

# Logistic regression with more than two outcomes

```
/* mathlogreg4.sas */
%include 'readmath.sas';
title2 'Logistic regression with more than 2 resp. categories using
proc catmod';

/***** Data step continues *****/
if sex = 'Female' then gender=1; else if sex = 'Male' then gender=0;
if tongue = 'English' then mtongue=1; else if tongue='Other' then mtongue=0;
if course=. then course2=0;
    else if course=4 then course2=0; /* 0=No Data */
    else course2=course;
hsutil = hsgpa+hscalch+hsengl;
if hsutil = . then hsmiss=1; else hsmiss=0;
label hsmiss = 'Missing Any High School Data'; format hsmiss ynfmt.;

if (0<=mark<=49) then outcome = 'Fail';
    else if (50<=mark<=100) then outcome = 'Pass';
    else outcome = 'Gone';
/*****/
options pagesize=100;

proc freq;
    tables outcome*passed / norow nocol nopercnt missing;

proc freq;
    title3 'One at a time cat IVs with proc freq';
    tables (course2 sex ethnic tongue hsmiss) * outcome
        / nocol nopercnt chisq;

proc logistic descending order=internal;
    title3 'Simple logistic regression: Reproduce this';
    model passed = hsgpa;

proc catmod;
    title3 'Hsgpa Reproduce b1 = 0.2089, Wald Chisq = 76.5326';
    direct hsgpa; /* Direct means no dummy vars please */
    model passed = hsgpa / noprofile; /* Always suppress the profile
        when there are quantitative IVs */

/* The last response category is the reference (denominator)
category so the sign of the regression coefficient
is reversed, but we can live with this. */

proc catmod;
    title3 'Hsmiss by outcome';
    direct hsmiss; /* It's already a dummy variable. */
    model outcome = hsmiss;
    contrast 'HS Missing method 1' hsmiss 1;
    contrast 'HS Missing method 2' all_parms 0 0 1 0,
        all_parms 0 0 0 1;
```

/\*\*\*\*\*\* Estimate the probabilities \*\*\*\*\*/

Here is a copy of the parameter estimates from proc catmod.

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|-----------------|----------|----------------|------------|------------|
| Intercept | 1               | -1.3594  | 0.1380         | 97.05      | <.0001     |
|           | 2               | -0.8306  | 0.1132         | 53.81      | <.0001     |
| hsmiss    | 1               | 0.6663   | 0.2856         | 5.44       | 0.0196     |
|           | 2               | 1.2360   | 0.2180         | 32.14      | <.0001     |

\*\*\*\*\*/

```
proc iml;
  title3 'Estimate Probabilities using output from proc catmod';
  b01 = -1.3594; b11 = 0.6663;
  b02 = -0.8306; b12 = 1.2360;
  hsmiss = 0;
  L1 = b01 + b11*hsmiss;
  L2 = b02 + b12*hsmiss;
  denom = 1 + exp(L1) + exp(L2);
  Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
  print "No Missing HS Data:" Fail Gone Pass;
  hsmiss = 1;
  L1 = b01 + b11*hsmiss;
  L2 = b02 + b12*hsmiss;
  denom = 1 + exp(L1) + exp(L2);
  Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
  print "Yes Missing HS Data:" Fail Gone Pass;
  print " ";

proc freq;
  title3 'Hsmiss by outcome again for comparison';
  tables hsmiss * outcome / nocol nopercent;
```

