

Regression: Part One

```
/* mathregr_1.sas */
%include 'readmath.sas';
title2 'Regression to predict university calculus mark';

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

/* Dummy variables for ethnic background */
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;

/* Ethnic
      1 = 'Asian'
      2 = 'Eastern European'
      3 = 'European not Eastern'
      4 = 'Middle-Eastern and Pakistani'
      5 = 'East Indian'
      6 was deleted.
*/
label e1 = 'Asian vs East Ind.'
      e2 = 'East Eur. vs East Ind.'
      e3 = 'Other Eur. vs East Ind.'
      e4 = 'Mid. East & Pak. vs East Ind.';

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;

proc freq;
  title3 'Check dummy variables';
  tables sex*gender / norow nocol nopercnt missing;
  tables tongue*mtongue / norow nocol nopercnt missing;
  tables (e1-e4) * ethnic / norow nocol nopercnt missing;

proc reg;
  title3 'Predict University Calculus Grade from HS Information';
  model grade = hsgpa hscalcalc hsenl;

proc reg;
  title3 'Predict University Calculus Grade from Diagnostic Test';
  model grade = precalc calc;

proc reg;
  title3 'Do the diagnostic test and HS info both contribute?';
  model grade = hsgpa hscalcalc hsenl precalc calc;
```

```

Diagnostic_Test: test precalc=calc=0;
HS_Information: test hsgpa=hscalcalc=hsengl=0;

proc iml;
  title3 'Proportion of remaining variation';
  print "Diag. controlling for HS";
  a = 2*8.40/(276+2*8.40); print a;
  print "HS controlling for Diag.";
  a = 3*44.95/(276+3*44.95); print a;

proc glm;
  title3 'Repeat part of the one-way example';
  class ethnic;
  model grade = ethnic;
  contrast 'East Indian vs. Middle-Eastern and Pakistani'
           ethnic 0 1 0 0 -1;
  contrast 'East Indian vs. Eastern European'
           ethnic 0 1 -1 0 0;
  contrast 'European_vs_Other' ethnic -2 -2 3 3 -2;
  contrast 'Non-Europ_Equal?' ethnic 1 -1 0 0 0,
           ethnic 1 0 0 0 -1;

proc reg;
  title3 'Replicate the one-way for ethnic background';
  model grade = e1-e4;
  European_vs_Other: test 2*e1 - 3*e2 - 3*e3 + 2*e4 = 0;
  Non_Europ_Equal: test e1=e4=0;
  /* Could separate the two statements with commas, like e1=0, e4=0; */

```

/* Here's how we get the European_vs_Other null hypothesis.

	e1	e2	e3	e4	Predicted Y = Y-bar	
1 = 'Asian'	1	0	0	0	$b_0 + b_1$	= Y-bar1
2 = 'Eastern European'	0	1	0	0	$b_0 + b_2$	= Y-bar2
3 = 'European not Eastern'	0	0	1	0	$b_0 + b_3$	= Y-bar3
4 = 'Middle-Eastern and Pakistani'	0	0	0	1	$b_0 + b_4$	= Y-bar4
5 = 'East Indian'	0	0	0	0	b_0	= Y-bar5

So for the contrast European_vs_Other, $H_0: 2\mu_1 - 3\mu_2 - 3\mu_3 + 2\mu_4 = 0$

is the same as

$$\begin{aligned}
 & 2(\beta_0 + \beta_1) \\
 & -3(\beta_0 + \beta_2) \\
 & -3(\beta_0 + \beta_3) \\
 & +2(\beta_0 + \beta_4) \\
 & +2(\beta_0)
 \end{aligned}
 = 0$$

If and only if

$$2\beta_1 - 3\beta_2 - 3\beta_3 + 2\beta_4 = 0$$

The variable names stand for the regression coefficients, and so we get our test statement. But all this algebra is inconvenient. We will see later that it's MUCH nicer with a different dummy variable coding scheme.

Some special contrasts are not so bad. Non_Europ_Equal is pretty straightforward because the others are all equal to East Indian, the reference category.

