Logistic regression with multinomial outcome

```
/* mathtest.sas  */
%include 'readmath.sas';
title2 'Test logistic regression with more than 2 resp. categories';
options pagesize=100;
/*************** Data step continues ***********************/
if sex = 'Female' then gender=1; else if sex = 'Male' then gender=0;
if (0<=mark<=49) then outcome = 'Fail';
   else if (50<=mark<=100) then outcome = 'Pass';
   else outcome = 'Gone';
/*************************************************************/

proc freq;
     title3 'Sex by Outcome with proc freq';
     tables sex * outcome  / nocol nopercent chisq;

proc catmod;
     title3 'Sex by outcome with proc catmod';
     direct gender; /* Already a dummy var */
     model outcome = gender;

proc logistic; /* Matches proc catmod and proc freq */
      title3 'Sex by Outcome with proc logistic';
      class sex / param=ref ref=last ;
      model outcome (ref='Pass') = sex / link=glogit; /* Generalized logits */
      /* ref='Pass' is redundant because it's alphabetically the last anyway */
      Gender: test sexFemale_Fail = sexFemale_Gone = 0;
      /* Ho: beta_11 = beta_12 = 0 */

proc logistic;
      title3 'Sex by Outcome with proc logistic: Simpler code';
      class sex / param=ref;
      model outcome = sex / link=glogit;
      Gender: test sexFemale_Fail = sexFemale_Gone = 0;

proc iml;
     title3 'Reproduce Conditional Probs using output from proc logistic';
      b01 = -1.2292; b11 = 0.0387;
      b02 = -0.4366; b12 = -0.2702;
      sex = 1; /* Females */
      L1 = b01 + b11*sex;
      L2 = b02 + b12*sex;
      denom = 1 + exp(L1) + exp(L2);
      Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
      print "Females:" Fail Gone Pass;
      sex = 0; /* Males */
      L1 = b01 + b11*sex;
      L2 = b02 + b12*sex;
      denom = 1 + exp(L1) + exp(L2);
      Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
      print "Males:" Fail Gone Pass;
      print "  ";

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

The FREQ Procedure

Table of sex by outcome

sex / outcome Frequency Row Pct	Fail	Gone	Pass	Total
Female	45 16.92	73 27.44	148 55.64	266
Male	43 15.09	95 33.33	147 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

Skipping `proc catmod` output ...

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH
Response Variable	outcome
Number of Response Levels	3
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	579
Number of Observations Used	551

Response Profile

Ordered Value	outcome	Total Frequency
1	Fail	88
2	Gone	168
3	Pass	295

Logits modeled use outcome='Pass' as the reference category.

NOTE: 28 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables
sex	Female	1
	Male	0

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	1094.553	1096.270
SC	1103.177	1113.517
-2 Log L	1090.553	1088.270

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2.2828	2	0.3194
Score	2.2773	2	0.3202
Wald	2.2718	2	0.3211

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
sex	2	2.2718	0.3211

Analysis of Maximum Likelihood Estimates

Parameter		outcome	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		Fail	1	-1.2292	0.1734	50.2690	<.0001
Intercept		Gone	1	-0.4366	0.1316	10.9977	0.0009
sex	Female	Fail	1	0.0387	0.2430	0.0253	0.8735
sex	Female	Gone	1	-0.2702	0.1944	1.9322	0.1645

Odds Ratio Estimates

Effect	outcome	Point Estimate	95% Wald Confidence Limits	
sex Female vs Male	Fail	1.039	0.646	1.673
sex Female vs Male	Gone	0.763	0.521	1.117

The LOGISTIC Procedure

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
Gender	2.2718	2	0.3211

Skipping output from simpler code ...

	Fail	Gone	Pass
Females:	0.1691821	0.2744244	0.5563935

	Fail	Gone	Pass
Males:	0.1508836	0.3333219	0.5157946

The FREQ Procedure

Table of sex by outcome

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

```
/* multimathlogreg.sas  */
%include 'readmath.sas';
title2 'Logistic regression with multinomial outcome';
options pagesize=100;

proc logistic;
      title3 'Full model (not really)';
      class sex / param=ref;
      model outcome = hsgpa hscalc hsengl sex / link=glogit;
      Gender_and_English: test sexFemale_Fail = sexFemale_Gone =
                                 hsengl_Fail = hsengl_Gone = 0;

proc logistic;
      title3 'Smaller model for generating predictions';
      model outcome = hsgpa hscalc / link=glogit;

proc means; var hsgpa hscalc;