$$\pi_1 = \frac{e^{L_1}}{1 + e^{L_1} + e^{L_2}}$$

$$\pi_2 = \frac{e^{L_2}}{1 + e^{L_1} + e^{L_2}}$$

$$\pi_k = \frac{1}{1 + e^{L_1} + e^{L_2}}$$

```

/* Now seek a good predictive model */

proc catmod;
  title3 'HS variables';
  direct hsgpa hscalcalc hsengl;
  model outcome = hsgpa hscalcalc hsengl / noprofile;

/* Drop HS English */

proc catmod;
  title3 'HS gpa and calc + course2 ';
  direct hsgpa hscalcalc;
  model outcome = hsgpa hscalcalc course2 / noprofile;
  /* Dummy vars for course2 use effect coding */

/* Forget course2 */

proc catmod;
  title3 'HS gpa and calc + diagnostic test';
  direct hsgpa hscalcalc precalc calc;
  model outcome = hsgpa hscalcalc precalc calc / noprofile;

/* Drop calc subtest, keep precalc */

proc catmod;
  title3 'Try gender, ethnic and mother tongue controlling for good stuff';
  direct hsgpa hscalcalc precalc calc gender mtongue;
  model outcome = hsgpa hscalcalc precalc ethnic gender mtongue / noprofile;
  contrast 'Demographics'
    ethnic 1 0 0 0 0,
    ethnic 0 1 0 0 0,
    ethnic 0 0 1 0 0,
    ethnic 0 0 0 1 0,
    ethnic 0 0 0 0 1,
    gender 1,
    mtongue 1;
  contrast 'Ethnic and Gender'
    ethnic 1 0 0 0 0,
    ethnic 0 1 0 0 0,
    ethnic 0 0 1 0 0,
    ethnic 0 0 0 1 0,
    ethnic 0 0 0 0 1,
    gender 1;

/* Got this in the log file: "WARNING: The formatted values of one or more
variables are truncated to 16" */

/* Mother tongue is significant. Still true when we drop ethnic and gender? */

proc catmod;
  title3 'hsgpa hscalcalc precalc calc mtongue';
  direct hsgpa hscalcalc precalc calc mtongue;
  model outcome = hsgpa hscalcalc precalc mtongue / noprofile;

/* Allowing for academic background, students whose first language is English
are more likely to fail the course as opposed to passing, and less likely to
disappear as opposed to passing. If this is replicated, it will be very
interesting. Now explore in more detail. */

```

```

proc catmod;
  title3 'Different coefficients for Gone and Fail?';
  direct hsgpa hscalcalc precalc calc mtongue;
  model outcome = hsgpa hscalcalc precalc calc mtongue / noprofile;
  contrast 'Diff Relationships Overall'
    all_parms 0 0 1 -1 0 0 0 0 0 0 0,
    all_parms 0 0 0 0 0 1 -1 0 0 0 0,
    all_parms 0 0 0 0 0 0 0 1 -1 0 0,
    all_parms 0 0 0 0 0 0 0 0 0 1 -1;
  contrast 'Diff Relationships for hsgpa'
    all_parms 0 0 1 -1 0 0 0 0 0 0 0;
  contrast 'Diff Relationships for hscalcalc'
    all_parms 0 0 0 0 0 1 -1 0 0 0 0;
  contrast 'Diff Relationships for precalc'
    all_parms 0 0 0 0 0 0 0 1 -1 0 0;
  contrast 'Diff Relationships for mtongue'
    all_parms 0 0 0 0 0 0 0 0 0 1 -1;

  /***** Cross-validation *****/

  %include 'readmath2.sas';
  if (0<=mark<=49) then outcome = 'Fail';
  else if (50<=mark<=100) then outcome = 'Pass';
  else outcome = 'Gone';

proc catmod data=mathrep;
  title3 'Replicate hsgpa hscalcalc precalc calc mtongue 0.05/7 = 0.00714';
  direct hsgpa hscalcalc precalc calc mtongue;
  model outcome = hsgpa hscalcalc precalc calc mtongue / noprofile;
  contrast 'Diff Relationships for mtongue'
    all_parms 0 0 0 0 0 0 0 0 0 1 -1;

```

Gender, Ethnicity and Math performance  
 Logistic regression with more than 2 resp. categories using proc catmod

1

The FREQ Procedure

Table of outcome by passed

| outcome   |     | passed(Passed the course) |  | Total |
|-----------|-----|---------------------------|--|-------|
| Frequency | No  | Yes                       |  |       |
| Fail      | 90  | 0                         |  | 90    |
| Gone      | 184 | 0                         |  | 184   |
| Pass      | 0   | 305                       |  | 305   |
| Total     | 274 | 305                       |  | 579   |