*/

```

proc reg;
  title3 'Does ethnic background help predict grade controlling for HS,
D?';
  model grade = hsgpa hscalcl hsenl precalc calc e1-e4;
  Ethnic_Background: test e1=e2=e3=e4=0;

proc glm;
  title3 'Repeat the last analysis with proc glm';
  class ethnic;
  model grade = hsgpa hscalcl hsenl precalc calc ethnic;
  lsmeans ethnic;

proc reg;
  title3 'Try gender and mother tongue as well as ethnic';
  model grade = hsgpa hscalcl hsenl precalc calc e1-e4 gender mtongue;
  Ethnic_Background: test e1=e2=e3=e4=0;

proc glm;
  title3 'Do the last analysis with proc glm';
  class sex ethnic tongue;
  model grade = hsgpa hscalcl hsenl precalc calc ethnic sex tongue;

proc reg;
  title3 'Finally, a nice prediction model, with residuals';
  model grade = hsgpa hscalcl hsenl precalc mtongue;
  output out=resdata predicted=pregrade residual=regrade;
  /* Creates new SAS data set called resdata (name arbitrary) with
     Y-hat and e as additional variables*/

/* By default, all SAS procedures use the most recently created data set.
   This can be over-ridden with the data= option. */

options pagesize=35;
proc univariate normal plot;
  title3 'Examine residuals';
  var regrade;
proc plot;
  plot grade*pregrade;
proc corr;
  var grade pregrade;

```

Skipping out put from the proc freq that check the dummy variables ...

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Predict University Calculus Grade from HS Information

3

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

Number of Observations Read	539
Number of Observations Used	316
Number of Observations with Missing Values	223

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	44549	14850	70.94	<.0001
Error	312	65311	209.33079		
Corrected Total	315	109860			

Root MSE	14.46827	R-Square	0.4055
Dependent Mean	59.91772	Adj R-Sq	0.3998
Coeff Var	24.14689		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-77.28860	11.11074	-6.96	<.0001
hsgpa	High School GPA	1	1.77965	0.22469	7.92	<.0001
hscal	HS Calculus	1	0.30806	0.09906	3.11	0.0020
hsengl	HS English	1	-0.39625	0.12045	-3.29	0.0011

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Predict University Calculus Grade from Diagnostic Test

4

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

Number of Observations Read	539
Number of Observations Used	338
Number of Observations with Missing Values	201

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	24337	12168	40.71	<.0001
Error	335	100128	298.88908		
Corrected Total	337	124465			

Root MSE	17.28841	R-Square	0.1955
Dependent Mean	59.65385	Adj R-Sq	0.1907
Coeff Var	28.98121		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	37.12530	2.93070	12.67
precalc	Number precalculus correct	1	3.24835	0.64850	5.01
calc	Number calculus correct	1	1.82839	0.41517	4.40

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
precalc	Number precalculus correct	1	<.0001
calc	Number calculus correct	1	<.0001

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Do the diagnostic test and HS info both contribute?

5

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

Number of Observations Read	539
Number of Observations Used	282
Number of Observations with Missing Values	257

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	43579	8715.77813	45.93	<.0001
Error	276	52380	189.78149		
Corrected Total	281	95959			

Root MSE	13.77612	R-Square	0.4541
Dependent Mean	60.73759	Adj R-Sq	0.4443
Coeff Var	22.68137		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-71.56814	11.36993	-6.29
hsgpa	High School GPA	1	1.59779	0.22503	7.10
hscal	HS Calculus	1	0.25559	0.10336	2.47
hsengl	HS English	1	-0.36046	0.12310	-2.93
precalc	Number precalculus correct	1	1.62221	0.56984	2.85
calc	Number calculus correct	1	0.66064	0.38357	1.72

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0140
hsengl	HS English	1	0.0037
precalc	Number precalculus correct	1	0.0047
calc	Number calculus correct	1	0.0861

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Do the diagnostic test and HS info both contribute?

6

The REG Procedure
 Model: MODEL1

Test Diagnostic_Test Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	1594.63412	8.40	0.0003
Denominator	276	189.78149		

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Do the diagnostic test and HS info both contribute?

