Categorical Data: Ordinal

- Ordered discrete response variable with a fixed number of classes c.

- For convenience we will number the categories $1 \dots c$.

- Consider calving difficulty:

Difficulty	Code
Easy	1
Moderate	2
Difficult	3

- In this case it is reasonable to think of a moderately difficult calving being somewhere between easy and difficult calving.

- However, it doesn't make sense to say that the difference between a moderately difficult and easy calving is the same as the difference between a difficult and an moderately difficult calving.

Response Variable

- The response on individual i will either be recorded as

$$y_i \in \{1 \ldots c\}$$

or as the $c \times 1$ vector

$$\boldsymbol{z}_i = \begin{pmatrix} z_{i1} \\ \vdots \\ z_{ic} \end{pmatrix}$$

with

$$z_{ij} = \begin{cases} 1 & y_i = j \\ 0 & \text{otherwise.} \end{cases}$$

Multinomial distribution

$$y_i \sim \text{Mult}(1, \pi_{i1} \ldots \pi_{ic})$$

- Mean

$$\text{E}(\boldsymbol{z}_i) = \boldsymbol{\mu}_i = \boldsymbol{\pi}_i = \begin{pmatrix} \pi_{i1} \\ \vdots \\ \pi_{ic} \end{pmatrix}$$

- Covariance Matrix

$$\text{var}(\boldsymbol{z}_i) = \boldsymbol{R}_i = \text{Diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i \boldsymbol{\pi}_i'$$

Link Function

- Recall: Logit

$$\begin{aligned} \boldsymbol{x}_i'\boldsymbol{\alpha} &= \ln\left(\frac{\mu_i}{1-\mu_i}\right) \\ \mu_i &= \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\alpha})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\alpha})} \\ \Pr(y_i = 0) &= \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\alpha})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\alpha})} \end{aligned}$$

- Generalize this by modeling

$$\Pr(y_i \leq j) = \frac{\exp(I_j + \boldsymbol{x}_i'\boldsymbol{\alpha})}{1 + \exp(I_j + \boldsymbol{x}_i'\boldsymbol{\alpha})}$$

$$I_j + \boldsymbol{x}_i'\boldsymbol{\alpha} = \ln\left(\frac{\Pr(y_i \leq j)}{1 - \Pr(y_i \leq j)}\right)$$