Gender, Ethnicity and Math performance  
 Logistic regression with more than 2 resp. categories using proc catmod  
 One at a time cat IVs with proc freq

2

The FREQ Procedure

Table of course2 by outcome

| course2              | outcome     |             |              | Total |
|----------------------|-------------|-------------|--------------|-------|
| Frequency<br>Row Pct | Fail        | Gone        | Pass         |       |
| 0                    | 14<br>12.96 | 59<br>54.63 | 35<br>32.41  | 108   |
| 1                    | 9<br>15.25  | 35<br>59.32 | 15<br>25.42  | 59    |
| 2                    | 61<br>16.35 | 88<br>23.59 | 224<br>60.05 | 373   |
| 3                    | 6<br>15.38  | 2<br>5.13   | 31<br>79.49  | 39    |
| Total                | 90          | 184         | 305          | 579   |

Statistics for Table of course2 by outcome

| Statistic                   | DF | Value   | Prob   |
|-----------------------------|----|---------|--------|
| Chi-Square                  | 6  | 74.9535 | <.0001 |
| Likelihood Ratio Chi-Square | 6  | 76.9104 | <.0001 |

Table of sex by outcome

| sex       | outcome     |             |              |       |
|-----------|-------------|-------------|--------------|-------|
| Frequency | Fail        | Gone        | Pass         | Total |
| Row Pct   |             |             |              |       |
| Female    | 45<br>16.92 | 73<br>27.44 | 148<br>55.64 | 266   |
| Male      | 43<br>15.09 | 95<br>33.33 | 147<br>51.58 | 285   |
| Total     | 88          | 168         | 295          | 551   |

Frequency Missing = 28

Statistics for Table of sex by outcome

| Statistic                   | DF | Value  | Prob   |
|-----------------------------|----|--------|--------|
| Chi-Square                  | 2  | 2.2773 | 0.3202 |
| Likelihood Ratio Chi-Square | 2  | 2.2828 | 0.3194 |
| Mantel-Haenszel Chi-Square  | 1  | 0.1233 | 0.7254 |
| Phi Coefficient             |    | 0.0643 |        |
| Contingency Coefficient     |    | 0.0642 |        |
| Cramer's V                  |    | 0.0643 |        |

Effective Sample Size = 551

Frequency Missing = 28

The FREQ Procedure

Table of ethnic by outcome

| ethnic(Judged Nationality of name) |             | outcome     |              |  |       |
|------------------------------------|-------------|-------------|--------------|--|-------|
| Frequency                          |             |             |              |  |       |
| Row Pct                            | Fail        | Gone        | Pass         |  | Total |
| Asian                              | 21<br>16.03 | 44<br>33.59 | 66<br>50.38  |  | 131   |
| Eastern European                   | 13<br>20.63 | 17<br>26.98 | 33<br>52.38  |  | 63    |
| European not Eastern               | 35<br>17.95 | 53<br>27.18 | 107<br>54.87 |  | 195   |
| Middle-Eastern and Pakistani       | 11<br>15.28 | 22<br>30.56 | 39<br>54.17  |  | 72    |
| East Indian                        | 6<br>7.69   | 25<br>32.05 | 47<br>60.26  |  | 78    |
| Other and DK                       | 4<br>10.00  | 23<br>57.50 | 13<br>32.50  |  | 40    |
| Total                              | 90          | 184         | 305          |  | 579   |

Statistics for Table of ethnic by outcome

| Statistic                   | DF | Value   | Prob   |
|-----------------------------|----|---------|--------|
| Chi-Square                  | 10 | 20.2180 | 0.0273 |
| Likelihood Ratio Chi-Square | 10 | 19.8317 | 0.0309 |
| Mantel-Haenszel Chi-Square  | 1  | 0.4698  | 0.4931 |
| Phi Coefficient             |    | 0.1869  |        |
| Contingency Coefficient     |    | 0.1837  |        |
| Cramer's V                  |    | 0.1321  |        |