7

The REG Procedure
 Model: MODEL1

Test HS_Information Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	8531.43803	44.95	<.0001
Denominator	276	189.78149		

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Proportion of remaining variation

8

Diag. controlling for HS

a

0.057377

HS controlling for Diag.

a

0.328222

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Repeat part of the one-way example

9

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 378

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Repeat part of the one-way example

10

The GLM Procedure

Dependent Variable: grade Final mark (if any)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	3640.6607	910.1652	2.45	0.0458
Error	373	138545.4478	371.4355		
Corrected Total	377	142186.1085			

R-Square	Coeff Var	Root MSE	grade Mean
0.025605	32.80823	19.27266	58.74339

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ethnic	4	3640.660655	910.165164	2.45	0.0458

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ethnic	4	3640.660655	910.165164	2.45	0.0458

Contrast	DF	Contrast SS
East Indian vs. Middle-Eastern and Pakistani	1	874.075530
East Indian vs. Eastern European	1	2188.870763
European_vs_Other	1	2377.061841
Non-Europ_Equal?	2	1123.948886

Contrast	Mean Square	F Value	Pr > F
East Indian vs. Middle-Eastern and Pakistani	874.075530	2.35	0.1259
East Indian vs. Eastern European	2188.870763	5.89	0.0157
European_vs_Other	2377.061841	6.40	0.0118
Non-Europ_Equal?	561.974443	1.51	0.2216

11

Gender, Ethnicity and Math performance
Regression to predict university calculus mark
Replicate the one-way for ethnic background

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

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
Intercept	Intercept	1	65.18868	2.64730	24.62
e1	Asian vs East Ind.	1	-5.13121	3.35821	-1.53
e2	East Eur. vs East Ind.	1	-9.42781	3.88367	-2.43
e3	Other Eur. vs East Ind.	1	-8.90699	3.10225	-2.87
e4	Mid. East & Pak. vs East Ind.	1	-5.82868	3.79960	-1.53

For East Indian vs. Eastern European, $F = 5.89 = (-2.43)^2$, except for rounding error. And for East Indian vs. Middle-Eastern and Pakistani, $F = 2.35 = (-1.53)^2$, close enough. p-values are identical.

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
e1	Asian vs East Ind.	1	0.1274
e2	East Eur. vs East Ind.	1	0.0157
e3	Other Eur. vs East Ind.	1	0.0043
e4	Mid. East & Pak. vs East Ind.	1	0.1259

Gender, Ethnicity and Math performance 12
 Regression to predict university calculus mark
 Replicate the one-way for ethnic background

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 13
 Regression to predict university calculus mark
 Replicate the one-way for ethnic background

The REG Procedure
 Model: MODEL1

Test Non_Europ_Equal Results for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	561.97444	1.51	0.2216
Denominator	373	371.43552		

Repeating proc glm output, ...

Contrast	Mean Square	F Value	Pr > F
East Indian vs. Middle-Eastern and Pakistani	874.075530	2.35	0.1259
East Indian vs. Eastern European	2188.870763	5.89	0.0157
European_vs_Other	2377.061841	6.40	0.0118
Non-Europ_Equal?	561.974443	1.51	0.2216

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Does ethnic background help predict grade controlling for HS, D?

14

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

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	44579	4953.17604	26.22	<.0001
Error	272	51380	188.89705		
Corrected Total	281	95959			

Root MSE	13.74398	R-Square	0.4646
Dependent Mean	60.73759	Adj R-Sq	0.4468
Coeff Var	22.62846		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-63.98252	11.84759	-5.40
hsgpa	High School GPA	1	1.59382	0.22550	7.07
hscal	HS Calculus	1	0.24563	0.10405	2.36
hsengl	HS English	1	-0.38863	0.12596	-3.09
precalc	Number precalculus correct	1	1.65613	0.56910	2.91
calc	Number calculus correct	1	0.63872	0.39008	1.64
e1	Asian vs East Ind.	1	-5.39910	2.78520	-1.94
e2	East Eur. vs East Ind.	1	-4.91422	3.21062	-1.53
e3	Other Eur. vs East Ind.	1	-4.86493	2.52155	-1.93
e4	Mid. East & Pak. vs East Ind.	1	-6.17280	3.25801	-1.89

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0189
hsengl	HS English	1	0.0022
precalc	Number precalculus correct	1	0.0039
calc	Number calculus correct	1	0.1027
e1	Asian vs East Ind.	1	0.0536
e2	East Eur. vs East Ind.	1	0.1270
e3	Other Eur. vs East Ind.	1	0.0547
e4	Mid. East & Pak. vs East Ind.	1	0.0592

Gender, Ethnicity and Math performance 15
 Regression to predict university calculus mark
 Does ethnic background help predict grade controlling for HS, D?