$$\Pr(y_i = j) = \mu_{ij} = \Pr(y_i \leq j) - \Pr(y_i \leq j - 1)$$

where

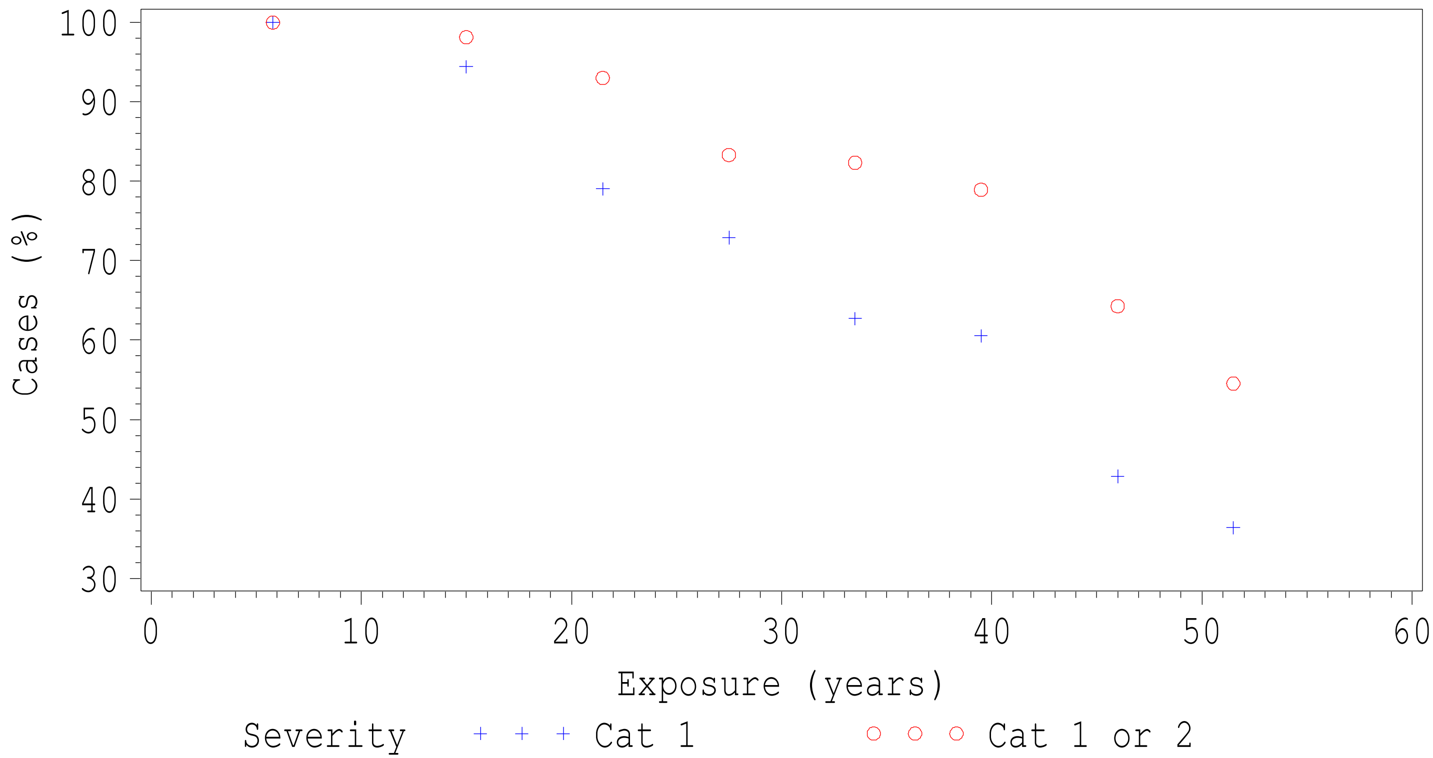
$$\Pr(y_i \leq 0) = 0$$

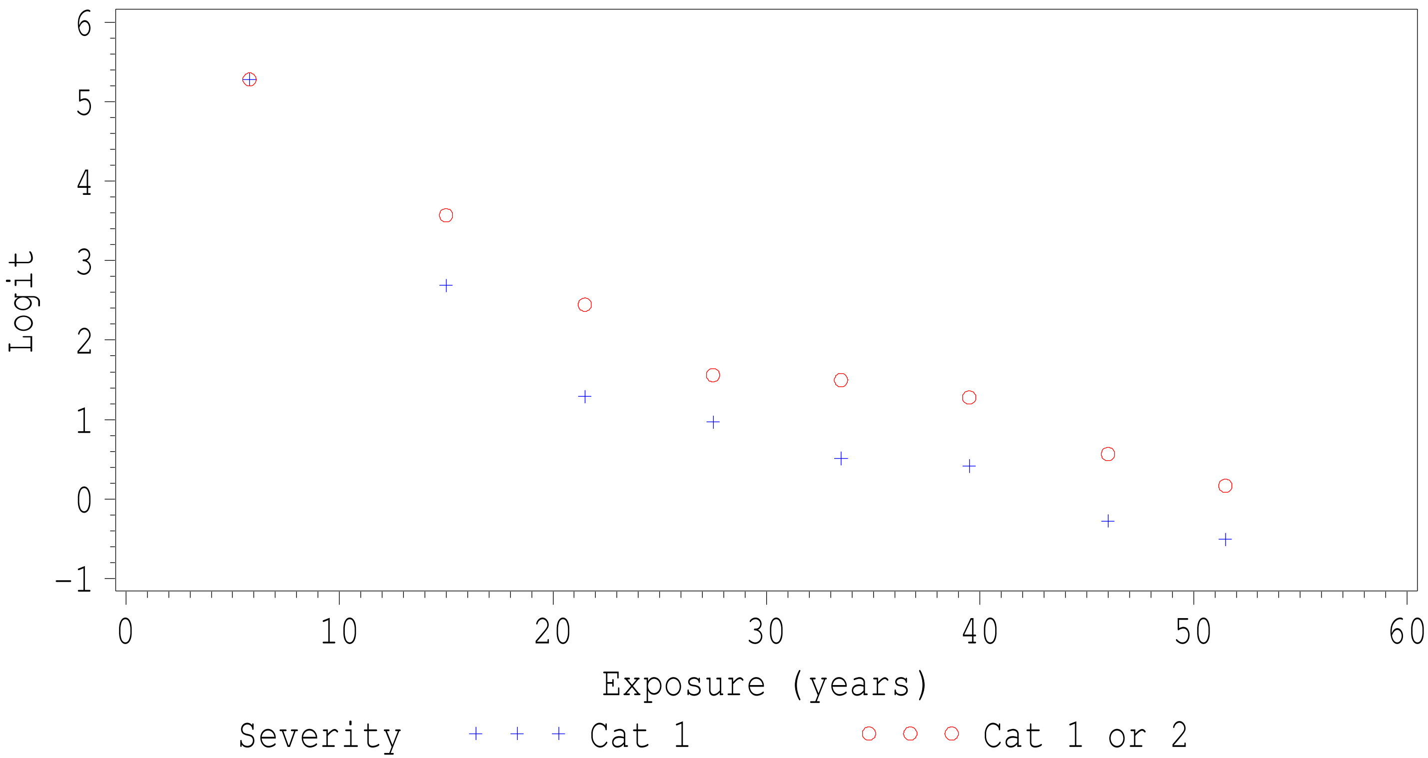
$$\Pr(y_i \leq c) = 1$$

Coal miner pneumoconiosis data set

Exposure	Diagnosis		
Years	Normal	Moderate	Severe
5.8	98	0	0
15.0	51	2	1
21.5	34	6	3
27.5	35	5	8
33.5	32	10	9
39.5	23	7	8
46.0	12	6	10
51.5	4	2	5

McCullagh and Nelder (1989) Generalized Linear Models, pg. 179


100
90
80
70
60
50
40
30
Cases (%)
0
10
20
30
40
50
60
Exposure (years)
Severity
Cat 1
Cat 1 or 2



Logit
6
5
4
3
2
1
0
-1
0
10
20
30
40
50
60
Exposure (years)
Severity
Cat 1
Cat 1 or 2

