Logistic Regression with Dummy Variables

```
/* mathlogreg1.sas  */
%include 'readmath.sas'; options pagesize=100;
title2 'Logistic Regression with dummy variables';

/* Recall definition of passed
     if (50<=mark<=100) then passed=1; else passed=0;    */

if course=. then course2=0;
   else if course=4 then course2=0; /* 0=No Data */
   else course2=course;
/* Dummy vars for course2: Course 3 (high end) is reference category  */
if course2=. then c0=.; else if course2=0 then c0=1; else c0=0;
if course2=. then c1=.; else if course2=1 then c1=1; else c1=0;
if course2=. then c2=.; else if course2=2 then c2=1; else c2=0;

hsutil = hsgpa+hscalc+hsengl;
if hsutil = . then hsmiss=1; else hsmiss=0;
label hsmiss = 'Missing Any High School Data'; format hsmiss ynfmt.;

proc freq;
     title3 'Check course2 and dummy vars -- and why so many no course?';
     tables (course c0-c2 totscore hsmiss) * course2
            / norow nocol nopercent missing;

proc freq;
     title3 'A few simple Chisquare tests to predict passed';
     tables (course2 sex ethnic tongue hsmiss) * passed / nocol nopercent chisq;

proc logistic descending order=internal; /* To model Y=1 */
     title3 'Course2 by passed with dummy vars';
     model passed = c0 c1 c2;
     Course0_vs_3: test c0=0;

proc iml;
     title3 'Estimate prob. of passing for course=2: Compare 224/373=0.6005';
     b0 = 1.3545; b1 = -2.0896; b2 = -2.4307; b3 = -0.9468;
     c0 = 0; c1=0; c2=1;
     lcombo = b0 + b1*c0 + b2*c1 + b3*c2;
     probpass = exp(lcombo) / (1+exp(lcombo));
     print "Estimated probability of passing course 2 is " probpass;
     print "  ";

proc logistic descending order=internal;
     title3 'Use the Class statement';
     class course2 / param=ref;  /* This param option makes last category
                                    (course=3) the reference category */
     model passed = course2;
     contrast 'Course 0 vs 3' course2  1 0 0 ;

/* Contrast is a little tricky in proc logistic. It lets you specify a
   set of linear combinations (not necessarily contrasts) to test on the
   regression coefficients of the dummy variables for a categorical
   independent variable. It is essential to know exactly what the dummy
   variable coding scheme is. This can still be more convenient than
   defining your own dummy variables in the data step. */
```

Logistic Regression with dummy variables

Check course2 and dummy vars -- and why so many no course?

The FREQ Procedure

Table of course by course2

course	course2				
Frequency	0	1	2	3	Total
.	99	0	0	0	99
1	0	59	0	0	59
2	0	0	373	0	373
3	0	0	0	39	39
No Resp	9	0	0	0	9
Total	108	59	373	39	579

Table of c0 by course2

c0	course2				
Frequency	0	1	2	3	Total
0	0	59	373	39	471
1	108	0	0	0	108
Total	108	59	373	39	579

Table of c1 by course2

c1	course2				
Frequency	0	1	2	3	Total
0	108	0	373	39	520
1	0	59	0	0	59
Total	108	59	373	39	579

etc.

Table of totscore by course2

totscore(Total # right on diagnostic test)
course2

Frequency	0	1	2	3	Total
.	99	0	0	0	99
0	1	2	0	0	3
1	0	1	2	0	3
2	2	8	10	1	21
3	0	12	17	1	30
4	0	13	23	0	36
5	2	6	41	1	50
6	0	6	45	7	58
7	0	5	47	1	53
8	2	2	41	7	52
9	0	1	29	1	31
10	0	2	32	5	39
11	0	0	23	2	25
12	0	0	22	1	23
13	0	1	11	2	14
14	0	0	16	4	20
15	1	0	6	3	10
16	0	0	3	2	5
17	1	0	1	0	2
18	0	0	1	0	1
19	0	0	1	1	2
20	0	0	2	0	2
Total	108	59	373	39	579

Table of hsmiss by course2

hsmiss(Missing Any High School Data)	course2				
Frequency	0	1	2	3	Total
No	62	21	322	30	435
Yes	46	38	51	9	144
Total	108	59	373	39	579

The FREQ Procedure

Table of course2 by passed

course2	passed(Passed the course)	
Frequency / Row Pct	No	Yes	Total
0	73	35	108
	67.59	32.41	
1	44	15	59
	74.58	25.42	
2	149	224	373
	39.95	60.05	
3	8	31	39
	20.51	79.49	
Total	274	305	579

Statistics for Table of course2 by passed

Statistic	DF	Value	Prob
Chi-Square	3	54.7664	<.0001
Likelihood Ratio Chi-Square	3	56.5621	<.0001
Mantel-Haenszel Chi-Square	1	45.6447	<.0001
Phi Coefficient		0.3076	
Contingency Coefficient		0.2940	
Cramer's V		0.3076	

Sample Size = 579

Table of sex by passed

sex passed(Passed the course)

Frequency / Row Pct	No	Yes	Total
Female	118	148	266
	44.36	55.64	
Male	138	147	285
	48.42	51.58	
Total	256	295	551

Frequency Missing = 28

Statistics for Table of sex by passed

Statistic	DF	Value	Prob
Chi-Square	1	0.9118	0.3396
Likelihood Ratio Chi-Square	1	0.9122	0.3395
Continuity Adj. Chi-Square	1	0.7559	0.3846
Mantel-Haenszel Chi-Square	1	0.9101	0.3401
Phi Coefficient		-0.0407	
Contingency Coefficient		0.0406	
Cramer's V		-0.0407	

