

Cell means coding and effect coding

```
/* mathregr_3.sas */
%include 'readmath.sas';
title2 'Illustrate more dummy variable coding schemes';

/* The data step continues */
if ethnic ne 6; /* Otherwise, throw the case out */

/* Indicator dummy variables for ethnic background
   1 = 'Asian'
   2 = 'Eastern European'
   3 = 'European not Eastern'
   4 = 'Middle-Eastern and Pakistani'
   5 = 'East Indian'
   6 was deleted. */

if ethnic=. then e1=.;
  else if ethnic=1 then e1=1;
  else e1=0;
if ethnic=. then e2=.;
  else if ethnic=2 then e2=1;
  else e2=0;
if ethnic=. then e3=.;
  else if ethnic=3 then e3=1;
  else e3=0;
if ethnic=. then e4=.;
  else if ethnic=4 then e4=1;
  else e4=0;
if ethnic=. then e5=.;
  else if ethnic=5 then e5=1;
  else e5=0;

proc glm;
  title3 'Repeat part of the one-way example';
  class ethnic;
  model grade = ethnic;
  contrast 'European_vs_Other' ethnic -2 -2 3 3 -2;

proc reg;
  title3 'Indicator dummy variables with intercept';
  model grade = e1-e4;
  European_vs_Other: test 2*e1 - 3*e2 - 3*e3 + 2*e4 = 0;

/* From both proc glm and indicator dummy variable coding with intercept,
   got overall F = 2.45, and for European vs. other, F = 6.40 */

proc reg;
  title3 'Cell means coding: overall F = 2.45, EvsOther F = 6.40';
  model grade = e1-e5 / noint;
  Overall: test e1=e2=e3=e4=e5;
  European_vs_Other: test 2*e1 - 3*e2 - 3*e3 + 2*e4 + 2*e5 = 0;

/* Careful: GLM uses alphabetical order. */
```

```

proc reg;
  title3 'Control for HS GPA';
  model grade = e1-e5 hsgpa / noint;
  Ethnic: test e1=e2=e3=e4=e5;
  European_vs_Other: test 2*e1 - 3*e2 - 3*e3 + 2*e4 + 2*e5 = 0;

proc glm;
  title3 'Ethnic controlling for HS GPA using proc glm';
  class ethnic;
  model grade = hsgpa ethnic;
  contrast 'European_vs_Other' ethnic -2 -2 3 3 -2;

/***** Effect coding *****/

%include 'readmath.sas';
title2 'Illustrate more dummy variable coding schemes';

/* The data step continues */
if ethnic ne 6; /* Otherwise, throw the case out */

if ethnic=. then e1=.;
  else if ethnic=1 then e1=1;
  else if ethnic=5 then e1 = -1;
  else e1=0;
if ethnic=. then e2=.;
  else if ethnic=2 then e2=1;
  else if ethnic=5 then e2 = -1;
  else e2=0;
if ethnic=. then e3=.;
  else if ethnic=3 then e3=1;
  else if ethnic=5 then e3 = -1;
  else e3=0;
if ethnic=. then e4=.;
  else if ethnic=4 then e4=1;
  else if ethnic=5 then e4 = -1;
  else e4=0;

proc freq;
  title3 'Check effect coding dummy vars';
  tables (e1-e4)*ethnic / norow nocol nopercent missing;

proc reg;
  title3 'Effect coding: Overall F = 2.45';
  model grade = e1-e4;

proc reg;
  title3 'Effect coding: Ethnic controlling for HSGPA F = 2.90';
  model grade = hsgpa e1-e4;
  ethnic: test e1=e2=e3=e4=0;

```

Skipping output for the `proc glm` and also the regression using indicator dummy variables with intercept. Both yielded overall $F = 2.45$, and for European vs. other, $F = 6.40$.

5

Gender, Ethnicity and Math performance
 Illustrate more dummy variable coding schemes
 Cell means coding: overall $F = 2.45$, EvsOther $F = 6.40$

The REG Procedure
 Model: MODEL1
 Dependent Variable: grade Final mark (if any)

Number of Observations Read	539
Number of Observations Used	378
Number of Observations with Missing Values	161

NOTE: No intercept in model. R-Square is redefined.

$$\text{Explaining } \sum_{i=1}^n Y_i^2 \text{ instead of } \sum_{i=1}^n (Y_i - \bar{Y})^2$$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1308038	261608	704.31	<.0001
Error	373	138545	371.43552		
Uncorrected Total	378	1446583			

Root MSE	19.27266	R-Square	0.9042
Dependent Mean	58.74339	Adj R-Sq	0.9029
Coeff Var	32.80823		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
e1		1	60.05747	2.06625	29.07	<.0001
e2		1	55.76087	2.84160	19.62	<.0001
e3		1	56.28169	1.61733	34.80	<.0001
e4		1	59.36000	2.72557	21.78	<.0001
e5		1	65.18868	2.64730	24.62	<.0001

Gender, Ethnicity and Math performance 6
Illustrate more dummy variable coding schemes
Cell means coding: overall F = 2.45, EvsOther F = 6.40

The REG Procedure
Model: MODEL1

Test Overall Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	910.16516	2.45	0.0458
Denominator	373	371.43552		

Gender, Ethnicity and Math performance 7
Illustrate more dummy variable coding schemes
Cell means coding: overall F = 2.45, EvsOther F = 6.40

The REG Procedure
Model: MODEL1

Test European_vs_Other Results
for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	2377.06184	6.40	0.0118
Denominator	373	371.43552		

Gender, Ethnicity and Math performance
 Illustrate more dummy variable coding schemes
 Control for HS GPA

8

The REG Procedure
 Model: MODEL1
 Dependent Variable: grade Final mark (if any)

Number of Observations Read	539
Number of Observations Used	329
Number of Observations with Missing Values	210