Model

$$y_i \text{ ind } \mathrm{Mult}(1, \pi_{iN}, \pi_{iM}, \pi_{iS})$$

$$\ln(\pi_{iN}/(1-\pi_{iN})) = I_N + b * Log(Exp)_i$$

$$\ln(\pi_{iN} + \pi_{iM}/(1-\pi_{iN}-\pi_{iM})) = I_M + b * Log(Exp)_i$$

$$\ln((1-\pi_{iS})/\pi_{iS}) = I_M + b * Log(Exp)_i$$

$$\ln(\pi_{iS}/(1-\pi_{iS})) = -I_M - b * Log(Exp)_i$$

- Parameters
  - Intercepts: I_N and I_M
  - Slope: b

Program

```
proc genmod data=miner;
    freq n;
    model score=logtime/dist=multinomial link=cumlogit type3;
    run;
```

Results

Model Information

Data Set	WORK.MINER
Distribution	Multinomial
Link Function	Cumulative Logit
Dependent Variable	score
Frequency Weight Variable	n

Number of Observations Read	22
Number of Observations Used	22
Sum of Frequencies Read	371
Sum of Frequencies Used	371

Response Profile

Ordered Value	score	Total Frequency
1	1	289
2	2	38
3	3	44

PROC GENMOD is modeling the probabilities of levels of score having LOWER Ordered Values in the response profile table. One way to change this to model the probabilities of HIGHER Ordered Values is to specify the DESCENDING option in the PROC statement.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Log Likelihood		-204.2742	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept1	1	9.6761	1.3233	7.0826	12.2696	53.47	<.0001
Intercept2	1	10.5817	1.3437	7.9481	13.2154	62.02	<.0001
logtime	1	-2.5968	0.3810	-3.3435	-1.8502	46.47	<.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
logtime	1	96.61	<.0001

Lack of Fit

```
proc genmod data=miner;
    freq n;
    class time;
    model score=logtime time/
          dist=multinomial link=cumlogit type1;
    run;
```

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Log Likelihood		-202.6940	

Algorithm converged.

LR Statistics For Type 1 Analysis

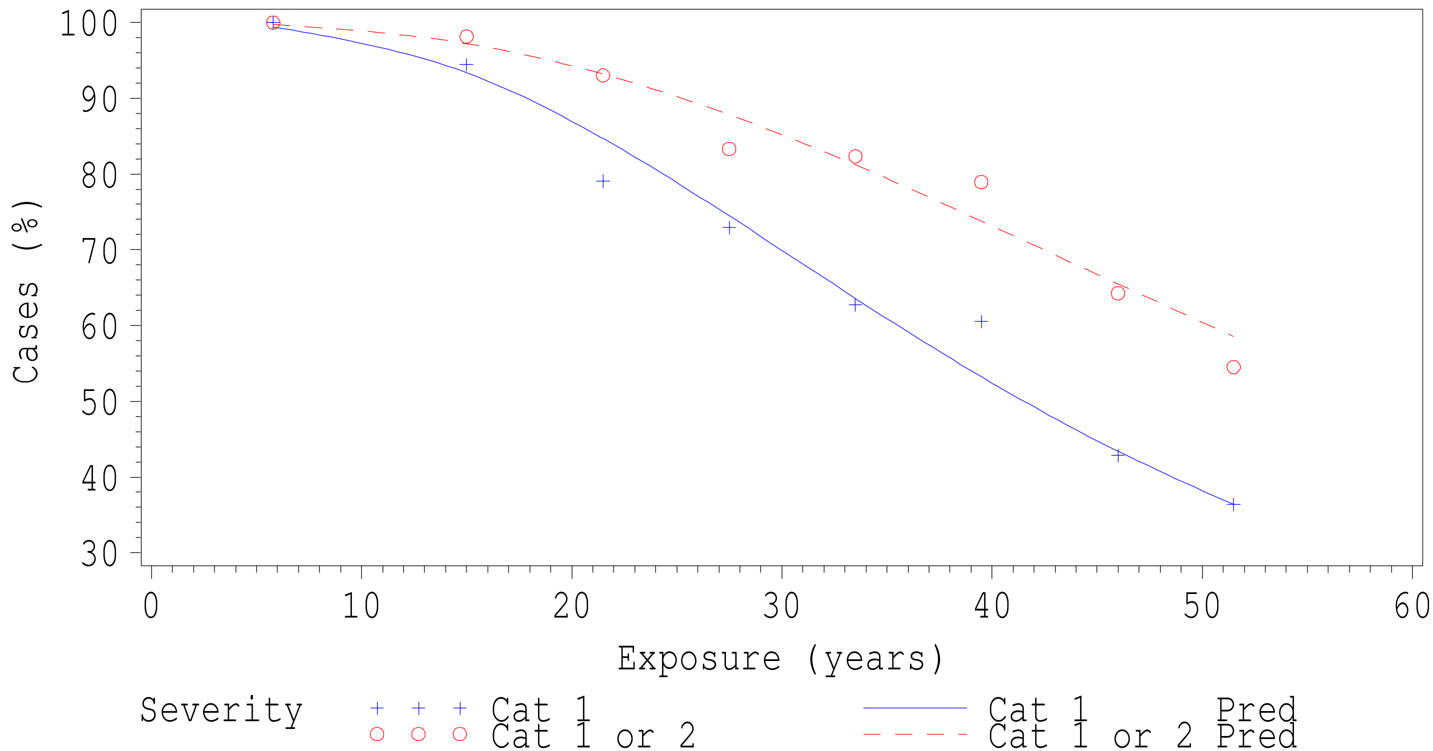
Source	Deviance	DF	Chi-Square	Pr > ChiSq
Intercepts	1010.3241			
logtime	817.0967	1	96.61	<.0001
time	810.7760	6	3.16	0.7885