Sample Size = 579

Table of tongue by outcome

tongue(Mother Tongue (Eng or Other))  
outcome

| Frequency<br>Row Pct | Fail        | Gone         | Pass         | Total |
|----------------------|-------------|--------------|--------------|-------|
| English              | 74<br>18.41 | 113<br>28.11 | 215<br>53.48 | 402   |
| Other                | 14<br>9.40  | 55<br>36.91  | 80<br>53.69  | 149   |
| Total                | 88          | 168          | 295          | 551   |

Frequency Missing = 28

Statistics for Table of tongue by outcome

| Statistic                   | DF | Value  | Prob   |
|-----------------------------|----|--------|--------|
| Chi-Square                  | 2  | 8.2920 | 0.0158 |
| Likelihood Ratio Chi-Square | 2  | 8.8198 | 0.0122 |
| Mantel-Haenszel Chi-Square  | 1  | 1.6654 | 0.1969 |
| Phi Coefficient             |    | 0.1227 |        |
| Contingency Coefficient     |    | 0.1218 |        |
| Cramer's V                  |    | 0.1227 |        |

Effective Sample Size = 551

Frequency Missing = 28



Table of hsmis by outcome

hsmis(Missing Any High School Data)  
outcome

| Frequency<br>Row Pct | Fail        | Gone         | Pass         | Total |
|----------------------|-------------|--------------|--------------|-------|
| No                   | 66<br>15.17 | 112<br>25.75 | 257<br>59.08 | 435   |
| Yes                  | 24<br>16.67 | 72<br>50.00  | 48<br>33.33  | 144   |
| Total                | 90          | 184          | 305          | 579   |

Gender, Ethnicity and Math performance  
Logistic regression with more than 2 resp. categories using proc catmod  
One at a time cat IVs with proc freq

4

The FREQ Procedure

Statistics for Table of hsmis by outcome

| Statistic                   | DF | Value   | Prob   |
|-----------------------------|----|---------|--------|
| Chi-Square                  | 2  | 33.7946 | <.0001 |
| Likelihood Ratio Chi-Square | 2  | 33.3038 | <.0001 |
| Mantel-Haenszel Chi-Square  | 1  | 14.7239 | 0.0001 |
| Phi Coefficient             |    | 0.2416  |        |
| Contingency Coefficient     |    | 0.2348  |        |
| Cramer's V                  |    | 0.2416  |        |

Sample Size = 579

Simple logistic regression: Reproduce this

Skipping most of the output ...

Testing Global Null Hypothesis: BETA=0

| Test             | Chi-Square | DF | Pr > ChiSq |
|------------------|------------|----|------------|
| Likelihood Ratio | 113.5689   | 1  | <.0001     |
| Score            | 97.2566    | 1  | <.0001     |
| Wald             | 76.5326    | 1  | <.0001     |

Analysis of Maximum Likelihood Estimates

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|-----------|----|----------|----------------|-----------------|------------|
| Intercept | 1  | -16.1468 | 1.8664         | 74.8415         | <.0001     |
| hsgpa     | 1  | 0.2089   | 0.0239         | 76.5326         | <.0001     |

---

Logistic regression with more than 2 resp. categories using proc catmod  
Hsgpa Reproduce b1 = 0.2089, Wald Chisq = 76.5326

The CATMOD Procedure

Data Summary

|                   |        |                 |     |
|-------------------|--------|-----------------|-----|
| Response          | passed | Response Levels | 2   |
| Weight Variable   | None   | Populations     | 137 |
| Data Set          | MATH   | Total Frequency | 466 |
| Frequency Missing | 113    | Observations    | 466 |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| Intercept        | 1   | 74.84      | <.0001     |
| hsgpa            | 1   | 76.53      | <.0001     |
| Likelihood Ratio | 135 | 110.99     | 0.9353     |

Analysis of Maximum Likelihood Estimates

| Parameter | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|----------|----------------|------------|------------|
| Intercept | 16.1468  | 1.8664         | 74.84      | <.0001     |
| hsgpa     | -0.2089  | 0.0239         | 76.53      | <.0001     |