The REG Procedure
 Model: MODEL1

Test Ethnic_Background Results
 for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	249.92343	1.32	0.2616
Denominator	272	188.89705		

Gender, Ethnicity and Math performance 16
 Regression to predict university calculus mark
 Repeat the last analysis with 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 282

Gender, Ethnicity and Math performance 17
 Regression to predict university calculus mark
 Repeat the last analysis with proc glm

The GLM Procedure

Dependent Variable: grade		Final mark (if any)			
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	44578.58439	4953.17604	26.22	<.0001
Error	272	51379.99717	188.89705		
Corrected Total	281	95958.58156			

R-Square Coeff Var Root MSE grade Mean
 0.464561 22.62846 13.74398 60.73759

Source	DF	Type I SS	Mean Square	F Value	Pr > F
hsgpa	1	33751.72367	33751.72367	178.68	<.0001
hscal	1	4714.48430	4714.48430	24.96	<.0001
hsengl	1	1923.41444	1923.41444	10.18	0.0016
precalc	1	2626.28543	2626.28543	13.90	0.0002
calc	1	562.98282	562.98282	2.98	0.0854
ethnic	4	999.69374	249.92343	1.32	0.2616

Source	DF	Type III SS	Mean Square	F Value	Pr > F
hsgpa	1	9436.196126	9436.196126	49.95	<.0001
hscal	1	1052.698571	1052.698571	5.57	0.0189
hsengl	1	1798.073019	1798.073019	9.52	0.0022
precalc	1	1599.703454	1599.703454	8.47	0.0039
calc	1	506.463378	506.463378	2.68	0.1027
ethnic	4	999.693738	249.923435	1.32	0.2616

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Repeat the last analysis with proc glm

18

The GLM Procedure
 Least Squares Means

ethnic	grade LSMEAN
Asian	59.7306552
East Indian	65.1297518
Eastern European	60.2155366
European not Eastern	60.2648246
Middle-Eastern and Pakistani	58.9569552

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Try gender and mother tongue as well as ethnic

19

The REG Procedure

Model: MODEL1

Dependent Variable: grade Final mark (if any)

Number of Observations Read	539
Number of Observations Used	280
Number of Observations with Missing Values	259

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	45834	4166.72192	22.34	<.0001
Error	268	49982	186.49830		
Corrected Total	279	95815			

Root MSE	13.65644	R-Square	0.4784
Dependent Mean	60.75714	Adj R-Sq	0.4569
Coeff Var	22.47709		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-59.95382	12.01497	-4.99
hsgpa	High School GPA	1	1.58242	0.22547	7.02
hscal	HS Calculus	1	0.21935	0.10433	2.10
hsengl	HS English	1	-0.38188	0.13107	-2.91
precalc	Number precalculus correct	1	1.83615	0.57567	3.19
calc	Number calculus correct	1	0.64267	0.38948	1.65
e1	Asian vs East Ind.	1	-5.57921	2.79547	-2.00
e2	East Eur. vs East Ind.	1	-5.59024	3.24184	-1.72
e3	Other Eur. vs East Ind.	1	-4.14163	2.55826	-1.62
e4	Mid. East & Pak. vs East Ind.	1	-6.01137	3.25759	-1.85

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0365
hsengl	HS English	1	0.0039
precalc	Number precalculus correct	1	0.0016
calc	Number calculus correct	1	0.1001
e1	Asian vs East Ind.	1	0.0470
e2	East Eur. vs East Ind.	1	0.0858
e3	Other Eur. vs East Ind.	1	0.1066
e4	Mid. East & Pak. vs East Ind.	1	0.0661