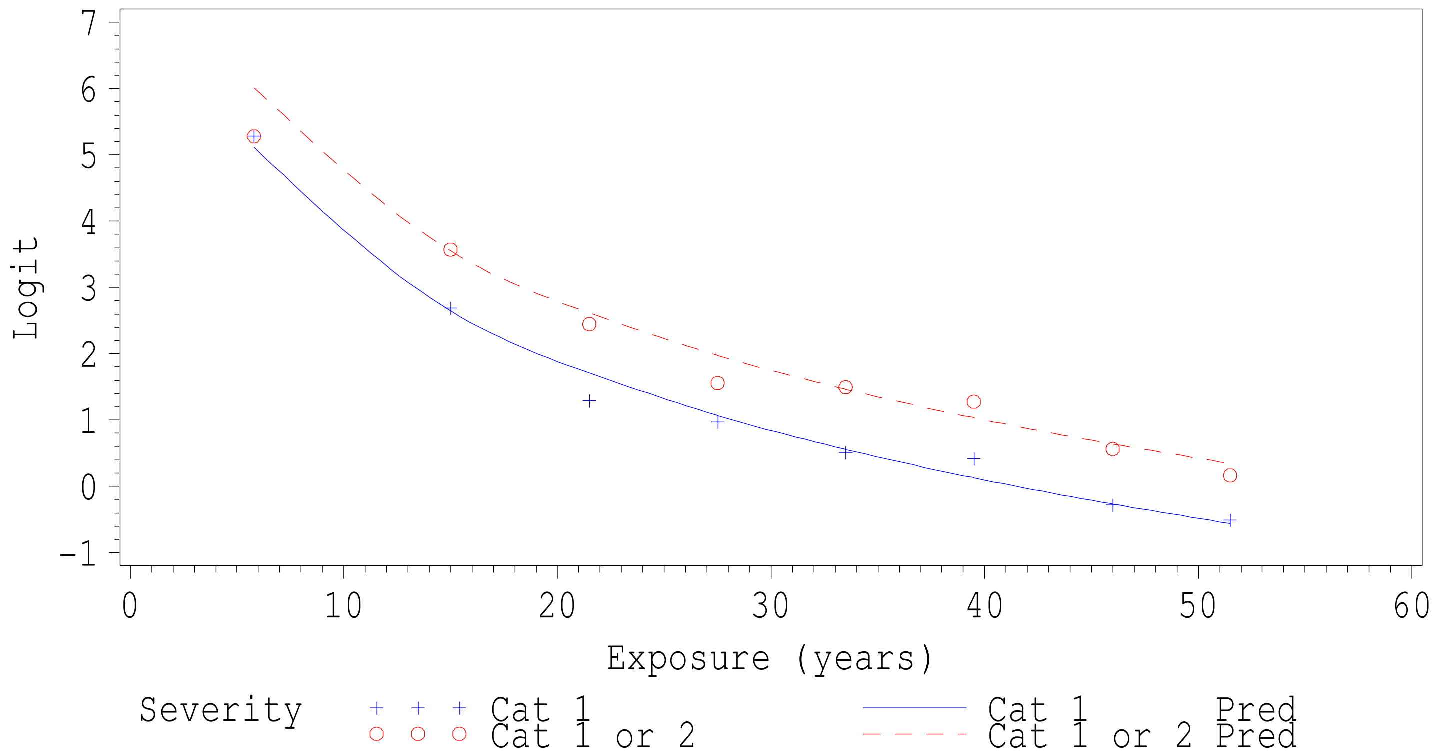
$$2(-202.694 - (-204.2742)) = 3.1604$$

Fitted Values

```
%macro invlogit(xbeta);
    exp(&xbeta)/(1+exp(&xbeta));
    %mend;

data minerlogit;set miner3;
xbeta=-2.5968*logtime;
int=9.6761;
if Severity="Cat 1 or 2" then int=10.5817;
output;
Severity=substr(Severity,1,10) || " Pred";
pestl=int+xbeta;
prob=100*%invlogit(int+xbeta);
output;
```


Cases (%)
100
90
80
70
60
50
40
30
0
10
20
30
40
50
60
Exposure (years)
Severity
Cat 1
Cat 1 or 2
Cat 1 Pred
Cat 1 or 2 Pred


7
6
5
4
3
2
1
0
−1
Logit
0
10
20
30
40
50
60
Exposure (years)
Severity
Cat 1
Cat 1 or 2
Cat 1 Pred
Cat 1 or 2 Pred

Effect of Surface and Vision on Balance

http://www.statsci.org/data/oz/ctsib.html

- Factors
  - Sex (Female and Male)
  - Surface (Normal and Foam)
  - Vision (Closed, Dome, and Open)

- 40 Subjects
  - Age (yr)
  - Weight (kg)
  - Height (cm)
- Each treatment repeated twice for each subject

Effects

- Fixed
  - Sex|Surface|Vision
  - Age, Weight, and Height as Covariates

- Random
  - Subject
  - Surface*Vision*Subject

SAS

```
%let DIR=h:/mixed-model;
data balance;
    infile "&DIR/ctsibuni.txt" firstobs=2 expandtabs;
        length sex $6 Vision $ 6;
    input Subject Sex  Age Height Weight Surface $ Vision  CTSIB;
    CTSIB=min(CTSIB,3)
run;
proc print;
run;
proc means data=balance;
var age weight height;
run;
```

```
proc genmod data=balance;
class sex Vision Subject Surface;
model score=age weight height sex|Vision|Surface/dist=mult type3;
```