---

Gender, Ethnicity and Math performance  
Logistic regression with more than 2 resp. categories using proc catmod  
Hsmis by outcome

7

The CATMOD Procedure

Data Summary

|                   |         |                 |     |
|-------------------|---------|-----------------|-----|
| Response          | outcome | Response Levels | 3   |
| Weight Variable   | None    | Populations     | 2   |
| Data Set          | MATH    | Total Frequency | 579 |
| Frequency Missing | 0       | Observations    | 579 |

Population Profiles

| Sample | hsmis | Sample Size |
|--------|-------|-------------|
| 1      | No    | 435         |
| 2      | Yes   | 144         |

Response Profiles

| Response | outcome |
|----------|---------|
| 1        | Fail    |
| 2        | Gone    |
| 3        | Pass    |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF | Chi-Square | Pr > ChiSq |
|------------------|----|------------|------------|
| Intercept        | 2  | 122.46     | <.0001     |
| hsmiss           | 2  | 32.14      | <.0001     |
| Likelihood Ratio | 0  | .          | .          |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|-----------------|----------|----------------|------------|------------|
| Intercept | 1               | -1.3594  | 0.1380         | 97.05      | <.0001     |
|           | 2               | -0.8306  | 0.1132         | 53.81      | <.0001     |
| hsmiss    | 1               | 0.6663   | 0.2856         | 5.44       | 0.0196     |
|           | 2               | 1.2360   | 0.2180         | 32.14      | <.0001     |

Contrasts of Maximum Likelihood Estimates

| Contrast            | DF | Chi-Square | Pr > ChiSq |
|---------------------|----|------------|------------|
| HS Missing method 1 | 2  | 32.14      | <.0001     |
| HS Missing method 2 | 2  | 32.14      | <.0001     |

---

8

Gender, Ethnicity and Math performance  
 Logistic regression with more than 2 resp. categories using proc catmod  
 Estimate Probabilities using output from proc catmod

Fail            Gone            Pass

No Missing HS Data: 0.1517278 0.2574661 0.5908062

Fail            Gone            Pass

Yes Missing HS Data: 0.1666786 0.4999798 0.3333416

---

9

Gender, Ethnicity and Math performance  
 Logistic regression with more than 2 resp. categories using proc catmod  
 Hsmisss by outcome again for comparison

The FREQ Procedure

Table of hsmisss by outcome

hsmisss(Missing Any High School Data)  
 outcome

| Frequency<br>Row Pct | Fail        | Gone         | Pass         | Total |
|----------------------|-------------|--------------|--------------|-------|
| No                   | 66<br>15.17 | 112<br>25.75 | 257<br>59.08 | 435   |
| Yes                  | 24<br>16.67 | 72<br>50.00  | 48<br>33.33  | 144   |
| Total                | 90          | 184          | 305          | 579   |

The CATMOD Procedure

Data Summary

|                   |         |                 |     |
|-------------------|---------|-----------------|-----|
| Response          | outcome | Response Levels | 3   |
| Weight Variable   | None    | Populations     | 434 |
| Data Set          | MATH    | Total Frequency | 435 |
| Frequency Missing | 144     | Observations    | 435 |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| Intercept        | 2   | 56.62      | <.0001     |
| hsgpa            | 2   | 20.48      | <.0001     |
| hscal            | 2   | 22.20      | <.0001     |
| hsengl           | 2   | 0.88       | 0.6454     |
| Likelihood Ratio | 860 | 690.18     | 1.0000     |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|-----------------|----------|----------------|------------|------------|
| Intercept | 1               | 13.3872  | 2.5867         | 26.78      | <.0001     |
|           | 2               | 16.5898  | 2.3505         | 49.82      | <.0001     |
| hsgpa     | 1               | -0.1487  | 0.0425         | 12.23      | 0.0005     |
|           | 2               | -0.1502  | 0.0375         | 16.03      | <.0001     |
| hscal     | 1               | -0.0485  | 0.0161         | 9.04       | 0.0026     |
|           | 2               | -0.0643  | 0.0141         | 20.77      | <.0001     |
| hsengl    | 1               | 0.00782  | 0.0214         | 0.13       | 0.7152     |
|           | 2               | -0.0119  | 0.0184         | 0.42       | 0.5171     |