20

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Try gender and mother tongue as well as ethnic

The REG Procedure

Model: MODEL1

Dependent Variable: grade Final mark (if any)

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
gender		1	2.23345	1.76445	1.27
mtongue		1	-4.67856	2.21309	-2.11

Parameter Estimates

Variable	Label	DF	Pr > t
gender		1	0.2067
mtongue		1	0.0354

Try gender and mother tongue as well as ethnic

The REG Procedure
Model: MODEL1

Test Ethnic_Background Results
for Dependent Variable grade

Source	DF	Mean Square	F Value	Pr > F
Numerator	4	246.74959	1.32	0.2616
Denominator	268	186.49830		

Gender, Ethnicity and Math performance 22
Regression to predict university calculus mark
Do the last analysis with proc glm

The GLM Procedure

Class Level Information

Class	Levels	Values
sex	2	Female Male
ethnic	5	Asian East Indian Eastern European European not Eastern Middle-Eastern and Pakistani
tongue	2	English Other

Number of Observations Read 539
Number of Observations Used 280

Gender, Ethnicity and Math performance 23
Regression to predict university calculus mark
Do the last analysis with proc glm

The GLM Procedure

Dependent Variable: grade Final mark (if any)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	45833.94113	4166.72192	22.34	<.0001
Error	268	49981.54459	186.49830		
Corrected Total	279	95815.48571			

R-Square	Coeff Var	Root MSE	grade Mean
0.478356	22.47709	13.65644	60.75714

Source	DF	Type I SS	Mean Square	F Value	Pr > F
hsgpa	1	33894.43672	33894.43672	181.74	<.0001
hscalc	1	4553.50249	4553.50249	24.42	<.0001
hsengl	1	1958.84352	1958.84352	10.50	0.0013
precalc	1	2755.46153	2755.46153	14.77	0.0002
calc	1	501.71761	501.71761	2.69	0.1021
ethnic	4	919.69067	229.92267	1.23	0.2972
sex	1	416.79877	416.79877	2.23	0.1361
tongue	1	833.48981	833.48981	4.47	0.0354

Source	DF	Type III SS	Mean Square	F Value	Pr > F
hsgpa	1	9186.039470	9186.039470	49.26	<.0001
hscalc	1	824.345755	824.345755	4.42	0.0365
hsengl	1	1583.016799	1583.016799	8.49	0.0039
precalc	1	1897.361570	1897.361570	10.17	0.0016
calc	1	507.786374	507.786374	2.72	0.1001
ethnic	4	986.998347	246.749587	1.32	0.2616
sex	1	298.819070	298.819070	1.60	0.2067
tongue	1	833.489811	833.489811	4.47	0.0354

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Finally, a nice prediction model, with residuals

24

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

Number of Observations Read	539
Number of Observations Used	280
Number of Observations with Missing Values	259

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	44143	8828.57493	46.81	<.0001
Error	274	51673	188.58617		
Corrected Total	279	95815			

Root MSE	13.73267	R-Square	0.4607
Dependent Mean	60.75714	Adj R-Sq	0.4509
Coeff Var	22.60256		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value
Intercept	Intercept	1	-70.60227	11.40402	-6.19
hsgpa	High School GPA	1	1.59555	0.22518	7.09
hscal	HS Calculus	1	0.25967	0.10170	2.55
hsengl	HS English	1	-0.31484	0.12548	-2.51
precalc	Number precalculus correct	1	2.02100	0.53182	3.80
mtongue		1	-4.91192	2.15404	-2.28

Parameter Estimates

Variable	Label	DF	Pr > t
Intercept	Intercept	1	<.0001
hsgpa	High School GPA	1	<.0001
hscal	HS Calculus	1	0.0112
hsengl	HS English	1	0.0127
precalc	Number precalculus correct	1	0.0002
mtongue		1	0.0234

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Examine residuals

26

The UNIVARIATE Procedure
 Variable: regrade (Residual)