Genmod

The GENMOD Procedure

Model Information

Data Set	WORK.BALANCE
Distribution	Multinomial
Link Function	Cumulative Logit
Dependent Variable	score

Number of Observations Read	480
Number of Observations Used	480

Class Level Information

Class	Levels	Values
sex	2	female male
Vision	3	closed dome open
Subject	40	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Surface	2	foam norm

Response Profile

Ordered Value	score	Total Frequency
1	1	114
2	2	292
3	3	74

PROC GENMOD is modeling the probabilities of levels of score having LOWER Ordered Values in the response profile table. One way to change this to model the probabilities of HIGHER Ordered Values is to specify the DESCENDING option in the PROC statement.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Log Likelihood		-243.1001	

Algorithm converged.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
Age	1	0.16	0.6912
Weight	1	23.87	<.0001
Height	1	24.88	<.0001
sex	1	7.30	0.0069
Vision	2	182.83	<.0001
sex*Vision	2	4.60	0.1001
Surface	1	269.08	<.0001
sex*Surface	1	0.44	0.5081
Vision*Surface	2	7.07	0.0291
sex*Vision*Surface	2	3.09	0.2134

GLIMMIX

```
proc glimmix data=balance;
class sex Vision Subject Surface;
model score=age weight height sex|Vision|Surface
    /dist=mult ddfm=satterthwaite;
random intercept surface*Vision/subject=subject(sex);
estimate 'closed vs dome' vision 1 -1 0;
estimate 'open vs rest'  vision -.5 -.5 1;
estimate 'foam vs norm' surface 1 -1;
estimate 'F vs M  closed' sex 1 -1  sex*vision 1 0 0 -1 0 0 ;
estimate 'F vs M  dome' sex 1 -1  sex*vision  0 1  0 0 -1  0;
estimate 'F vs M  open' sex 1 -1  sex*vision 0 0 1 0 0 -1;
estimate 'Weight' weight 1;
estimate 'Height' height 1;
```

The GLIMMIX Procedure

Model Information

Data Set	WORK.BALANCE
Response Variable	score
Response Distribution	Multinomial (ordered)
Link Function	Cumulative Logit
Variance Function	Default
Variance Matrix Blocked By	Subject(sex)
Estimation Technique	Residual PL
Degrees of Freedom Method	Satterthwaite

Class Level Information

Class	Levels	Values
sex	2	female male
Vision	3	closed dome open
Subject	40	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Surface	2	foam norm

Number of Observations Read	480
Number of Observations Used	480

Response Profile

Ordered Value	score	Total Frequency
1	1	114
2	2	292
3	3	74

The GLIMMIX procedure is modeling the probabilities of levels of score having lower Ordered Values in the Response Profile table.

Dimensions

G-side Cov. Parameters	2
Columns in X	40
Columns in Z per Subject	7
Subjects (Blocks in V)	40
Max Obs per Subject	12