---

The CATMOD Procedure

Data Summary

|                   |         |                 |     |
|-------------------|---------|-----------------|-----|
| Response          | outcome | Response Levels | 3   |
| Weight Variable   | None    | Populations     | 418 |
| Data Set          | MATH    | Total Frequency | 437 |
| Frequency Missing | 142     | Observations    | 437 |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| Intercept        | 2   | 52.28      | <.0001     |
| hsgpa            | 2   | 28.90      | <.0001     |
| hscal            | 2   | 20.87      | <.0001     |
| course2          | 6   | 8.13       | 0.2285     |
| Likelihood Ratio | 824 | 656.57     | 1.0000     |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|-----------------|----------|----------------|------------|------------|
| Intercept | 1               | 13.7838  | 2.6169         | 27.74      | <.0001     |
|           | 2               | 15.6840  | 2.3742         | 43.64      | <.0001     |
| hsgpa     | 1               | -0.1441  | 0.0366         | 15.46      | <.0001     |
|           | 2               | -0.1583  | 0.0327         | 23.36      | <.0001     |
| hscal     | 1               | -0.0515  | 0.0156         | 10.89      | 0.0010     |
|           | 2               | -0.0580  | 0.0138         | 17.70      | <.0001     |
| course2   | 0 1             | 0.0131   | 0.3750         | 0.00       | 0.9721     |
|           | 0 2             | 0.7556   | 0.3306         | 5.22       | 0.0223     |
|           | 1 1             | -0.1934  | 0.5388         | 0.13       | 0.7197     |
|           | 1 2             | 0.2597   | 0.4682         | 0.31       | 0.5791     |
|           | 2 1             | 0.0649   | 0.2712         | 0.06       | 0.8110     |
|           | 2 2             | 0.1083   | 0.2740         | 0.16       | 0.6928     |

The CATMOD Procedure

Data Summary

|                   |         |                 |     |
|-------------------|---------|-----------------|-----|
| Response          | outcome | Response Levels | 3   |
| Weight Variable   | None    | Populations     | 374 |
| Data Set          | MATH    | Total Frequency | 375 |
| Frequency Missing | 204     | Observations    | 375 |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| Intercept        | 2   | 37.36      | <.0001     |
| hsgpa            | 2   | 14.57      | 0.0007     |
| hscal            | 2   | 19.65      | <.0001     |
| precalc          | 2   | 8.76       | 0.0126     |
| calc             | 2   | 3.57       | 0.1678     |
| Likelihood Ratio | 738 | 548.17     | 1.0000     |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|-----------------|----------|----------------|------------|------------|
| Intercept | 1               | 13.2650  | 2.7880         | 22.64      | <.0001     |
|           | 2               | 14.4536  | 2.6768         | 29.16      | <.0001     |
| hsgpa     | 1               | -0.1247  | 0.0388         | 10.33      | 0.0013     |
|           | 2               | -0.1133  | 0.0365         | 9.64       | 0.0019     |
| hscal     | 1               | -0.0492  | 0.0170         | 8.33       | 0.0039     |
|           | 2               | -0.0664  | 0.0156         | 18.04      | <.0001     |
| precalc   | 1               | -0.2425  | 0.1118         | 4.70       | 0.0301     |
|           | 2               | -0.2782  | 0.1049         | 7.03       | 0.0080     |
| calc      | 1               | -0.0113  | 0.0804         | 0.02       | 0.8886     |
|           | 2               | -0.1469  | 0.0802         | 3.35       | 0.0670     |



Gender, Ethnicity and Math performance 13  
 Logistic regression with more than 2 resp. categories using proc catmod  
 Try gender, ethnic and mother tongue controlling for good stuff