Moments

N	280	Sum Weights	280
Mean	0	Sum Observations	0
Std Deviation	13.6090592	Variance	185.206491
Skewness	-0.6648322	Kurtosis	0.69501632
Uncorrected SS	51672.6111	Corrected SS	51672.6111
Coeff Variation	.	Std Error Mean	0.81329684

Basic Statistical Measures

Location		Variability	
Mean	0.000000	Std Deviation	13.60906
Median	1.005467	Variance	185.20649
Mode	.	Range	76.97547
		Interquartile Range	16.29163

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 0	Pr > t	1.0000
Sign	M 7	Pr >= M	0.4373
Signed Rank	S 1400	Pr >= S	0.3028

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Examine residuals

27

The UNIVARIATE Procedure
 Variable: regrade (Residual)

Tests for Normality

Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.972669	Pr < W	<0.0001
Kolmogorov-Smirnov	D 0.068705	Pr > D	<0.0100
Cramer-von Mises	W-Sq 0.286016	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 1.866825	Pr > A-Sq	<0.0050

Quantiles (Definition 5)

Quantile	Estimate
100% Max	29.53659
99%	26.50443
95%	20.44278
90%	16.16616
75% Q3	9.43571
50% Median	1.00547
25% Q1	-6.85592
10%	-17.86273
5%	-27.36028
1%	-38.04585
0% Min	-47.43888

Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Examine residuals

28

The UNIVARIATE Procedure
 Variable: regrade (Residual)

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-47.4389	8	25.6063	31
-46.6746	163	25.8727	188
-38.0459	213	26.5044	428
-37.8947	180	26.9083	328
-34.9631	156	29.5366	136

Missing Values

Missing Value	Count	-----Percent Of-----	
		All Obs	Missing Obs
.	259	48.05	100.00

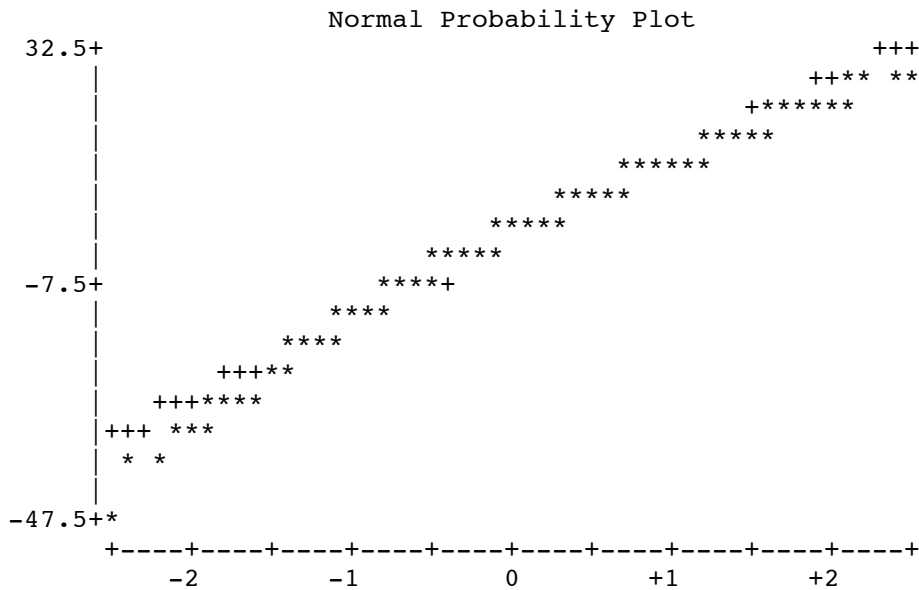
Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Examine residuals

The UNIVARIATE Procedure
 Variable: regrade (Residual)

Stem Leaf	#	Boxplot
3 0	1	
2 56677	5	
2 0011123344	10	
1 5666677777788889	16	
1 00000000001111111122222233333344444	37	
0 55555555666666666677777778888888999999999999999	46	+-----+
0 0011111111112222223333333344444	32	*---*---
-0 4444444444433333332222221111110000000000	42	
-0 999998877777777766666655555555555	34	+-----+
-1 44443222221111111000	20	
-1 998777766655	12	
-2 3211110	7	
-2 9999777765	10	
-3 321	3	0
-3 885	3	0
-4		
-4 77	2	0