Optimization Information

Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	2
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

```
Convergence criterion (PCONV=1.11022E-8) satisfied.

                Fit Statistics

-2 Res Log Pseudo-Likelihood         6888.28
Pseudo-AIC  (smaller is better)      6892.28
Pseudo-AICC (smaller is better)      6892.30
Pseudo-BIC  (smaller is better)      6895.65
Pseudo-CAIC (smaller is better)      6897.65
Pseudo-HQIC (smaller is better)      6893.50
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	Subject(sex)	4.0077	1.4750
Vision*Surface	Subject(sex)	1.6234	0.5589

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Age	1	25.64	0.01	0.9046
Weight	1	27.15	4.18	0.0507
Height	1	27.18	4.91	0.0353
sex	1	27.23	1.93	0.1756
Vision	2	360.4	29.62	<.0001
sex*Vision	2	229.6	2.25	0.1080
Surface	1	464	63.34	<.0001
sex*Surface	1	221.6	0.32	0.5696
Vision*Surface	2	362.9	1.18	0.3083
sex*Vision*Surface	2	266.7	0.55	0.5784

Estimates

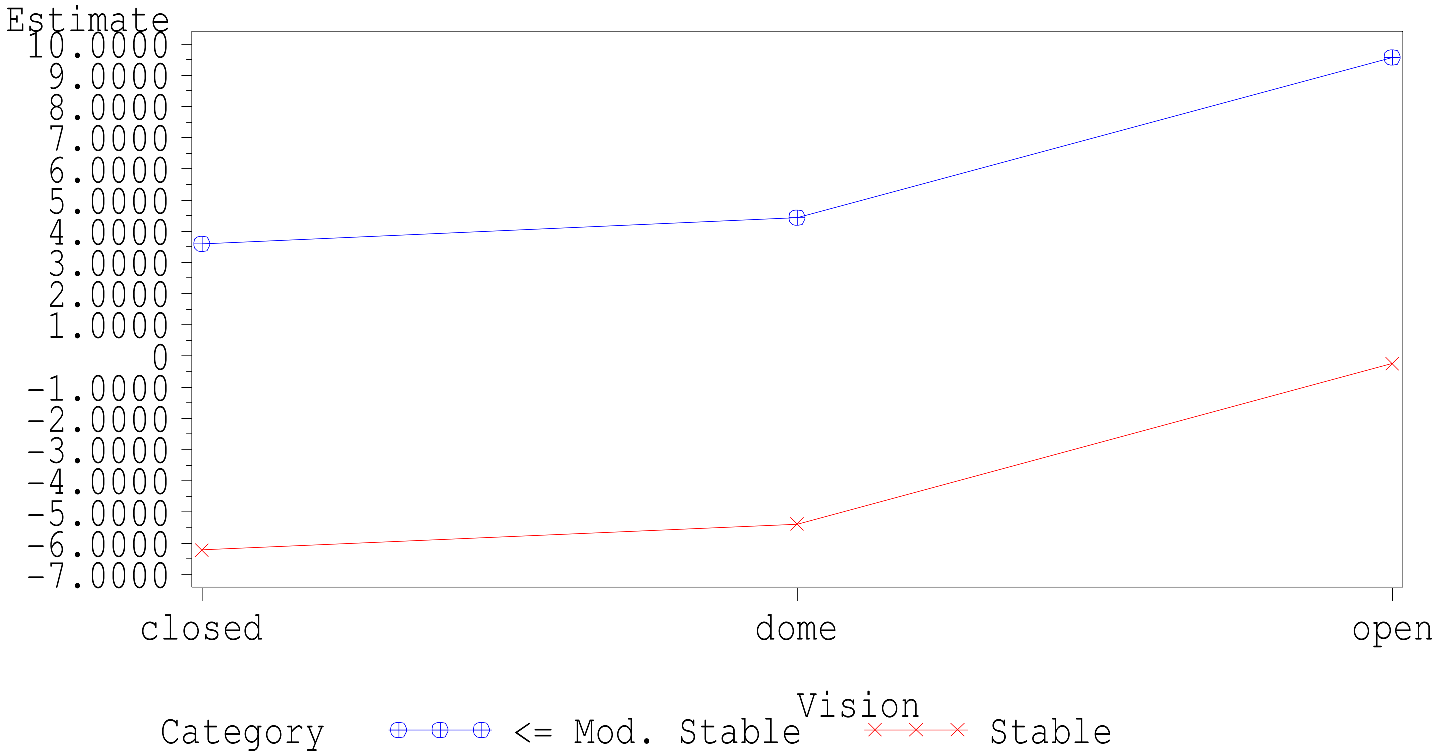
Label	Estimate	Standard Error	DF	t Value	Pr > |t|
closed vs dome	-0.8305	0.4339	214.7	-1.91	0.0569
open vs rest	5.5587	0.7270	464	7.65	<.0001
foam vs norm	-6.7815	0.8521	464	-7.96	<.0001