proc iml;
     title3 'Estimate Some Probabilities';
      b01 = 13.5876; b11 = -0.1420; b21 = -0.0503;
      b02 = 16.5405; b12 = -0.1638; b22 = -0.0615;
      print "-------------------------------------------";
      print "A great student: hsgpa=99, hscalc=100";
      hsgpa=99; hscalc=100;
      L1 = b01 + b11*hsgpa + b21*hscalc;
      L2 = b02 + b12*hsgpa + b22*hscalc;
      denom = 1 + exp(L1) + exp(L2);
      Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
      print "Probability " Fail Gone Pass;
      print "-------------------------------------------";
      print "An average student: hsgpa=80, hscalc=75";
      hsgpa=80; hscalc=75;
      L1 = b01 + b11*hsgpa + b21*hscalc;
      L2 = b02 + b12*hsgpa + b22*hscalc;
      denom = 1 + exp(L1) + exp(L2);
      Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
      print "Probability " Fail Gone Pass;
      print "-------------------------------------------";
      print "A poor student: hsgpa=65, hscalc=50";
      hsgpa=65; hscalc=50;
      L1 = b01 + b11*hsgpa + b21*hscalc;
      L2 = b02 + b12*hsgpa + b22*hscalc;
      denom = 1 + exp(L1) + exp(L2);
      Fail = exp(L1)/denom; Gone = exp(L2)/denom; Pass = 1/denom;
      print "Probability " Fail Gone Pass;
      print "-------------------------------------------";

proc freq;
     title3 'Overall Percentages for comparison';
     tables outcome;
```

Logistic regression with multinomial outcome
Full model (not really)

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH
Response Variable	outcome
Number of Response Levels	3
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	579
Number of Observations Used	428

Response Profile

Ordered Value	outcome	Total Frequency
1	Fail	66
2	Gone	107
3	Pass	255

Logits modeled use outcome='Pass' as the reference category.

NOTE: 151 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables
sex	Female	1
	Male	0

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	811.545	699.050
SC	819.663	739.641
-2 Log L	807.545	679.050

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	128.4957	8	<.0001
Score	108.8063	8	<.0001
Wald	82.5853	8	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
hsgpa	2	20.3799	<.0001
hscalc	2	20.7198	<.0001
hsengl	2	0.2063	0.9020
sex	2	1.2613	0.5323

Analysis of Maximum Likelihood Estimates

Parameter		outcome	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		Fail	1	13.4964	2.5937	27.0768	<.0001
Intercept		Gone	1	16.0871	2.3580	46.5450	<.0001
hsgpa		Fail	1	-0.1483	0.0425	12.1771	0.0005
hsgpa		Gone	1	-0.1503	0.0377	15.9302	<.0001
hscalc		Fail	1	-0.0471	0.0161	8.5401	0.0035
hscalc		Gone	1	-0.0622	0.0142	19.2642	<.0001
hsengl		Fail	1	0.00391	0.0221	0.0314	0.8594
hsengl		Gone	1	-0.00606	0.0192	0.0997	0.7522
sex	Female	Fail	1	0.1200	0.3079	0.1518	0.6968
sex	Female	Gone	1	-0.2263	0.2721	0.6917	0.4056

Gender, Ethnicity and Math performance 2
Logistic regression with multinomial outcome
Full model (not really)

The LOGISTIC Procedure

Odds Ratio Estimates

Effect		outcome	Point Estimate	95% Wald Confidence Limits	
hsgpa		Fail	0.862	0.793	0.937
hsgpa		Gone	0.860	0.799	0.926
hscalc		Fail	0.954	0.924	0.985
hscalc		Gone	0.940	0.914	0.966
hsengl		Fail	1.004	0.961	1.048
hsengl		Gone	0.994	0.957	1.032
sex	Female vs Male	Fail	1.127	0.617	2.062
sex	Female vs Male	Gone	0.797	0.468	1.359

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
Gender_and_English	1.8075	4	0.7711

Logistic regression with multinomial outcome
Smaller model for generating predictions

Skipping ...

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
hsgpa	2	31.0740	<.0001
hscalc	2	24.0423	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	outcome	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	Fail	1	13.5876	2.5787	27.7645	<.0001
Intercept	Gone	1	16.5405	2.3389	50.0098	<.0001
hsgpa	Fail	1	-0.1420	0.0360	15.5757	<.0001
hsgpa	Gone	1	-0.1638	0.0320	26.1264	<.0001
hscalc	Fail	1	-0.0503	0.0151	11.1151	0.0009
hscalc	Gone	1	-0.0615	0.0132	21.5931	<.0001

Skipping ...

A great student: hsgpa=99, hscalc=100

	Fail	Gone	Pass
Probability	0.0040566	0.0029301	0.9930133

An average student: hsgpa=80, hscalc=75

	Fail	Gone	Pass
Probability	0.1401808	0.2027214	0.6570978

A poor student: hsgpa=65, hscalc=50

	Fail	Gone	Pass
Probability	0.2623323	0.696113	0.0415548

The FREQ Procedure

outcome	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Fail	90	15.54	90	15.54
Gone	184	31.78	274	47.32
Pass	305	52.68	579	100.00