The CATMOD Procedure

Data Summary

|                   |         |                 |     |
|-------------------|---------|-----------------|-----|
| Response          | outcome | Response Levels | 3   |
| Weight Variable   | None    | Populations     | 370 |
| Data Set          | MATH    | Total Frequency | 370 |
| Frequency Missing | 209     | Observations    | 370 |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| Intercept        | 2   | 30.46      | <.0001     |
| hsgpa            | 2   | 10.43      | 0.0054     |
| hscal            | 2   | 25.99      | <.0001     |
| precalc          | 2   | 15.31      | 0.0005     |
| ethnic           | 10  | 8.71       | 0.5594     |
| gender           | 2   | 0.76       | 0.6851     |
| mtongue          | 2   | 13.41      | 0.0012     |
| Likelihood Ratio | 718 | 512.06     | 1.0000     |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number  | Estimate | Standard Error | Chi-Square | Pr > ChiSq |        |
|-----------|------------------|----------|----------------|------------|------------|--------|
| Intercept | 1                | 10.8891  | 3.0227         | 12.98      | 0.0003     |        |
|           | 2                | 14.4497  | 2.8149         | 26.35      | <.0001     |        |
| hsgpa     | 1                | -0.1176  | 0.0416         | 7.97       | 0.0047     |        |
|           | 2                | -0.0924  | 0.0387         | 5.69       | 0.0170     |        |
| hscal     | 1                | -0.0500  | 0.0177         | 7.98       | 0.0047     |        |
|           | 2                | -0.0816  | 0.0163         | 24.91      | <.0001     |        |
| precalc   | 1                | -0.2555  | 0.1096         | 5.44       | 0.0197     |        |
|           | 2                | -0.3940  | 0.1056         | 13.91      | 0.0002     |        |
| ethnic    | Asian            | 1        | 0.4901         | 0.3799     | 1.66       | 0.1970 |
|           | Asian            | 2        | 0.3606         | 0.3744     | 0.93       | 0.3354 |
|           | Eastern European | 1        | 0.2206         | 0.4267     | 0.27       | 0.6052 |
|           | Eastern European | 2        | -0.0339        | 0.4417     | 0.01       | 0.9388 |
|           | European not Eas | 1        | -0.2127        | 0.3132     | 0.46       | 0.4971 |
|           | European not Eas | 2        | -0.2316        | 0.3339     | 0.48       | 0.4880 |
|           | Middle-Eastern   | 1        | 0.5776         | 0.4495     | 1.65       | 0.1988 |
|           | Middle-Eastern   | 2        | 0.4759         | 0.4455     | 1.14       | 0.2854 |
|           | East Indian      | 1        | -0.7675        | 0.5653     | 1.84       | 0.1745 |
|           | East Indian      | 2        | 0.4496         | 0.4094     | 1.21       | 0.2721 |
| gender    | 1                | -0.1684  | 0.3421         | 0.24       | 0.6227     |        |
|           | 2                | -0.2690  | 0.3188         | 0.71       | 0.3988     |        |
| mtongue   | 1                | 2.1901   | 0.7748         | 7.99       | 0.0047     |        |
|           | 2                | -0.6472  | 0.3836         | 2.85       | 0.0916     |        |

Contrasts of Maximum Likelihood Estimates

| Contrast          | DF | Chi-Square | Pr > ChiSq |
|-------------------|----|------------|------------|
| Demographics      | 14 | 25.48      | 0.0301     |
| Ethnic and Gender | 12 | 10.07      | 0.6101     |

Got this in the log file: "WARNING: The formatted values of one or more variables are truncated to 16"

The CATMOD Procedure

Data Summary

|                   |         |                 |     |
|-------------------|---------|-----------------|-----|
| Response          | outcome | Response Levels | 3   |
| Weight Variable   | None    | Populations     | 368 |
| Data Set          | MATH    | Total Frequency | 370 |
| Frequency Missing | 209     | Observations    | 370 |

Maximum Likelihood Analysis

Maximum likelihood computations converged.