F vs M closed	-2.8479	1.3687	38.82	-2.08	0.0441
F vs M dome	-1.3362	1.3274	34.9	-1.01	0.3211
F vs M open	-1.0147	1.3608	38.9	-0.75	0.4604
Weight	0.09269	0.04533	27.15	2.04	0.0507
Height	-0.1464	0.06605	27.18	-2.22	0.0353

LSMEANS

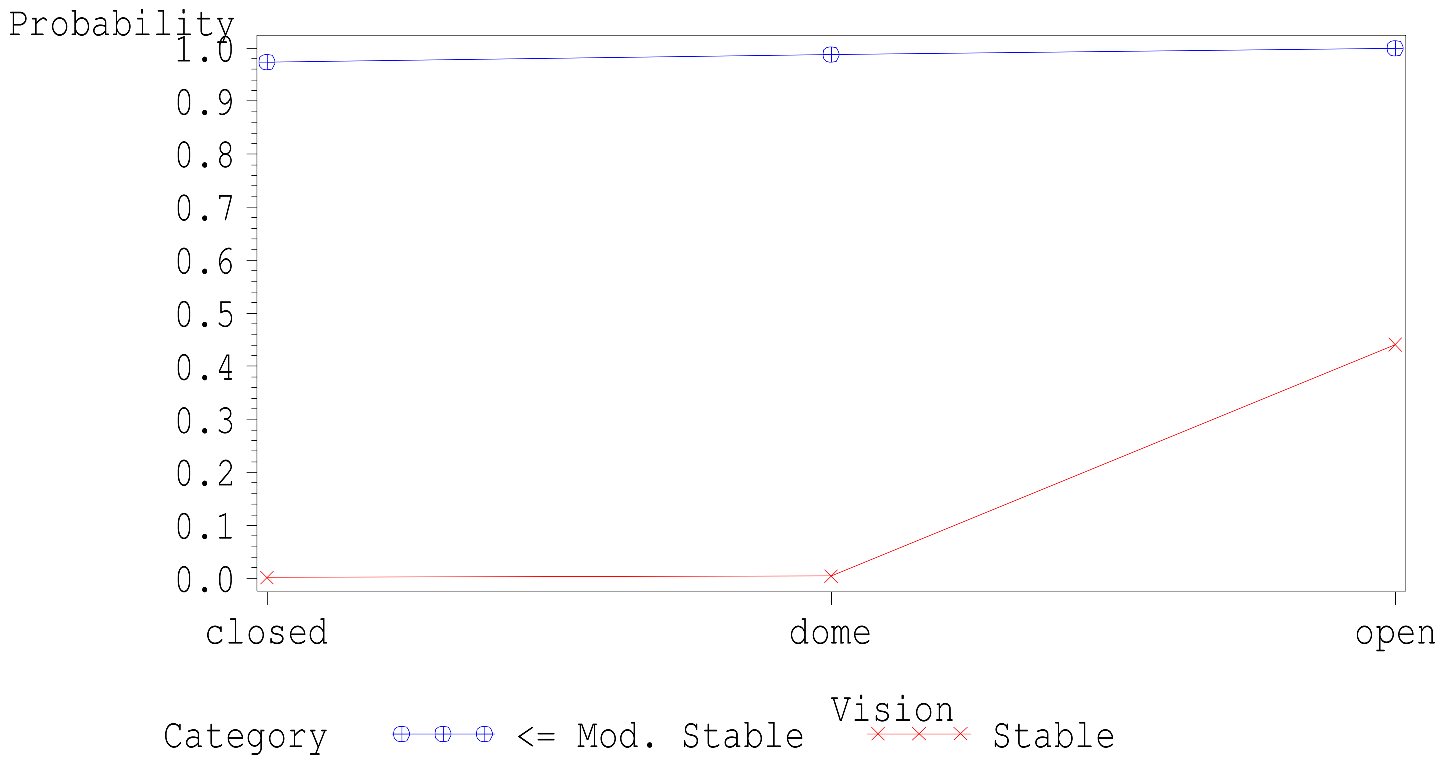
```
estimate 'LSM Vision  closed 1' intercept 1 0 sex .5 .5  age 21.8
          weight 71.145 height 172.05
          Vision 1 0 0 Surface .5 .5 sex*vision .5 0 0 .5  0 0
          surface*vision .5 .5  0 0 0 0
          sex*surface .25 .25 .25 .25
          sex*vision*surface .25 .25 0 0 0 0 .25 .25 0 0 0 0 ;
estimate 'LSM Vision  dome   1' intercept 1 0 sex .5 .5 age 21.8
          weight 71.145 height 172.05
          Vision 0 1 0 Surface .5 .5 sex*vision 0 .5 0 0 .5  0
          surface*vision 0 0 .5 .5  0 0
          sex*surface .25 .25 .25 .25
          sex*vision*surface 0 0 .25 .25 0 0 0 0 .25 .25 0 0 ;
ods output estimates=lsm;
```

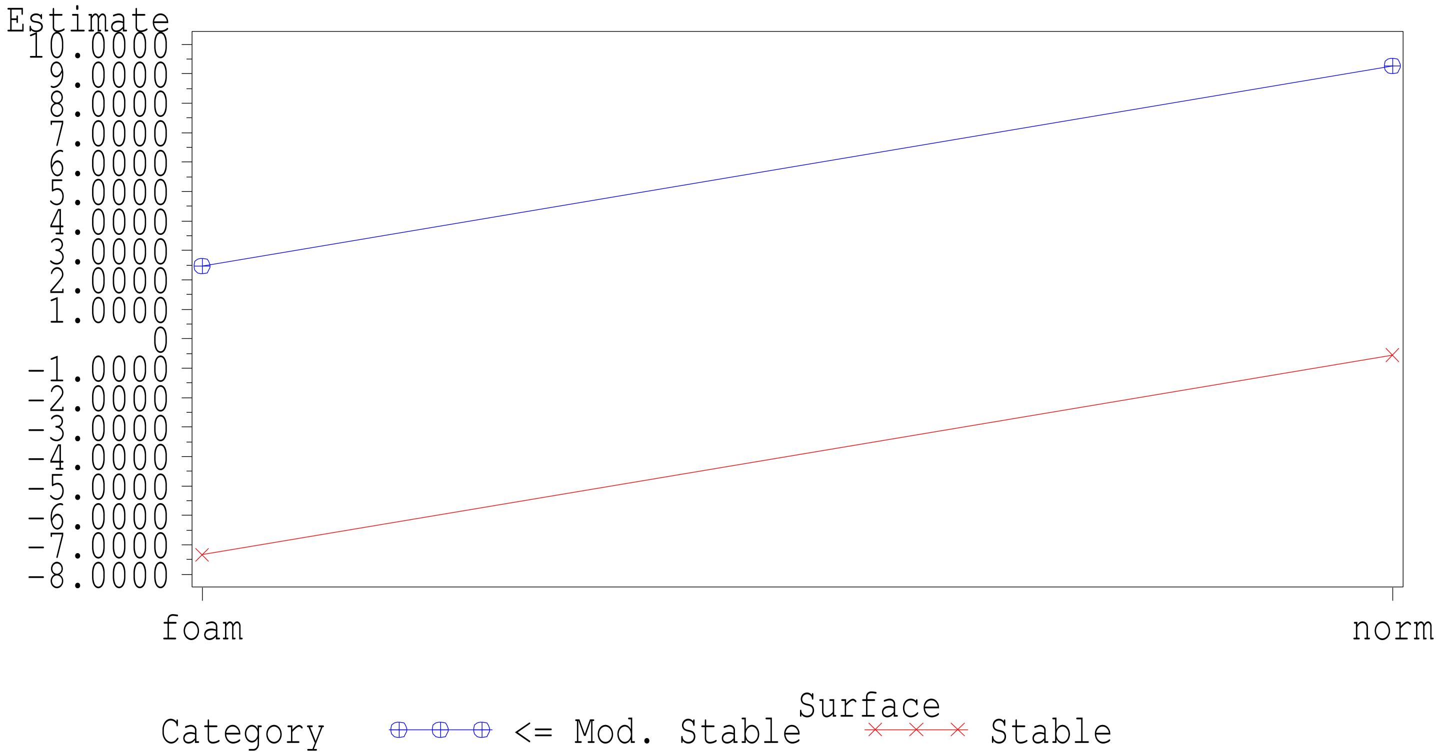
```
data lsm;set lsm;
length cat $ 15;
if substr(label,1,3) = "LSM" then do;
   var=scan(label,2);
   lev=scan(label,3);
   c=scan(label,4);
   if c=1 then cat="Stable";
   if c=2 then cat="<= Mod. Stable";
   label cat="Category";
