020	0.00632	0.00318
female	0.000300	0.0832	0.00360
kids	0.000364	0.0907	0.00402
cage	0.000017	0.00418	0.00415




Diagnostics for cdoc
cdoc
0.04
0.02
0.00
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.00
0.02
0.04
0.06
cdoc


Diagnostics for female
female
2.6
2.4
2.2
2.0
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
2.0
2.2
2.4
2.6
female


Diagnostics for kids
kids
0.0
0.2
0.4
0.6
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.0
0.2
0.4
0.6
0.8
kids


Diagnostics for cage
cage
0.04
0.05
0.06
0.07
0
20000
40000
60000
80000
100000
Iteration
Autocorrelation
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
50
Lag
Posterior Density
0.04
0.05
0.06
0.07
cage