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 Multiply Stem.Leaf by 10**+1

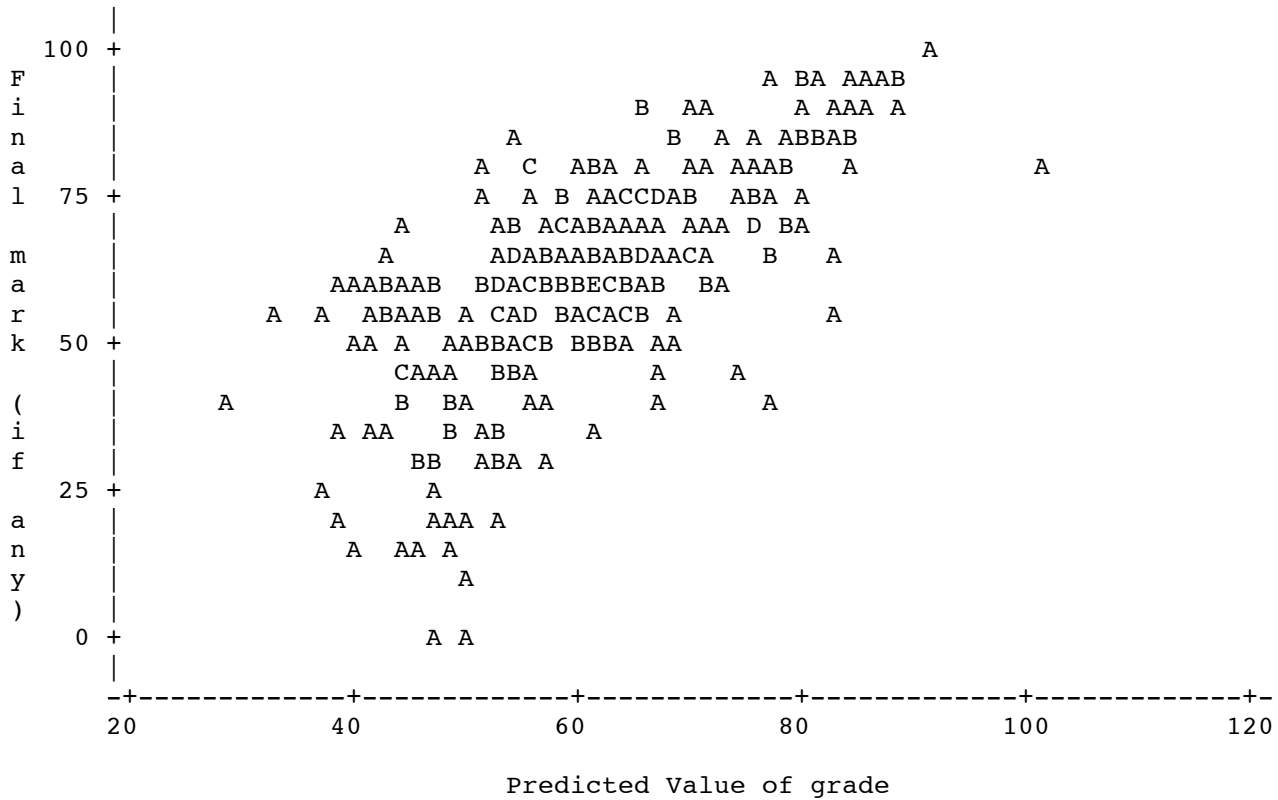
The UNIVARIATE Procedure
 Variable: regrade (Residual)



Gender, Ethnicity and Math performance
 Regression to predict university calculus mark
 Examine residuals

31

Plot of grade*pregrade. Legend: A = 1 obs, B = 2 obs, etc.



NOTE: 259 obs had missing values.

The CORR Procedure

Pearson Correlation Coefficients
 Prob > |r| under H0: Rho=0
 Number of Observations

	grade	pregrade
grade	1.00000	0.67875
Final mark (if any)		<.0001
	378	280
pregrade	0.67875	1.00000
Predicted Value of grade	<.0001	
	280	360