   prob=exp(estimate)/(1+exp(estimate));
   label prob="Probability";
   output;
end;
```

Obs	var	lev	cat	Estimate	prob
1	Vision	closed	Stable	-6.2103	0.00200
2	Vision	dome	Stable	-5.3798	0.00459
3	Vision	open	Stable	-0.2364	0.44118
4	Vision	closed	<= Mod. Stable	3.6023	0.97346
5	Vision	dome	<= Mod. Stable	4.4329	0.98826
6	Vision	open	<= Mod. Stable	9.5763	0.99993
7	Surface	foam	Stable	-7.3329	0.00065
8	Surface	norm	Stable	-0.5514	0.36554
9	Surface	foam	<= Mod. Stable	2.4797	0.92271
10	Surface	norm	<= Mod. Stable	9.2613	0.99990

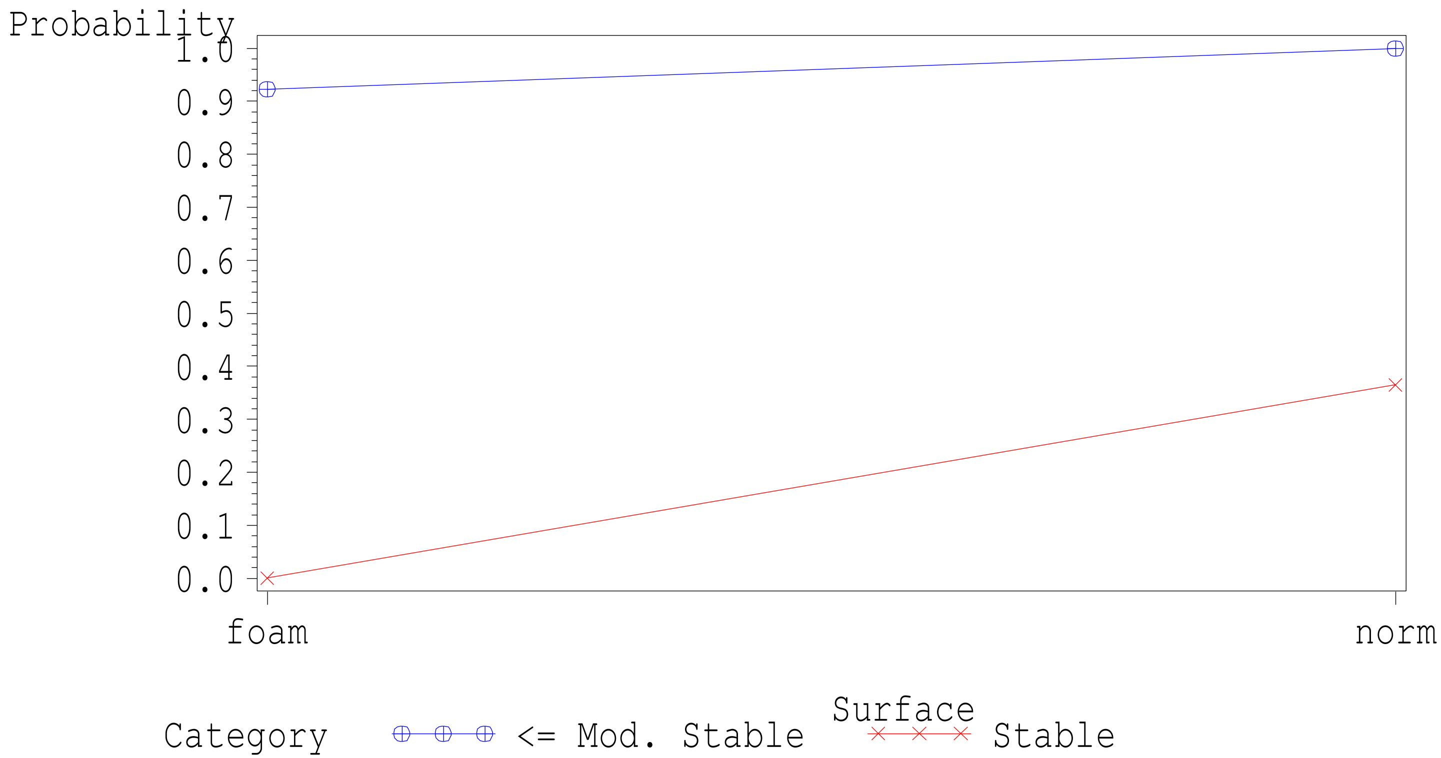


Estimate
10.0000
9.0000
8.0000
7.0000
6.0000
5.0000
4.0000
3.0000
2.0000
1.0000
0
-1.0000
-2.0000
-3.0000
-4.0000
-5.0000
-6.0000
-7.0000
closed
dome
open
Vision
Category
<= Mod. Stable
Stable


Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
closed
dome
open
Vision
Category
<= Mod. Stable
Stable



Estimate
10.0000
9.0000
8.0000
7.0000
6.0000
5.0000
4.0000
3.0000
2.0000
1.0000
0
-1.0000
-2.0000
-3.0000
-4.0000
-5.0000
-6.0000
-7.0000
-8.0000
foam
norm
Surface
Category
<= Mod. Stable
Stable


Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
foam
norm
Surface
Category
<= Mod. Stable
Stable