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq |
|------------------|-----|------------|------------|
| Intercept        | 2   | 34.45      | <.0001     |
| hsgpa            | 2   | 14.43      | 0.0007     |
| hscalc           | 2   | 25.02      | <.0001     |
| precalc          | 2   | 14.16      | 0.0008     |
| mtongue          | 2   | 16.00      | 0.0003     |
| Likelihood Ratio | 726 | 517.41     | 1.0000     |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq |
|-----------|-----------------|----------|----------------|------------|------------|
| Intercept | 1               | 12.0789  | 2.9530         | 16.73      | <.0001     |
|           | 2               | 14.5402  | 2.7213         | 28.55      | <.0001     |
| hsgpa     | 1               | -0.1385  | 0.0403         | 11.83      | 0.0006     |
|           | 2               | -0.0975  | 0.0370         | 6.93       | 0.0085     |
| hscalc    | 1               | -0.0430  | 0.0169         | 6.46       | 0.0110     |
|           | 2               | -0.0772  | 0.0156         | 24.43      | <.0001     |
| precalc   | 1               | -0.2372  | 0.1068         | 4.93       | 0.0263     |
|           | 2               | -0.3679  | 0.1025         | 12.89      | 0.0003     |
| mtongue   | 1               | 1.9594   | 0.7554         | 6.73       | 0.0095     |
|           | 2               | -0.8614  | 0.3523         | 5.98       | 0.0145     |

Very interesting ...

Different coefficients for Gone and Fail?

Skipping all but the very last part of the output ...

Contrasts of Maximum Likelihood Estimates

| Contrast                       | DF | Chi-Square | Pr > ChiSq |
|--------------------------------|----|------------|------------|
| Diff Relationships Overall     | 4  | 17.06      | 0.0019     |
| Diff Relationships for hsgpa   | 1  | 0.83       | 0.3630     |
| Diff Relationships for hscal   | 1  | 3.50       | 0.0612     |
| Diff Relationships for precalc | 1  | 1.15       | 0.2842     |
| Diff Relationships for mtongue | 1  | 13.60      | 0.0002     |

Gender, Ethnicity and Math performance

16

Replicate hsgpa hscal precalc calc mtongue 0.05/7 = 0.00714

Maximum Likelihood Analysis

Maximum Likelihood Analysis of Variance

| Source           | DF  | Chi-Square | Pr > ChiSq    |
|------------------|-----|------------|---------------|
| Intercept        | 2   | 40.43      | <.0001        |
| hsgpa            | 2   | 17.66      | <b>0.0001</b> |
| hscal            | 2   | 16.72      | <b>0.0002</b> |
| precalc          | 2   | 9.69       | <b>0.0079</b> |
| mtongue          | 2   | 0.70       | <b>0.7051</b> |
| Likelihood Ratio | 744 | 619.87     | 0.9997        |

Analysis of Maximum Likelihood Estimates

| Parameter | Function Number | Estimate | Standard Error | Chi-Square | Pr > ChiSq    |
|-----------|-----------------|----------|----------------|------------|---------------|
| Intercept | 1               | 12.8640  | 2.5073         | 26.32      | <.0001        |
|           | 2               | 12.3389  | 2.2829         | 29.21      | <.0001        |
| hsgpa     | 1               | -0.1391  | 0.0364         | 14.60      | 0.0001        |
|           | 2               | -0.0980  | 0.0324         | 9.13       | 0.0025        |
| hscal     | 1               | -0.0317  | 0.0167         | 3.61       | 0.0575        |
|           | 2               | -0.0609  | 0.0149         | 16.67      | <.0001        |
| precalc   | 1               | -0.2068  | 0.1005         | 4.23       | 0.0397        |
|           | 2               | -0.2615  | 0.0914         | 8.18       | 0.0042        |
| mtongue   | 1               | 0.2323   | 0.3582         | 0.42       | <b>0.5166</b> |
|           | 2               | 0.2330   | 0.3265         | 0.51       | <b>0.4753</b> |

Contrasts of Maximum Likelihood Estimates

| Contrast                       | DF | Chi-Square | Pr > ChiSq    |
|--------------------------------|----|------------|---------------|
| Diff Relationships for mtongue | 1  | 0.00       | <b>0.9986</b> |