NOTE: No intercept in model. R-Square is redefined.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1201998	200333	883.25	<.0001
Error	323	73261	226.81399		
Uncorrected Total	329	1275259			

Root MSE	15.06034	R-Square	0.9426
Dependent Mean	59.38906	Adj R-Sq	0.9415
Coeff Var	25.35879		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
e1		1	-81.89552	11.31897	-7.24	<.0001
e2		1	-85.95513	11.29739	-7.61	<.0001
e3		1	-86.94550	11.29315	-7.70	<.0001
e4		1	-83.74110	11.49744	-7.28	<.0001
e5		1	-79.42828	11.59274	-6.85	<.0001
hsgpa	High School GPA	1	1.78130	0.13914	12.80	<.0001

Gender, Ethnicity and Math performance
Illustrate more dummy variable coding schemes
Control for HS GPA

9

The REG Procedure
Model: MODEL1

Test Ethnic Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	657.46812	2.90	0.0222
Denominator	323	226.81399		

Gender, Ethnicity and Math performance
Illustrate more dummy variable coding schemes
Control for HS GPA

10

The REG Procedure
Model: MODEL1

Test European_vs_Other Results
for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	1	1463.96963	6.45	0.0115
Denominator	323	226.81399		

Gender, Ethnicity and Math performance
 Ethnic controlling for HS GPA using proc glm
 The GLM Procedure

Class Level Information

Class	Levels	Values
ethnic	5	Asian East Indian Eastern European European not Eastern Middle-Eastern and Pakistani

Number of Observations Read	539
Number of Observations Used	329

Gender, Ethnicity and Math performance 12

The GLM Procedure

Dependent Variable: grade Final mark (if any)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	41595.2824	8319.0565	36.68	<.0001
Error	323	73260.9182	226.8140		
Corrected Total	328	114856.2006			

R-Square	Coeff Var	Root MSE	grade Mean
0.362151	25.35879	15.06034	59.38906

Source	DF	Type I SS	Mean Square	F Value	Pr > F
hsgpa	1	38965.40989	38965.40989	171.79	<.0001
ethnic	4	2629.87247	657.46812	2.90	0.0222

Source	DF	Type III SS	Mean Square	F Value	Pr > F
hsgpa	1	37174.97070	37174.97070	163.90	<.0001
ethnic	4	2629.87247	657.46812	2.90	0.0222

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
European_vs_Other	1	1463.969633	1463.969633	6.45	0.0115

Check effect coding dummy vars

The FREQ Procedure

Table of e1 by ethnic

e1		ethnic(Judged Nationality of name)					Total
Frequency	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian		
-1	0	0	0	0	78	78	
0	0	63	195	72	0	330	
1	131	0	0	0	0	131	
Total	131	63	195	72	78	539	

Table of e2 by ethnic

e2		ethnic(Judged Nationality of name)					Total
Frequency	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian		
-1	0	0	0	0	78	78	
0	131	0	195	72	0	398	
1	0	63	0	0	0	63	
Total	131	63	195	72	78	539	

Table of e3 by ethnic

e3		ethnic(Judged Nationality of name)					Total
Frequency	Asian	Eastern European	European not Eastern	Middle-Eastern and Pakistani	East Indian		
-1	0	0	0	0	78	78	
0	131	63	0	72	0	266	
1	0	0	195	0	0	195	
Total	131	63	195	72	78	539	

Gender, Ethnicity and Math performance
 Illustrate more dummy variable coding schemes
 Check effect coding dummy vars

14

The FREQ Procedure

Table of e4 by ethnic

```
e4      ethnic(Judged Nationality of name)
```

Frequency	Asian	Eastern European	European not Eas tern	Middle-E astern and Paki stani	East Ind ian	Total
-1	0	0	0	0	78	78
0	131	63	195	0	0	389
1	0	0	0	72	0	72
Total	131	63	195	72	78	539

Gender, Ethnicity and Math performance
 Illustrate more dummy variable coding schemes
 Effect coding: Overall F = 2.45

15

The REG Procedure

Model: MODEL1

Dependent Variable: grade Final mark (if any)

Number of Observations Read	539
Number of Observations Used	378
Number of Observations with Missing Values	161

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	3640.66065	910.16516	2.45	0.0458
Error	373	138545	371.43552		
Corrected Total	377	142186			

Root MSE	19.27266	R-Square	0.0256
Dependent Mean	58.74339	Adj R-Sq	0.0152
Coeff Var	32.80823		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	59.32974	1.08438	54.71	<.0001
e1		1	0.72773	1.93326	0.38	0.7068
e2		1	-3.56887	2.45371	-1.45	0.1467
e3		1	-3.04805	1.65690	-1.84	0.0666
e4		1	0.03026	2.37341	0.01	0.9898

Effect coding: Ethnic controlling for HSGPA F = 2.90

The REG Procedure
 Dependent Variable: grade Final mark (if any)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	41595	8319.05647	36.68	<.0001
Error	323	73261	226.81399		
Corrected Total	328	114856			

Root MSE 15.06034 R-Square 0.3622
 Dependent Mean 59.38906 Adj R-Sq 0.3523
 Coeff Var 25.35879

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-83.59311	11.24743	-7.43	<.0001
hsgpa	High School GPA	1	1.78130	0.13914	12.80	<.0001
e1		1	1.69759	1.64421	1.03	0.3026
e2		1	-2.36203	2.13927	-1.10	0.2704
e3		1	-3.35239	1.37935	-2.43	0.0156
e4		1	-0.14799	2.13123	-0.07	0.9447

Effect coding: Ethnic controlling for HSGPA F = 2.90

The REG Procedure

Test ethnic Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	657.46812	2.90	0.0222
Denominator	323	226.81399		