Fisher's Exact Test

Cell (1,1) Frequency (F)	118
Left-sided Pr <= F	0.1923
Right-sided Pr >= F	0.8509
Table Probability (P)	0.0432
Two-sided Pr <= P	0.3484

Effective Sample Size = 551
Frequency Missing = 28

The FREQ Procedure

Table of ethnic by passed

ethnic(Judged Nationality of name)
passed(Passed the course)

Frequency Row Pct	No	Yes	Total
Asian	65 49.62	66 50.38	131
Eastern European	30 47.62	33 52.38	63
European not Eastern	88 45.13	107 54.87	195
Middle-Eastern and Pakistani	33 45.83	39 54.17	72
East Indian	31 39.74	47 60.26	78
Other and DK	27 67.50	13 32.50	40
Total	274	305	579

Statistics for Table of ethnic by passed

Statistic	DF	Value	Prob
Chi-Square	5	9.0500	0.1071
Likelihood Ratio Chi-Square	5	9.1556	0.1030
Mantel-Haenszel Chi-Square	1	0.0788	0.7789
Phi Coefficient		0.1250	
Contingency Coefficient		0.1241	
Cramer's V		0.1250	

Sample Size = 579

Table of tongue by passed

tongue(Mother Tongue (Eng or Other))
passed(Passed the course)

Frequency Row Pct	No	Yes	Total
English	187 46.52	215 53.48	402
Other	69 46.31	80 53.69	149
Total	256	295	551

Frequency Missing = 28

Statistics for Table of tongue by passed

Statistic	DF	Value	Prob
Chi-Square	1	0.0019	0.9652
Likelihood Ratio Chi-Square	1	0.0019	0.9652
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.0019	0.9652
Phi Coefficient		0.0019	
Contingency Coefficient		0.0019	
Cramer's V		0.0019	

Fisher's Exact Test

Cell (1,1) Frequency (F)	187
Left-sided Pr <= F	0.5552
Right-sided Pr >= F	0.5214
Table Probability (P)	0.0765
Two-sided Pr <= P	1.0000

Effective Sample Size = 551
Frequency Missing = 28

The FREQ Procedure

Table of hsmiss by passed

hsmiss(Missing Any High School Data)
passed(Passed the course)

Frequency Row Pct	No	Yes	Total
No	178 40.92	257 59.08	435
Yes	96 66.67	48 33.33	144
Total	274	305	579

Statistics for Table of hsmiss by passed

Statistic	DF	Value	Prob
Chi-Square	1	28.7698	<.0001
Likelihood Ratio Chi-Square	1	29.0768	<.0001
Continuity Adj. Chi-Square	1	27.7463	<.0001
Mantel-Haenszel Chi-Square	1	28.7202	<.0001
Phi Coefficient		-0.2229	
Contingency Coefficient		0.2176	
Cramer's V		-0.2229	

Fisher's Exact Test

Cell (1,1) Frequency (F)	178
Left-sided Pr <= F	6.015E-08
Right-sided Pr >= F	1.0000
Table Probability (P)	4.001E-08
Two-sided Pr <= P	9.310E-08

Sample Size = 579

Logistic Regression with dummy variables
Course2 by passed with dummy vars

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	579

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	305
2	No	274

Probability modeled is passed='Yes'.

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	803.004	752.442
SC	807.365	769.887
-2 Log L	801.004	744.442

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	56.5621	3	<.0001
Score	54.7664	3	<.0001
Wald	50.4015	3	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.3545	0.3966	11.6672	0.0006
c0	1	-2.0896	0.4467	21.8848	<.0001
c1	1	-2.4307	0.4966	23.9535	<.0001
c2	1	-0.9468	0.4104	5.3225	0.0211

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
c0	0.124	0.052	0.297
c1	0.088	0.033	0.233
c2	0.388	0.174	0.867

Association of Predicted Probabilities and Observed Responses

Percent Concordant	43.1	Somers' D	0.302
Percent Discordant	12.8	Gamma	0.540
Percent Tied	44.1	Tau-a	0.151
Pairs	83570	c	0.651

Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
Course0_vs_3	21.8848	1	<.0001

Gender, Ethnicity and Math performance 7
Logistic Regression with dummy variables
Estimate prob. of passing for course=2: Compare 224/373=0.6005

	probpass
Estimated probability of passing course 2 is	0.6005363

Logistic Regression with dummy variables
Use the Class statement

The LOGISTIC Procedure

Model Information

Data Set	WORK.MATH	
Response Variable	passed	Passed the course
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	579
Number of Observations Used	579

Response Profile

Ordered Value	passed	Total Frequency
1	Yes	305
2	No	274

Probability modeled is passed='Yes'.

Class Level Information

Class	Value	Design Variables		
course2	0	1	0	0
	1	0	1	0
	2	0	0	1
	3	0	0	0

Model Convergence Status

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

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	803.004	752.442
SC	807.365	769.887
-2 Log L	801.004	744.442

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	56.5621	3	<.0001
Score	54.7664	3	<.0001
Wald	50.4015	3	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
course2	3	50.4015	<.0001

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.3545	0.3966	11.6672	0.0006
course2	0	1	-2.0896	0.4467	21.8848	<.0001
course2	1	1	-2.4307	0.4966	23.9535	<.0001
course2	2	1	-0.9468	0.4104	5.3225	0.0211

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
course2 0 vs 3	0.124	0.052	0.297
course2 1 vs 3	0.088	0.033	0.233
course2 2 vs 3	0.388	0.174	0.867

Gender, Ethnicity and Math performance 9
Logistic Regression with dummy variables
Use the Class statement

The LOGISTIC Procedure

Contrast Test Results

Contrast	DF	Wald Chi-Square	Pr > ChiSq
Course 0 vs 3	1	21.8848	<.0001